

N°

N° 464

**ASSEMBLÉE NATIONALE**

**SÉNAT**

CONSTITUTION DU 4 OCTOBRE 1958  
QUATORZIÈME LÉGISLATURE

SESSION ORDINAIRE 2016 - 2017

Enregistré à la présidence de l'Assemblée nationale

Enregistré à la présidence du Sénat

le 15 mars 2017

le 15 mars 2017

**RAPPORT**

*au nom de*

**L'OFFICE PARLEMENTAIRE D'ÉVALUATION  
DES CHOIX SCIENTIFIQUES ET TECHNOLOGIQUES**

**POUR UNE INTELLIGENCE ARTIFICIELLE MAÎTRISÉE, UTILE ET DÉMYSTIFIÉE**

PAR

M. Claude de GANAY, député, et Mme Dominique GILLOT, sénatrice

---

Déposé sur le Bureau de l'Assemblée nationale  
par M. Jean-Yves LE DÉAUT,  
*Président de l'Office*

---

Déposé sur le Bureau du Sénat  
par M. Bruno SIDO,  
*Premier vice-président de l'Office*

## **Composition de l'Office parlementaire d'évaluation des choix scientifiques et technologiques**

### **Président**

M. Jean-Yves LE DÉAUT, député

### **Premier vice-président**

M. Bruno SIDO, sénateur

### **Vice-présidents**

M. Christian BATAILLE, député  
Mme Anne-Yvonne LE DAIN, députée  
M. Jean-Sébastien VIALATTE, député

M. Roland COURTEAU, sénateur  
M. Christian NAMY, sénateur  
Mme Catherine PROCACCIA, sénateur

### **DÉPUTÉS**

M. Bernard ACCOYER  
M. Gérard BAPT  
M. Alain CLAEYS  
M. Claude de GANAY  
Mme Françoise GUÉGOT  
M. Patrick HETZEL  
M. Laurent KALINOWSKI  
M. Alain MARTY  
M. Philippe NAUCHE  
Mme Maud OLIVIER  
Mme Dominique ORLIAC  
M. Bertrand PANCHER  
M. Jean-Louis TOURAINE  
N.

### **SÉNATEURS**

M. Patrick ABATE  
M. Gilbert BARBIER  
Mme Delphine BATAILLE  
M. Michel BERSON  
M. François COMMEINHES  
Mme Catherine GÉNISSON  
Mme Dominique GILLOT  
M. Alain HOUPERT  
Mme Fabienne KELLER  
M. Jean-Pierre LELEUX  
M. Gérard LONGUET  
M. Pierre MÉDEVIELLE  
M. Franck MONTAUGÉ  
M. Hervé POHER

*« Science sans conscience n'est que ruine de l'âme »*

**Rabelais**

*« Dans la vie, rien est à craindre, tout est à comprendre »*

**Marie Curie**

*« L'intelligence, ça n'est pas ce que l'on sait, mais  
ce que l'on fait quand on ne sait pas »*

**Jean Piaget**



---

## SOMMAIRE

	<u>Pages</u>
<b>SYNTHÈSE DU RAPPORT</b> .....	11
<b>I. SYNTHÈSE : LES OBSERVATIONS DU RAPPORT</b> .....	11
<b>II. SYNTHÈSE : LES PROPOSITIONS DU RAPPORT</b> .....	13
<b>INTRODUCTION</b> .....	15
<b>PREMIÈRE PARTIE : ÉLÉMENTS DE CONTEXTE</b> .....	19
<b>I. LA DÉMARCHE DE VOS RAPPORTEURS</b> .....	19
<b>A. DE LA PROCÉDURE DE SAISINE À L'ADOPTION D'UN CALENDRIER DE TRAVAIL</b> .....	19
1. <i>L'origine et l'instruction de la saisine</i> .....	19
2. <i>Un calendrier de travail enfermé dans un délai contraint de six mois</i> .....	21
<b>B. LE CHAMP DES INVESTIGATIONS DE L'ÉTUDE</b> .....	21
1. <i>L'étude de faisabilité du rapport conclut à la pertinence d'une étude spécifique de l'OPECST</i> .....	21
2. <i>Un ciblage délibéré sur un nombre limité de problématiques et de pistes d'investigation</i> .....	24
<b>C. LA MÉTHODE DE TRAVAIL</b> .....	27
1. <i>Une méthode de travail fondée sur des auditions bilatérales et des déplacements en France et à l'étranger</i> .....	27
2. <i>L'organisation d'une journée entière d'auditions publiques au Sénat le 19 janvier 2017</i> .....	28
3. <i>La consultation d'ouvrages, de rapports et d'articles parus sur le sujet</i> .....	28
<b>II. L'HISTOIRE DES TECHNOLOGIES D'INTELLIGENCE ARTIFICIELLE ET DE LEURS USAGES</b> .....	31
<b>A. DES TECHNOLOGIES NÉES AU MILIEU DU XX<sup>E</sup> SIÈCLE</b> .....	31
1. <i>La préhistoire de l'intelligence artificielle et sa présence dans les œuvres de fiction</i> .....	31
2. <i>Les premières étapes de formation des technologies d'intelligence artificielle au XX<sup>e</sup> siècle, la notion d'algorithme et le débat sur la définition du concept d'intelligence artificielle</i> .....	33
3. <i>« L'âge d'or » des approches symboliques et des raisonnements logiques dans les années 1960 a été suivi d'un premier « hiver de l'intelligence artificielle » dans les années 1970</i> .....	38
4. <i>Un enthousiasme renouvelé dans les années 1980 autour des systèmes experts, de leurs usages et de l'ingénierie des connaissances précède un second « hiver de l'intelligence artificielle » dans les années 1990</i> .....	40
5. <i>Les autres domaines et technologies d'intelligence artificielle : robotique, systèmes multi-agents, machines à vecteur de support (SVM), réseaux bayésiens, apprentissage machine dont apprentissage par renforcement, programmation par contraintes, raisonnements à partir de cas, ontologies, logiques de description, algorithmes génétiques...</i> .....	42

<b>B. L'ACCÉLÉRATION RÉCENTE DE L'USAGE DES TECHNOLOGIES D'INTELLIGENCE ARTIFICIELLE GRÂCE AUX PROGRÈS EN APPRENTISSAGE AUTOMATIQUE (« MACHINE LEARNING »)</b> .....	52
1. <i>Les découvertes en apprentissage profond (« deep learning ») remontent surtout aux années 1980, par un recours aux « réseaux de neurones artificiels » imaginés dès les années 1940</i> .....	52
2. <i>L'apprentissage profond connaît un essor inédit dans les années 2010 avec l'émergence de données massives (« big data ») et l'accélération de la vitesse de calcul des processeurs</i> ....	53
3. <i>Les technologies d'intelligence artificielle conduisent d'ores et déjà à des applications dans de nombreux secteurs</i> .....	60
4. <i>Par leurs combinaisons en évolution constante, ces technologies offrent un immense potentiel et ouvrent un espace d'opportunités transversal inédit</i> .....	67
5. <i>L'apprentissage automatique reste encore largement supervisé et fait face au défi de l'apprentissage non-supervisé</i> .....	69
 <b>III. CARACTÉRISTIQUES GÉNÉRALES DE LA RECHERCHE EN INTELLIGENCE ARTIFICIELLE ET ORGANISATION NATIONALE EN LA MATIÈRE</b> .....	70
<b>A. LES CARACTÉRISTIQUES DE LA RECHERCHE EN INTELLIGENCE ARTIFICIELLE</b> .....	70
1. <i>La place prépondérante de la recherche privée, dominée par les entreprises américaines et, potentiellement, chinoises</i> .....	70
2. <i>Une recherche essentiellement masculine</i> .....	74
3. <i>Une interdisciplinarité indispensable mais encore insuffisante</i> .....	75
4. <i>Une recherche soumise à une contrainte d'acceptabilité sociale assez forte sous l'effet de représentations catastrophistes de l'intelligence artificielle</i> .....	80
5. <i>Une recherche en intelligence artificielle qui s'accompagne de plus en plus d'interrogations et de démarches éthiques</i> .....	84
 <b>B. TABLEAU DE LA RECHERCHE FRANÇAISE EN INTELLIGENCE ARTIFICIELLE</b> .....	85
1. <i>De nombreux organismes publics intervenant dans la recherche en intelligence artificielle</i> .....	85
2. <i>Quelques exemples de centres, de laboratoires et de projets de recherche</i> .....	88
3. <i>Une reconnaissance internationale de la recherche française et qui s'accompagne d'un phénomène de rachat de start-ups et de fuite des cerveaux lié aux conditions attractives offertes à l'étranger</i> .....	89
4. <i>Une communauté française de l'intelligence artificielle encore insuffisamment organisée et visible</i> .....	90
5. <i>La sous-estimation des atouts considérables de la France et le risque de « décrochage » par rapport à la recherche internationale en intelligence artificielle</i> .....	92
 <b>DEUXIÈME PARTIE : LES ENJEUX DE L'INTELLIGENCE ARTIFICIELLE</b> .....	97
 <b>I. LES CONSÉQUENCES ÉCONOMIQUES ET SOCIALES DE L'INTELLIGENCE ARTIFICIELLE</b> .....	97
<b>A. D'IMPORTANTES TRANSFORMATIONS ÉCONOMIQUES EN COURS OU À VENIR</b> .....	97
1. <i>L'évolution vers une économie globalisée dominée par des « plateformes »</i> .....	97
2. <i>Un risque de redéfinition, sous l'effet de ce nouveau contexte économique, des rapports de force politiques à l'échelle mondiale</i> .....	101
3. <i>Des bouleversements annoncés dans le marché du travail : perspectives de créations, d'évolutions et de disparitions d'emplois</i> .....	102

<b>B. LA SOCIÉTÉ EN MUTATION SOUS L'EFFET DE L'INTELLIGENCE ARTIFICIELLE</b> .....	108
1. <i>Les défis lancés par l'intelligence artificielle aux politiques d'éducation et de formation continue</i> .....	108
2. <i>Une révolution potentielle de notre cadre de vie et de l'aide aux personnes</i> .....	109
3. <i>Le défi de la cohabitation progressive avec des systèmes d'intelligence artificielle dans la vie quotidienne</i> .....	111
<b>II. LES QUESTIONS ÉTHIQUES ET JURIDIQUES POSÉES PAR LES PROGRÈS EN INTELLIGENCE ARTIFICIELLE</b> .....	114
<b>A. LES ANALYSES PRÉSENTÉES PAR D'AUTRES INSTANCES POLITIQUES</b> .....	114
1. <i>Les deux rapports issus des institutions de l'Union européenne : Parlement européen et Comité économique et social européen (CESE)</i> .....	114
2. <i>Les trois rapports de la Maison Blanche</i> .....	115
3. <i>Le rapport de la Chambre des Communes du Royaume-Uni</i> .....	119
4. <i>Les initiatives chinoises et japonaises en intelligence artificielle accordent une place contrastée aux questions éthiques</i> .....	121
5. <i>La stratégie du Gouvernement pour l'intelligence artificielle : un plan qui arrive trop tard pour être intégré dans les stratégies nationales destinées au monde de la recherche</i> .....	126
<b>B. DES « LOIS D'ASIMOV » À LA QUESTION CONTEMPORAINE DE LA RÉGULATION DES SYSTÈMES D'INTELLIGENCE ARTIFICIELLE</b> .....	127
1. <i>Dépasser les « lois d'Asimov » pour envisager un droit de la robotique</i> .....	127
2. <i>Les questions juridiques en matière de conception (design), de propriété intellectuelle et de protection des données personnelles et de la vie privée</i> .....	129
3. <i>Les divers régimes de responsabilité envisageables et ceux envisagés</i> .....	138
4. <i>Les différenciations du droit applicable selon le type d'agents autonomes : robots industriels, robots de service, voitures autonomes et dilemmes éthiques afférents</i> .....	142
<b>C. LA PRISE EN COMPTE GRANDISSANTE DES ENJEUX ÉTHIQUES</b> .....	146
1. <i>Le cadre national de la réflexion sur les enjeux éthiques de l'intelligence artificielle</i> .....	146
2. <i>Les nombreuses expériences anglo-saxonnes de réflexion sur les enjeux éthiques de l'intelligence artificielle</i> .....	157
3. <i>Le travail en cours sur les enjeux éthiques au sein de l'association mondiale des ingénieurs électriciens et électroniciens (Institute of Electrical and Electronics Engineers ou IEEE)</i> .....	165
4. <i>Une sensibilisation insuffisante du grand public à ces questions et un besoin de partage en temps réel de la culture scientifique et de ses enjeux éthiques</i> .....	168
<b>III. LES QUESTIONS TECHNOLOGIQUES ET SCIENTIFIQUES QUI SE POSENT EN MATIÈRE D'INTELLIGENCE ARTIFICIELLE</b> .....	170
<b>A. LES SUJETS D'INTERROGATION LIÉS AUX ALGORITHMES UTILISÉS PAR LES TECHNOLOGIES D'INTELLIGENCE ARTIFICIELLE</b> .....	170
1. <i>Les questions de sécurité et de robustesse</i> .....	170
2. <i>Les biais et les problèmes posés par les données nécessaires aux algorithmes d'apprentissage automatique</i> .....	171
3. <i>Le phénomène de « boîtes noires » des algorithmes de deep learning appelle un effort de recherche fondamentale vers leur transparence</i> .....	172
4. <i>La question des bulles d'information dites « bulles de filtres »</i> .....	173
<b>B. LES SUJETS D'INTERROGATION LIÉS À LA « SINGULARITÉ », À LA « CONVERGENCE NBIC » ET AU « TRANSHUMANISME »</b> .....	174

1. La « singularité », point de passage de l'IA faible à l'IA forte peut, à long terme, constituer un risque .....	174
2. Un prophétisme dystopique indémontrable scientifiquement .....	176
3. Les questions posées par la « convergence NBIC » .....	180
4. La tentation du « transhumanisme » .....	181
<b>TROISIÈME PARTIE : LES PROPOSITIONS DE VOS RAPPORTEURS .....</b>	<b>183</b>
<b>I. POUR UNE INTELLIGENCE ARTIFICIELLE MAÎTRISÉE .....</b>	<b>183</b>
1. Proposition n° 1 : Se garder d'une contrainte juridique trop forte sur la recherche en intelligence artificielle, qui - en tout état de cause - gagnerait à être, autant que possible, européenne, voire internationale, plutôt que nationale .....	183
2. Proposition n° 2 : Favoriser des algorithmes et des robots sûrs, transparents et justes, et prévoir une charte de l'intelligence artificielle et de la robotique .....	184
3. Proposition n° 3 : Former à l'éthique de l'intelligence artificielle et de la robotique dans certains cursus spécialisés de l'enseignement supérieur .....	185
4. Proposition n° 4 : Confier à un institut national de l'éthique de l'intelligence artificielle et de la robotique un rôle d'animation du débat public sur les principes éthiques qui doivent encadrer ces technologies .....	185
5. Proposition n° 5 : Accompagner les transformations du marché du travail sous l'effet de l'intelligence artificielle et de la robotique en menant une politique de formation continue ambitieuse visant à s'adapter aux exigences de requalification et d'amélioration des compétences .....	186
<b>II. POUR UNE INTELLIGENCE ARTIFICIELLE UTILE, AU SERVICE DE L'HOMME ET DES VALEURS HUMANISTES .....</b>	<b>187</b>
1. Proposition n° 6 : Redonner une place essentielle à la recherche fondamentale et revaloriser la place de la recherche publique par rapport à la recherche privée tout en encourageant leur coopération .....	187
2. Proposition n° 7 : Encourager la constitution de champions européens en intelligence artificielle et en robotique .....	189
3. Proposition n° 8 : Orienter les investissements dans la recherche en intelligence artificielle vers l'utilité sociale des découvertes .....	189
4. Proposition n° 9 : Élargir l'offre de cursus et de modules de formation aux technologies d'intelligence artificielle dans l'enseignement supérieur et créer - en France - au moins un pôle d'excellence international et interdisciplinaire en intelligence artificielle et en robotique .....	189
5. Proposition n° 10 : Structurer et mobiliser la communauté française de la recherche en intelligence artificielle en organisant davantage de concours primés à dimension nationale, destinés à dynamiser la recherche en intelligence artificielle, par exemple autour du traitement de grandes bases de données nationales labellisées .....	190
6. Proposition n° 11 : Assurer une meilleure prise en compte de la diversité et de la place des femmes dans la recherche en intelligence artificielle .....	191
<b>III. POUR UNE INTELLIGENCE ARTIFICIELLE DÉMYSTIFIÉE .....</b>	<b>191</b>
1. Proposition n° 12 : Organiser des formations à l'informatique dans l'enseignement primaire et secondaire faisant une place à l'intelligence artificielle et à la robotique .....	191
2. Proposition n° 13 : Former et sensibiliser le grand public à l'intelligence artificielle par des campagnes de communication, l'organisation d'un salon international de l'intelligence artificielle et de la robotique et la diffusion d'émissions de télévision pédagogiques .....	192
3. Proposition n° 14 : Former et sensibiliser le grand public aux conséquences pratiques de l'intelligence artificielle et de la robotisation .....	194
4. Proposition n° 15 : Être vigilant sur les usages spectaculaires et alarmistes du concept d'intelligence artificielle et de représentation des robots .....	195



---

CONCLUSION .....	197
ANNEXES .....	199
ANNEXE 1 : SAISINE DE L'OFFICE.....	199
ANNEXE 2 : RÉUNION DE L'OPECST DU 14 MARS 2017 : ADOPTION DU RAPPORT .....	201
ANNEXE 3 : LISTE DES PERSONNES RENCONTRÉES .....	229
I. PERSONNES RENCONTRÉES PAR LES RAPPORTEURS EN VUE DE L'ÉTUDE DE FAISABILITÉ .....	229
II. PERSONNES RENCONTRÉES PAR LES RAPPORTEURS EN VUE DE L'ÉTABLISSEMENT DU RAPPORT .....	231
A. EN FRANCE .....	231
B. À L'ÉTRANGER.....	237
ANNEXE 4 : PROGRAMMES DES DÉPLACEMENTS (ÉTATS-UNIS, ROYAUME- UNI, BELGIQUE, SUISSE, ARCACHON POUR UN SÉMINAIRE SUR L'ÉTHIQUE DE L'INTELLIGENCE ARTIFICIELLE) .....	245
I. DÉPLACEMENT AUX ÉTATS-UNIS .....	246
II. DÉPLACEMENT AU ROYAUME-UNI.....	247
III. DÉPLACEMENT EN BELGIQUE.....	249
IV. DÉPLACEMENT EN SUISSE .....	250
V. DÉPLACEMENT À ARCACHON POUR UN SÉMINAIRE SUR L'ÉTHIQUE DE L'INTELLIGENCE ARTIFICIELLE .....	251
ANNEXE 5 : PANORAMAS DE L'INTELLIGENCE ARTIFICIELLE .....	252
I. AUX ÉTATS-UNIS.....	252
II. EN CHINE.....	264
III. AU JAPON.....	271
IV. AU ROYAUME-UNI.....	281
V. EN SUISSE .....	288
ANNEXE 6 : SUR LE DROIT COMPARÉ DE LA ROBOTIQUE .....	293
ANNEXE 7 : CONTRIBUTIONS .....	316

<b>I. CONTRIBUTION DE RAJA CHATILA : ROBOTIQUE ET INTELLIGENCE ARTIFICIELLE</b> .....	317
<b>II. CONTRIBUTION DE PATRICK ALBERT</b> .....	320
<b>ANNEXE 8 : RÉPONSES ÉCRITES AUX RAPPORTEURS</b> .....	321
<b>I. RÉPONSES D'AXELLE LEMAIRE, SECRÉTAIRE D'ÉTAT CHARGÉE DU NUMÉRIQUE ET DE L'INNOVATION ET DE THIERRY MANDON, SECRÉTAIRE D'ÉTAT CHARGÉ DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE, AU QUESTIONNAIRE DE VOS RAPPORTEURS</b> .....	321
<b>II. RÉPONSES À LA CONSULTATION ORGANISÉE PAR LE PARIS MACHINE LEARNING MEETUP EN VUE DU RAPPORT DE L'OPECST</b> .....	334
<b>ANNEXE 9 : COMPTE RENDU DE L'AUDITION PUBLIQUE DU 19 JANVIER 2017</b> ....	341
<b>I. INTRODUCTION</b> .....	341
1. <i>M. Bruno Sido, sénateur, premier vice-président de l'OPECST</i> .....	341
2. <i>M. Jean-Yves Le Déaut, député, président de l'OPECST</i> .....	342
<b>II. PREMIÈRE TABLE RONDE, PRÉSIDIÉE PAR M. CLAUDE DE GANAY, RAPPORTEUR : LES TECHNOLOGIES RELEVANT DE L'INTELLIGENCE ARTIFICIELLE</b> .....	345
1. <i>Mme Dominique Gillot, sénatrice, membre de l'OPECST, rapporteure</i> .....	345
2. <i>M. Claude de Ganay, député, membre de l'OPECST, rapporteur</i> .....	345
3. <i>M. Jean-Gabriel Ganascia, professeur à l'université Pierre-et-Marie-Curie, Paris-VI</i> .....	346
4. <i>M. Gérard Sabah, directeur de recherche honoraire au CNRS</i> .....	349
5. <i>M. Yves Demazeau, président de l'Association française pour l'intelligence artificielle</i> .....	353
6. <i>M. Bertrand Braunschweig, directeur du centre de l'INRIA de Saclay</i> .....	360
7. <i>M. David Sadek, directeur de la recherche à l'Institut Mines-Télécom</i> .....	363
8. <i>M. Jean-Daniel Kant, maître de conférences à l'université Pierre-et-Marie-Curie Paris-VI</i> .....	366
9. <i>M. Benoît Le Blanc, directeur adjoint de l'École nationale supérieure de cognitive</i> .....	368
10. <i>Débat</i> .....	370
<b>III. DEUXIÈME TABLE RONDE, PRÉSIDIÉE PAR MME DOMINIQUE GILLOT, RAPPORTEURE : DIMENSIONS STRATÉGIQUES EN MATIÈRE DE RECHERCHE EN INTELLIGENCE ARTIFICIELLE</b> .....	375
1. <i>M. Jean-Marc Merriaux, directeur général de Canopé</i> .....	375
2. <i>M. François Taddei, directeur du Centre de recherche interdisciplinaire (CRI)</i> .....	377
3. <i>M. Olivier Esper, responsable des affaires publiques de Google France</i> .....	379
4. <i>Mme Delphine Reyre, directrice Europe des affaires publiques de Facebook</i> .....	381
5. <i>M. Laurent Massoulié, directeur du Centre de recherche commun INRIA-Microsoft</i> .....	383
6. <i>M. Dominique Cardon, professeur de sociologie à l'Institut d'Études Politiques de Paris/Médialab</i> .....	384
7. <i>M. Gilles Babinet, entrepreneur, digital champion auprès de la Commission européenne</i> .....	385
8. <i>Débat</i> .....	388
<b>IV. TROISIÈME TABLE RONDE, PRÉSIDIÉE PAR MME DOMINIQUE GILLOT, RAPPORTEURE : QUESTIONS POLITIQUES, SOCIÉTALES ET ÉCONOMIQUES LIÉES À L'IRRUPATION DES TECHNOLOGIES DE L'INTELLIGENCE ARTIFICIELLE</b> .....	396
1. <i>M. Henri Verdier, directeur interministériel du numérique</i> .....	396

2. Mme Marie-Claire Carrère-Gée, présidente du Conseil d'orientation pour l'emploi .....	398
3. M. Laurent Alexandre, entrepreneur (DNA vision) .....	401
4. M. Jean-Christophe Baillie, entrepreneur (Novaquark) .....	406
5. M. Jean-Claude Heudin, directeur de l'Institut de l'internet et du multimédia .....	408
6. Débat.....	411
<b>V. INTERVENTION DE LA MINISTRE : AXELLE LEMAIRE, SECRÉTAIRE D'ÉTAT CHARGÉE DU NUMÉRIQUE ET DE L'INNOVATION .....</b>	<b>416</b>
1. Débat.....	419
<b>VI. QUATRIÈME TABLE RONDE, PRÉSIDIÉE PAR M. CLAUDE DE GANAY, RAPPORTEUR : ENJEUX ÉTHIQUES DE L'INTELLIGENCE ARTIFICIELLE .....</b>	<b>422</b>
1. Gilles Dowek, directeur de recherche à l'INRIA, professeur attaché à l'ENS Paris-Saclay.....	422
2. Laurence Devillers, professeur à l'Université Paris-Sorbonne/LMSI-CNRS .....	425
3. Serge Abiteboul, directeur de recherche à l'INRIA .....	427
4. M. Jean Ponce, professeur à l'ENS .....	429
5. Serge Tisseron, psychiatre, chercheur associé à l'Université Paris Diderot-Paris VII .....	430
<b>VII. CINQUIÈME TABLE RONDE, PRÉSIDIÉE PAR MME DOMNIQUE GILLOT, RAPPORTEURE : DÉFIS JURIDIQUES INHÉRENTS AUX USAGES DE L'INTELLIGENCE ARTIFICIELLE .....</b>	<b>433</b>
1. Intervention de Mady Delvaux, députée européenne (Luxembourg - groupe S&D), rapporteuse du groupe de travail sur la robotique et l'intelligence artificielle .....	433
2. Mme Isabelle Falque-Pierrotin, présidente de la CNIL .....	435
3. M. Rand Hindi, membre du Conseil national du numérique, pilote du groupe de travail sur l'intelligence artificielle, président de SNIPS .....	437
4. M. Olivier Guilhem, directeur juridique chez SoftBank Robotics (ex Aldebaran) .....	438
5. M <sup>e</sup> Alain Bensoussan, avocat, président de l'Association du droit des robots .....	439
6. Débat.....	440
<b>VIII. CONCLUSION.....</b>	<b>442</b>
<b>ANNEXE 10 : COMPTES RENDUS DES AUDITIONS BILATÉRALES CONDUITES PAR LES RAPPORTEURS .....</b>	<b>443</b>
<b>I. AUDITIONS DU 25 OCTOBRE 2016 .....</b>	<b>443</b>
1. Stéphane Mallat, professeur à l'École normale supérieure (ENS), chercheur en mathématiques appliquées.....	443
2. M. Patrick Albert, entrepreneur (créateur de ILOG), chercheur et pionnier dans le domaine de l'intelligence artificielle .....	445
<b>II. AUDITIONS DU 8 NOVEMBRE 2016 .....</b>	<b>446</b>
1. M. Marc Mézard, directeur de l'École normale supérieure (ENS).....	446
2. M. Raja Chatila, directeur de recherche au Centre national de la recherche scientifique (CNRS), directeur de l'Institut des systèmes intelligents et de robotique (ISIR) .....	448
<b>III. AUDITIONS DU 9 NOVEMBRE 2016.....</b>	<b>450</b>
1. M <sup>e</sup> Alain Bensoussan, avocat, président de l'association pour les droits des robots et M <sup>e</sup> Marie Soulez, avocate spécialisée sur les TIC dans son cabinet .....	450
2. M. Henri Verdier, directeur interministériel du numérique, ancien entrepreneur et spécialiste du numérique.....	453
3. M. Laurent Alexandre, président de DNA Vision, fondateur de Doctissimo, chirurgien- urologue .....	454

---

4. M. Pierre-Yves Oudeyer, directeur de recherche à l'INRIA, directeur du laboratoire Flowers, président du comité technique des systèmes cognitifs et développementaux de l'IEEE (Institut des ingénieurs électriciens et électroniciens) .....	456
<b>IV. AUDITIONS DU 24 NOVEMBRE 2016</b> .....	463
1. Mme Flora Fischer, chargée de programme de recherche au CIGREF, Club informatique des grandes entreprises françaises, pilote du groupe de travail sur l'intelligence artificielle en entreprise .....	463
2. M. Max Dauchet, professeur émérite à l'Université de Lille, président de la Commission de Réflexion sur l'Éthique de la Recherche en sciences et technologies du numérique (CERNA) d'Allistene, alliance des sciences et technologies du numérique .....	466
3. M. Cédric Sauviat, ingénieur, président de l'association française contre l'intelligence artificielle (AFCIA) et Mme Marie David, ingénieur, éditrice, membre du bureau de l'association .....	469
<b>V. AUDITIONS DU 28 NOVEMBRE 2016</b> .....	479
1. Claude Berrou, professeur à Télécom Bretagne (Institut MINES-TELECOM), chercheur en électronique et informatique, membre de l'Académie des sciences .....	479
2. M. Nicolas Cointe et Mme Fiona Berreby, chercheurs en thèse de doctorat sur l'éthique de l'intelligence artificielle .....	481
3. Mme Laurence Devillers, professeur d'informatique à l'université Paris-Sorbonne et directrice de recherche du CNRS au Laboratoire d'informatique pour la mécanique et les sciences de l'ingénieur (Limsi de Saclay) .....	484
<b>VI. AUDITIONS DU 30 NOVEMBRE 2016</b> .....	486
1. M. David Sadek, directeur de la recherche de l'Institut Mines-Télécom, spécialiste en intelligence artificielle .....	486
2. M. Dominique Sciamma, directeur de l'école de design « Strate » à Sèvres .....	488
3. M. François Taddéi, directeur du Centre de recherches interdisciplinaires (Inserm, université Paris-Descartes), biologiste .....	491
4. M. Igor Carron, entrepreneur, organisateur du principal « meet-up » en intelligence artificielle en France intitulé « Paris Machine Learning » .....	493
5. M. Jill-Jênn Vie, chercheur en thèse de doctorat à l'École normale supérieure Paris-Saclay .....	494

---

## SYNTHÈSE DU RAPPORT

### I. SYNTHÈSE : LES OBSERVATIONS DU RAPPORT

L'essor récent des technologies d'intelligence artificielle représente un bouleversement de nature à **transformer profondément nos sociétés et nos économies** mais reste soumis à une contrainte d'acceptabilité sociale assez forte sous l'effet de **représentations souvent catastrophistes**. Le concept d'intelligence artificielle renvoie à des **technologies multiples**, nées dans la seconde moitié du XX<sup>e</sup> siècle, qui reposent sur l'**utilisation d'algorithmes**. Ces technologies, dont les combinaisons sont en évolution constante, conduisent d'ores et déjà à des **applications dans de nombreux secteurs** et ouvrent un **espace d'opportunités inédit**, à même de révolutionner notre cadre de vie et l'aide aux personnes.

Les progrès en intelligence artificielle posent des questions auxquelles toute la société doit être sensibilisée : quels sont **les opportunités et les risques** qui se dessinent ? **La France et l'Europe** sont-elles dans une position satisfaisante dans la **course mondiale** qui s'est engagée ? Quelles places respectives pour la **recherche publique et la recherche privée** ? Quelle **coopération** entre celles-ci ? Quelles **priorités pour les investissements** dans la recherche en intelligence artificielle ? Quels **principes éthiques, juridiques et politiques** doivent encadrer ces technologies ? La **régulation doit-elle se placer au niveau national, européen ou international** ?

L'irruption de l'intelligence artificielle au cœur du débat public remonte à un peu plus de deux ans, après la diffusion d'une **lettre d'avertissement sur ses dangers potentiels**, publiée en janvier 2015, signée par 700 chercheurs et entrepreneurs, lancée pour alerter l'opinion publique et insister sur l'urgence de définir des règles éthiques. Il est frappant de constater qu'**aucun argument sérieux ne venait étayer cette première mise en garde quant au risque présumé de dérive malveillante**. Pourtant, même sans justification, ni preuve, cette alerte a contribué à **renforcer les peurs et les angoisses irrationnelles face au déploiement des technologies d'intelligence artificielle**. Tout au long de l'année 2016, les initiatives en matière d'intelligence artificielle se sont multipliées à un rythme effréné. Après l'irruption de l'intelligence artificielle dans le débat public en 2015, l'année 2016 et le premier trimestre 2017 ont en effet été jalonnés de **nombreux événements et rapports**. devant cet emballement, alors que les progrès se font à une vitesse exponentielle et reposent de plus en plus sur un financement privé aux moyens considérables, il est indispensable **que la réflexion soit conduite de manière sereine et rationnelle**, afin de **mettre en avant les opportunités tout autant que les risques de l'intelligence artificielle**, de **rassurer le public et de démystifier les représentations biaisées** de l'intelligence artificielle.

Ces représentations excessives, qui peuvent être totalement opposées, sont accentuées par la phase générale de progrès dans laquelle nous nous situons : en effet, la période récente s'apparente à un véritable « Printemps de l'intelligence artificielle ». Cette période polarise donc les opinions, qui peuvent être **des angoisses excessives mais aussi des espoirs démesurés** : les cycles d'espoirs et de déceptions qui jalonnent l'histoire de l'intelligence artificielle invitent à **ne pas trop s'enthousiasmer et à faire preuve d'attentes réalistes** à l'égard de ces technologies.

Les **applications sectorielles présentes ou futures sont d'envergure considérable** et les évolutions peuvent cependant être rapides, que l'on pense par exemple à **l'éducation, à l'environnement, à l'énergie, aux transports, à l'aéronautique, à l'agriculture, au commerce, à la finance, à la défense, à la sécurité, à la sécurité informatique, à la communication, aux loisirs, à la santé, à la dépendance ou, encore, au handicap.**

Le présent rapport fournit un **état de la recherche quant à ce que l'on appelle « l'intelligence artificielle »** et fait le **point sur de nombreux autres rapports parus récemment** sur le sujet en France et dans le monde. Il présente aussi les **enjeux éthiques, juridiques, économiques, sociaux et scientifiques de ces technologies**, parmi lesquels la place prépondérante de la recherche privée, dominée par les entreprises américaines et, potentiellement, chinoises, l'accélération du passage à une économie globalisée dominée par des « plateformes », les transformations du marché du travail, les régimes de responsabilité, les biais et les problèmes posés par les données et les algorithmes, le phénomène de « boîtes noires » des algorithmes et la question des « bulles d'information ». Il évoque, par ailleurs, certains sujets d'interrogation liés à la « singularité », à la « convergence NBIC » et au « transhumanisme » ainsi que la nécessité d'une **prise en compte grandissante de règles éthiques.**

Les **progrès** en intelligence artificielle sont d'abord et avant tout **bénéfiques**. Ils comportent aussi des **risques**, qu'il serait malhonnête de nier. Mais ces risques peuvent et doivent être **identifiés, anticipés et maîtrisés**. L'avènement d'une **super-intelligence** ne fait pas partie de ces risques à court et moyen terme. À **long terme, la réalité de cette menace n'est pas certaine**. Quant à **son imminence à court ou moyen terme**, prophétisée par plusieurs figures médiatiques, elle **relève du fantasme**. Le présent rapport se veut une première contribution à un travail indispensable d'identification, d'anticipation et de maîtrise des risques réels. Ce travail de démystification et d'objectivation doit être **collectif, interdisciplinaire et international**. Afin de prévenir de futures désillusions, il est nécessaire d'assurer un **suivi continu de ces technologies et de leurs usages.**

Ni quête vaine ni projet de remplacement de l'homme par la machine, **l'intelligence artificielle représente une chance à saisir pour nos sociétés et nos économies. La France doit relever ce défi.** Il convient donc d'aller au-delà des apparences et de regarder la réalité scientifique derrière les espoirs et les angoisses s'exprimant en raison du développement de l'intelligence artificielle. **Le débat public ne peut pas s'engager sereinement dans l'ignorance des technologies mises en œuvre, des méthodes scientifiques et des principes de l'intelligence artificielle.**

Le rapport se prononce **pour une intelligence artificielle maîtrisée, utile et démystifiée** : maîtrisée, parce que ces technologies devront être **les plus sûres, les plus transparentes et les plus justes possibles** ; utile parce qu'elles doivent, dans le respect des valeurs humanistes, **profiter à tous** au terme d'un large débat public ; démystifiée, enfin, parce que les difficultés d'acceptabilité sociale de l'intelligence artificielle résultent largement de **visions catastrophistes sans fondement**. Plutôt qu'une hypothétique confrontation dans le futur entre les hommes et les machines, qui relève d'une forme de science-fiction dystopique, les rapporteurs sont convaincus du **bel avenir de la complémentarité homme-machine**. Nous allons bien **plus vers une intelligence humaine augmentée que vers une intelligence artificielle concurrençant l'homme.**

## II. SYNTHÈSE : LES PROPOSITIONS DU RAPPORT

### I. Pour une intelligence artificielle maîtrisée

*Proposition n° 1 : Se garder d'une contrainte juridique trop forte sur la recherche en intelligence artificielle, qui - en tout état de cause - gagnerait à être, autant que possible, européenne, voire internationale, plutôt que nationale*

*Proposition n° 2 : Favoriser des algorithmes et des robots sûrs, transparents et justes et prévoir une charte de l'intelligence artificielle et de la robotique*

*Proposition n° 3 : Former à l'éthique de l'intelligence artificielle et de la robotique dans certains cursus spécialisés de l'enseignement supérieur*

*Proposition n° 4 : Confier à un institut national de l'éthique de l'intelligence artificielle et de la robotique un rôle d'animation du débat public sur les principes éthiques qui doivent encadrer ces technologies*

*Proposition n° 5 : Accompagner les transformations du marché du travail sous l'effet de l'intelligence artificielle et de la robotique en menant une politique de formation continue ambitieuse visant à s'adapter aux exigences de requalification et d'amélioration des compétences*

### II. Pour une intelligence artificielle utile, au service de l'homme et des valeurs humanistes

*Proposition n° 6 : Redonner une place essentielle à la recherche fondamentale et revaloriser la place de la recherche publique par rapport à la recherche privée tout en encourageant leur coopération*

*Proposition n° 7 : Encourager la constitution de champions européens en intelligence artificielle et en robotique*

*Proposition n° 8 : Orienter les investissements dans la recherche en intelligence artificielle vers l'utilité sociale des découvertes*

*Proposition n° 9 : Élargir l'offre de cursus et de modules de formation aux technologies d'intelligence artificielle dans l'enseignement supérieur et créer - en France - au moins un pôle d'excellence international et interdisciplinaire en intelligence artificielle et en robotique*

*Proposition n° 10 : Structurer et mobiliser la communauté française de la recherche en intelligence artificielle en organisant davantage de concours primés à dimension nationale, destinés à dynamiser la recherche en intelligence artificielle, par exemple autour du traitement de grandes bases de données nationales labellisées*

*Proposition n° 11 : Assurer une meilleure prise en compte de la diversité et de la place des femmes dans la recherche en intelligence artificielle*

### III. Pour une intelligence artificielle démystifiée

*Proposition n° 12 : Organiser des formations à l'informatique dans l'enseignement primaire et secondaire faisant une place à l'intelligence artificielle et à la robotique*

*Proposition n° 13 : Former et sensibiliser le grand public à l'intelligence artificielle par des campagnes de communication, l'organisation d'un salon international de l'intelligence artificielle et de la robotique et la diffusion d'émissions de télévision pédagogiques*

*Proposition n° 14 : Former et sensibiliser le grand public aux conséquences pratiques de l'intelligence artificielle et de la robotisation*

*Proposition n° 15 : Être vigilant sur les usages spectaculaires et alarmistes du concept d'intelligence artificielle et de représentations des robots*





---

## INTRODUCTION

Mesdames,

Messieurs,

**L'intelligence artificielle n'est pas un simple terrain de jeu**, même si la victoire du système AlphaGo au jeu de Go face au champion Lee Sedol en mars 2016, celle du système Watson au jeu télévisé Jeopardy en 2011 ou, auparavant du système Deep Blue aux échecs face à Garry Kasparov en 1997 pourrait le laisser penser.

Après la révolution qu'ont représentée Internet et les technologies de l'information et de la communication au cours des vingt dernières années, **un nouveau bouleversement pourrait transformer profondément nos sociétés et nos économies : l'essor, l'accélération exponentielle et la diffusion massive des technologies d'intelligence artificielle.**

Ces **opportunités**, qui pourront apporter dans notre futur des **progrès dans de nombreux domaines**, ne font pas l'objet d'une **analyse sereine et objective**, sans doute sous l'effet d'une **opinion publique souvent mal informée, voire désinformée** en raison de représentations catastrophistes issues de la science-fiction et d'analyses médiatiques alarmistes.

L'irruption de l'intelligence artificielle au cœur du débat public remonte à un peu plus de deux ans, après la diffusion d'une **lettre d'avertissement sur les dangers potentiels de l'intelligence artificielle**, publiée en janvier 2015 et signée par 700 personnalités, le plus souvent des scientifiques et des chefs d'entreprises, rejoints par plus de 5 000 signataires en un an. Elle a été lancée pour alerter l'opinion publique et insister sur l'urgence de définir des règles éthiques et une charte déontologique pour cadrer la recherche scientifique dans ce domaine, qu'elle soit publique ou privée.

Il est frappant de constater qu'**aucun argument sérieux ne venait étayer cette première mise en garde quant au risque présumé de dérive malveillante**. Pourtant, même sans justification, ni preuve, cette alerte a contribué à **renforcer les peurs et les angoisses irrationnelles face au déploiement des technologies d'intelligence artificielle.**

De plus, cette naissance du débat public sur le sujet de l'intelligence artificielle a été suivie d'une certaine confusion en raison de la publication, en juillet 2015, d'une **autre lettre signée par plus de mille personnalités demandant l'interdiction des robots tueurs**, à savoir les armes autonomes

---

qui sélectionnent et combattent des cibles sans intervention humaine, en arguant du fait que l'intelligence artificielle pourrait à terme être plus dangereuse que des ogives nucléaires. Cette lettre a été publiée lors de l'ouverture de la Conférence internationale sur l'intelligence artificielle qui s'est tenue à Buenos Aires en 2015 et à la suite de deux réunions d'experts qui s'étaient tenues à Genève sur les armes autonomes.

Après ces deux événements marquants, 2016 a ensuite fait figure d'année de l'intelligence artificielle, marquée par la chaire d'informatique du Collège de France confiée à Yann LeCun en février 2016<sup>1</sup> ou par l'évènement largement commenté du 15 mars 2016, lorsque **le système d'intelligence artificielle AlphaGo** créé par l'entreprise britannique DeepMind, rachetée en 2014 par Google, **a battu le champion de Go, Lee Sedol**, avec un score final de 4 à 1, marquant **l'histoire des progrès en intelligence artificielle et contredisant la thèse de ceux qui estimaient une telle victoire impossible**, tant le jeu de Go exige une subtilité et une complexité propres à l'intelligence humaine.

Deux semaines plus tôt, **l'Office parlementaire d'évaluation des choix scientifiques et technologiques (OPECST) a été saisi le 29 février 2016**, en application de l'article 6 *ter* de l'ordonnance n° 58-1100 du 17 novembre 1958 relative au fonctionnement des assemblées parlementaires, **par la commission des affaires économiques du Sénat, d'une étude sur l'intelligence artificielle.**

Tout au long de l'année 2016, parallèlement aux investigations conduites par vos rapporteurs, les initiatives en matière d'intelligence artificielle se sont multipliées à un rythme effréné. Après l'irruption de l'intelligence artificielle dans le débat public en 2015, l'année 2016 et le premier trimestre 2017 ont en effet été jalonnés de **nombreux événements et rapports**, dont il serait difficile de faire ici une liste exhaustive.

Pour mémoire, peuvent être mentionnés les rapports sur l'intelligence artificielle du Parlement européen, de la Maison Blanche (trois rapports), de la Chambre des Communes, de l'association mondiale des ingénieurs électriciens et électroniciens (Institute of Electrical and Electronics Engineers ou IEEE), de la commission de réflexion sur l'éthique de la recherche en sciences et technologies du numérique (CERNA) de l'alliance des sciences et technologies du numérique (Allistene) (deux rapports), de l'Institut national de recherche en informatique et en automatique (INRIA), de l'Institut Mines-Télécom, du Club informatique des grandes entreprises françaises (Cigref), du Syndicat des machines et technologies de production (SYMOP), de l'association française pour l'intelligence artificielle (AFIA), de

---

<sup>1</sup> Yann LeCun, directeur de la recherche en intelligence artificielle de Facebook, professeur d'informatique et de neurosciences à l'université de New York, a ainsi prononcé en tant que titulaire de la chaire annuelle « Technologies informatiques et sciences numériques » du Collège de France, le 4 février 2016 à 18h, au Collège de France, sa leçon inaugurale intitulée « Le deep learning, une révolution en intelligence artificielle ». Il y faisait valoir que « comme toute technologie puissante, l'intelligence artificielle peut être utilisée pour le bénéfice de l'humanité entière ou pour le bénéfice d'un petit nombre aux dépens du plus grand nombre ».

---

l'association française contre l'intelligence artificielle (AFCIA) etc. Des conférences d'envergure nationale ou internationale ont aussi été organisées sur le sujet par les Nations Unies, l'OCDE, la Fondation pour le futur de la vie, le MEDEF, l'AFIA, la Commission Supérieure du Numérique et des Postes (CSNP) entre autres. Enfin, l'initiative « France IA » lancée par le Gouvernement en janvier 2017 s'est accompagnée de l'annonce d'un plan national pour l'intelligence artificielle en mars 2017.

Devant cet emballement, alors que les progrès se font à une vitesse exponentielle et reposent de plus en plus sur un financement privé aux moyens considérables, il est indispensable **que la réflexion soit conduite de manière sereine et rationnelle**, afin de **mettre en avant les opportunités tout autant que les risques de l'intelligence artificielle**, de **rassurer le public et de démystifier les représentations biaisées**. Comme le disait Marie Curie, « *dans la vie, rien est à craindre, tout est à comprendre* ».

L'intelligence artificielle suscite en effet **enthousiasme, espoir et intérêt, aussi bien que méfiance, incrédulité ou oppositions**. À l'heure où les impacts de ces technologies deviennent de plus en plus significatifs, y compris dans la vie quotidienne de chacun de nous, et où les frontières entre l'homme et la machine semblent s'effacer peu à peu, **les choix scientifiques et technologiques à opérer doivent, plus que jamais, pouvoir l'être en connaissance de cause**.

Ces représentations excessives de l'intelligence artificielle, qui peuvent être totalement opposées, sont accentuées par la phase générale d'enthousiasme dans laquelle nous nous situons : en effet, selon les observations cycliques observées depuis un demi-siècle, la période récente s'apparente à un véritable « Printemps de l'intelligence artificielle ». Cette période polarise donc les opinions, qui peuvent être **des angoisses excessives mais aussi des espoirs démesurés** : les cycles d'espoirs et de déceptions qui jalonnent l'histoire de l'intelligence artificielle invitent à **ne pas trop s'enthousiasmer en faisant preuve d'attentes irréalistes** à l'égard des technologies existantes et de celles mises à disposition dans un avenir proche.

Il convient donc d'aller au-delà des apparences et de regarder la réalité scientifique derrière les espoirs et les angoisses s'exprimant en raison du développement de l'intelligence artificielle. **Le débat public ne peut pas s'engager sereinement dans l'ignorance des technologies mises en œuvre, des méthodes scientifiques et des principes de l'intelligence artificielle**. C'est pourquoi vos rapporteurs ont entendu mettre en lumière l'état à un instant donné de ces technologies en constante évolution.

Ils se posent la question de savoir **comment développer une culture de la responsabilité et une prise en compte des questions éthiques** au sein de la communauté des chercheurs en intelligence artificielle et en robotique<sup>1</sup>

---

<sup>1</sup> Cette question, analysée plus loin, a fait l'objet du premier rapport de la commission de réflexion sur l'éthique de la recherche en sciences et technologies du numérique (CERNA) de l'alliance des sciences et technologies du numérique (Allistene). Allistene regroupe en effet les organismes de

---

et au-delà, parce qu'ils n'oublient jamais que « *science sans conscience n'est que ruine de l'âme* », ainsi que l'affirmait Rabelais par la bouche de Gargantua s'adressant à Pantagruel.

Les progrès en intelligence artificielle posent des questions auxquelles toute la société doit être sensibilisée : quels sont **les opportunités et les risques** qui se dessinent ? **La France et l'Europe** sont-elles dans une position satisfaisante dans la **course mondiale** qui s'est engagée ? Quelles places respectives pour la **recherche publique et la recherche privée** ? Quelle **coopération** entre celles-ci ? Quelles **priorités pour les investissements** dans la recherche en intelligence artificielle ? Quels **principes éthiques, juridiques et politiques** doivent encadrer ces technologies ? La **régulation doit-elle se placer au niveau national, européen ou international** ? Face à ces interrogations, vos rapporteurs estiment qu'il est de la responsabilité des pouvoirs publics de proposer un point d'équilibre qui devra toujours être remis en débat à proportion des découvertes scientifiques, de leurs transferts et de leurs usages. Tel est l'objet même du présent rapport.

Afin de **prévenir les discours catastrophistes mais aussi ce risque de futures désillusions**, il est nécessaire d'**opérer un bilan objectif** de l'état de l'art en matière scientifique et technologique quant à ce que l'on appelle « l'intelligence artificielle ». C'est le rôle de l'OPECST et c'est la mission que se sont donnée vos rapporteurs : ils ont souhaité faire **l'état de la recherche et des usages des technologies d'intelligence artificielle**, en souligner les enjeux multiples et les riches perspectives, car ils sont **animés d'une préoccupation pédagogique de faciliter le partage des connaissances scientifiques et technologiques**.

---

## PREMIÈRE PARTIE : ÉLÉMENTS DE CONTEXTE

### I. LA DÉMARCHE DE VOS RAPPORTEURS

#### A. DE LA PROCÉDURE DE SAISINE À L'ADOPTION D'UN CALENDRIER DE TRAVAIL

##### 1. L'origine et l'instruction de la saisine

L'Office parlementaire d'évaluation des choix scientifiques et technologiques (OPECST) a été saisi le 29 février 2016, en application de l'article 6 *ter* de l'ordonnance n° 58-1100 du 17 novembre 1958 relative au fonctionnement des assemblées parlementaires, par la commission des affaires économiques du Sénat, d'une étude sur l'intelligence artificielle. Deux semaines plus tard, le 15 mars 2016, le système AlphaGo de Google DeepMind, battait le champion de Go, Lee Sedol. Cet événement a marqué l'histoire des progrès en intelligence artificielle car il contredisait la thèse de ceux qui estimaient une telle victoire impossible, tant le jeu de Go exige une subtilité et une complexité propres à l'intelligence humaine.

Vos rapporteurs ont pu constater la pertinence de ce sujet dans les mois suivants, compte tenu de la forte visibilité du thème de **l'intelligence artificielle dans l'actualité et la vie publique**. Votre rapporteure Dominique Gillot avait été sensibilisée aux enjeux de ce sujet par le **colloque annuel de la conférence des présidents d'université (CPU)** organisé en 2015, autour du thème « **Université 3.0** », ouvert par Bernard Stiegler, directeur de l'Institut de recherche et d'innovation (IRI) du centre Georges Pompidou et professeur à l'université de Londres. L'idée d'« Université 4.0 » avait été évoquée dans la conclusion de ce colloque.

Sans se fier au seul traitement médiatique, plus ou moins exact scientifiquement et plus ou moins sensationnaliste, votre rapporteure avait également été initiée aux problématiques de l'intelligence artificielle par différents travaux de l'OPECST, dont plusieurs tables rondes avaient déjà partiellement abordé le sujet<sup>1</sup>. Ces éléments ont interpellé sa **responsabilité de faire connaître et partager la culture scientifique, technique et industrielle (CSTI)**<sup>2</sup>, objectif que partage votre rapporteur Claude de Ganay.

Lors de sa réunion du 18 mai 2016, l'Office a désigné vos deux rapporteurs pour conduire l'étude. Dans un contexte de délais très

---

<sup>1</sup> Ces travaux sont récapitulés un peu plus loin.

<sup>2</sup> Votre rapporteure Dominique Gillot préside le Conseil national de la culture scientifique, technique et industrielle (CNCSTI). Placé auprès du ministre chargé de la Culture et du ministre chargé de la Recherche, ce conseil « participe à l'élaboration d'une politique nationale en matière de développement de la CSTI, en cohérence avec les grandes orientations de la stratégie nationale de recherche ».

---

contraints, **une étude de faisabilité a été réalisée par vos rapporteurs** qui a été présentée le 28 juin 2016. En effet, à la suite d'une saisine et en amont de la publication d'un rapport, l'OPECST procède toujours à une **étude de faisabilité**, conformément aux articles 19 et 20 de son règlement intérieur relatifs à la mission du rapporteur, qu'il convient de rappeler au stade du présent rapport :

« **Article 19 :**

*Le rapporteur procède d'abord à une étude de faisabilité, qui a pour objet :*

- *d'établir un état des connaissances sur le sujet,*
- *de déterminer d'éventuels axes de recherche et d'apprécier les possibilités d'obtenir des résultats pertinents dans les délais requis,*
- *de déterminer les moyens nécessaires pour engager valablement un programme d'études.*

*Pour cette étude de faisabilité, le rapporteur peut demander le concours des membres du conseil scientifique, avec l'accord du vice-président de cet organisme. »*

« **Article 20 :**

*Le rapporteur soumet à la délégation les conclusions de son étude de faisabilité.*

*Il propose :*

- *soit de ne pas poursuivre les travaux,*
- *soit de suggérer à l'auteur de la saisine une nouvelle formulation de celle-ci,*
- *soit d'engager un programme d'études conduisant à l'établissement d'un rapport. »*

L'OPECST ayant pour vocation d'**anticiper les questions complexes d'ordre scientifique et technologique qui pourraient se poser au législateur**, il doit pouvoir lui fournir des **explications circonstanciées sur des enjeux dont les risques et les opportunités auraient été difficiles à identifier sans son éclairage.**

C'est ainsi que l'Office, à la différence des commissions parlementaires permanentes, n'a pas pour mission de réagir en quelques semaines en réponse, par exemple, au dépôt d'un projet de loi dont l'examen vient, en général, rapidement devant le Parlement.

La mission de l'Office implique de **se situer suffisamment en amont du processus législatif pour sensibiliser les parlementaires à des aspects complexes de questions d'ordre scientifique ou technologique.**

En effet, créé par la loi n° 83-609 du 8 juillet 1983, à la suite d'un vote unanime du Parlement, l'OPECST a pour mission, aux termes de la loi,

---

« d’informer le Parlement des conséquences des choix de caractère scientifique et technologique afin, notamment, d’éclairer ses décisions ».

## 2. Un calendrier de travail enfermé dans un délai contraint de six mois

Suite à l’adoption de l’étude de faisabilité le 28 juin 2016, vos rapporteurs ont organisé des auditions et des déplacements, dont la liste figure en annexe du présent rapport.

Ces auditions et ces déplacements n’ont pu commencer qu’en **septembre 2016**, après les deux mois d’été, et dans la perspective des élections présidentielles, législatives et sénatoriales à venir en 2017, année électorale particulièrement chargée, vos rapporteurs, dont les mandats sont tous deux renouvelables, ont dû interrompre leurs investigations au mois de **février 2017**, ce qui ne leur a laissé qu’une **période utile d’environ six mois de travail** pour conduire leurs investigations. Ceci est inhabituellement bref pour un rapport de l’OPECST, qui d’ordinaire nécessite au moins une année de travail.

Il a donc été nécessaire pour vos rapporteurs de **tirer les conséquences de ce calendrier de travail enfermé dans un délai contraint de six mois en précisant le champ des investigations de l’étude. Un tel contexte rendait en effet difficile l’élaboration d’un rapport faisant un point complet sur l’ensemble des questions posées par l’intelligence artificielle.** Vos rapporteurs ne veulent donc aucunement prétendre épuiser le sujet et entendent faire de ce rapport un travail exploratoire sur un thème aux enjeux très diversifiés.

### B. LE CHAMP DES INVESTIGATIONS DE L’ÉTUDE

#### 1. L’étude de faisabilité du rapport conclut à la pertinence d’une étude spécifique de l’OPECST

Vos rapporteurs ont d’abord dressé en juin 2016 dans l’étude de faisabilité un **bilan des précédents travaux de l’Office** et des **autres travaux** sur l’intelligence artificielle conduits récemment **en-dehors de son cadre**.

Leur revue des rapports de l’Office parlementaire d’évaluation des choix scientifiques et technologiques visait à voir si une étude analogue avait été conduite et à déterminer ce qu’il était pertinent d’analyser de manière plus spécifique concernant l’intelligence artificielle.

L’OPECST a, dernièrement, rendu en 2016 un **rapport sur les robots et la loi**, fondé sur le compte rendu d’une audition publique qu’il a organisé le 10 décembre 2015<sup>1</sup>. Auparavant, et en allant du plus récent au plus ancien, l’Office a rendu plusieurs travaux sur le thème du numérique, démentant la

---

<sup>1</sup> Sénat, n° 570 (2015-2016).

---

thèse du livre de Laure Belot, « *La déconnexion des élites. Comment Internet déränge l'ordre établi ?* »<sup>1</sup>.

L'Office a ainsi travaillé sur le numérique au service de la santé<sup>2</sup>, sur les risques et la sécurité numérique<sup>3</sup>, sur le *big data* dans l'agriculture<sup>4</sup>, sur les drones et la sécurité des installations nucléaires<sup>5</sup>, sur les nouveaux moyens de transports<sup>6</sup>, sur les nouvelles technologies d'exploration et de thérapie du cerveau<sup>7</sup>, et, précédemment, sur la gouvernance mondiale de l'Internet<sup>8</sup>, sur les conséquences de l'évolution scientifique et technique dans le secteur des télécommunications<sup>9</sup>, sur les enjeux de société posés par le monde virtuel<sup>10</sup>, sur les techniques d'apprentissages en matière informatique<sup>11</sup>, sur l'entrée dans la société de l'information<sup>12</sup> et sur les conséquences de la révolution numérique<sup>13</sup>.

L'Office n'a cependant **jamais travaillé directement sur l'intelligence artificielle**, ce qui plaide pour la poursuite du travail de vos rapporteurs. Certains enjeux de l'intelligence artificielle, tels que l'impact sur la santé, la protection des données ou les moyens de transport, ont pu être entrevus à l'occasion de ces précédentes études de l'Office parlementaire d'évaluation des choix scientifiques et technologiques. Ils ont légitimement alimenté la curiosité de vos rapporteurs. Il semblait toutefois nécessaire d'approfondir cette connaissance au moyen d'une étude plus poussée.

La raison même de l'étude de faisabilité précédant les rapports de l'Office a ensuite conduit à se demander si une étude analogue concernant l'intelligence artificielle avait déjà été conduite en-dehors de l'OPECST ou était parue récemment.

Pour répondre à cette question, différentes publications nationales et européennes ont été passées en revue. Vos rapporteurs n'ont pas, au printemps 2016, identifié **de rapport public faisant le point sur les**

---

<sup>1</sup> Ce livre fait valoir l'argument d'une prétendue faible connexion des élites avec les enjeux de transformation massive de nos sociétés sous l'impact de la révolution numérique.

<sup>2</sup> Rapport Sénat n° 465 (2014-2015).

<sup>3</sup> Rapport « Sécurité numérique et risques : enjeux et chances pour les entreprises », Sénat n° 271 (2014-2015).

<sup>4</sup> Rapport « La place du traitement massif des données (big data) dans l'agriculture : situation et perspectives », Sénat n° 614 (2014-2015).

<sup>5</sup> Rapport Sénat n° 267 (2014-2015).

<sup>6</sup> Rapport « Les nouvelles mobilités sereines et durables », Sénat n° 293 (2013-2014).

<sup>7</sup> Sénat n° 476 (2011-2012).

<sup>8</sup> Sénat n° 219 (2005-2006).

<sup>9</sup> Sénat n° 159 (2001-2002).

<sup>10</sup> Rapport « Images de synthèse et monde virtuel : techniques et enjeux de société », Sénat n° 526 (1997-1998).

<sup>11</sup> Rapport « Les techniques des apprentissages essentiels pour une bonne insertion dans la société de l'information », Sénat n° 45 (1996-1997).

<sup>12</sup> Rapport « La France et la société de l'information : un cri d'alarme et une croisade nécessaire », Sénat n° 335 (1996-1997).

<sup>13</sup> Rapport « Les nouvelles techniques d'information et de communication : l'Homme cybernétique », Sénat n° 47 (1994-1995).



---

**technologies d'intelligence artificielle, ainsi que sur les opportunités et les risques qu'elles incarnent.**

Alors que ces opportunités et ces risques, en lien avec l'évolution des techniques, l'accélération des capacités des machines et les usages de l'intelligence artificielle devaient de toute évidence être identifiés et évalués, **les pouvoirs publics paraissaient ne pas avoir pris toute la mesure de l'enjeu, bien que cela fût souhaitable.**

Deux exceptions avaient été relevées : une initiative émanant d'une députée européenne et une association missionnée par la Commission européenne. En effet, au sein des institutions de l'Union européenne, Mme Mady Delvaux, présidente du groupe de travail du Parlement européen sur la robotique et l'intelligence artificielle, a rendu public, le 31 mai 2016, un **projet de rapport contenant des recommandations à la Commission européenne relatives à des règles de droit civil pour la robotique, assorti d'une motion portant résolution du Parlement européen.** Antérieurement, l'association **EuRobotics** (« *European Robotics Coordination Action* »), en charge du programme de recherche de l'Union européenne en robotique, qui a pour objectif de favoriser le développement de la robotique en Europe, a proposé le 31 décembre 2012 un projet de **livre vert sur les aspects juridiques de la robotique**<sup>1</sup>.

D'autres initiatives, moins en rapport avec l'intelligence artificielle, ont également été mentionnées dans l'étude préalable de vos rapporteurs. Au niveau national, le Parlement a débattu en 2016 du projet de loi pour une République numérique<sup>2</sup>. Également en 2016, la commission « Technologies de l'information et de la communication » (TIC) de l'**Académie des Technologies** a produit une **note interne** répondant à la question suivante : « l'accélération des nouvelles technologies numériques produit-elle des inquiétudes et une difficulté d'acceptation de la société ? ». En 2015, l'Académie des technologies a rendu une **communication sur le « big data »** et a poursuivi les activités de son groupe de travail « Vers une technologie de la conscience ? ». Initiés en 2014 et animés par Gérard Sabah et Philippe Coiffet, les travaux de ce groupe de travail n'ont pas débouché sur un rapport. En 2009, l'Académie des technologies avait, en revanche, sur décision de son conseil académique, publié une **brochure de dix questions sur l'intelligence artificielle et la technologie** posées à l'académicien Gérard Sabah. Il ne s'agissait pas pour autant formellement d'un rapport, d'un avis ou d'une communication de l'Académie des technologies, le thème de l'intelligence artificielle n'ayant pas fait l'objet de telles publications, habituelles au sein de l'Académie. Enfin, vos rapporteurs mentionnent une

---

<sup>1</sup> Cf. [http://www.eu-robotics.net/cms/upload/PDF/euRobotics\\_Deliverable\\_D.3.2.1\\_Annex\\_Suggestion\\_GreenPaper\\_ELS\\_IssuesInRobotics.pdf](http://www.eu-robotics.net/cms/upload/PDF/euRobotics_Deliverable_D.3.2.1_Annex_Suggestion_GreenPaper_ELS_IssuesInRobotics.pdf)

<sup>2</sup> Cf. le rapport Sénat n° 534 (2015-2016) de M. Christophe-André Frassa, ainsi que les avis n° 524, 525, 526 et 528 respectivement de M. Philippe Dallier, au nom de la commission des finances, de Mme Colette Mélot, au nom de la commission de la culture, de l'éducation et de la communication, de M. Patrick Chaize, au nom de la commission de l'aménagement du territoire et du développement durable, et de M. Bruno Sido, au nom de la commission des affaires économiques.

---

**lettre de veille du pôle de compétitivité *Cap Digital*** sur l'intelligence artificielle en 2015<sup>1</sup> et le fait que **l'Agence nationale de la recherche** (ANR) ait consacré un de ses « cahiers »<sup>2</sup> au thème de l'intelligence artificielle et de la robotique en 2012, intitulé « *Intelligence Artificielle et Robotique : Confluences de l'Homme et des STIC* »<sup>3</sup>.

De manière plus significative, en 2015, la **commission de réflexion sur l'éthique de la recherche en sciences et technologies du numérique** (CERNA) de l'alliance des sciences et technologies du numérique (Allistene), sous la présidence de Max Dauchet, a rendu son **premier rapport public** consacré à l'éthique de la recherche en robotique<sup>4</sup>.

La même année, pour mémoire, le **colloque annuel de la conférence des présidents d'université** (CPU) ouvert par Bernard Stiegler, directeur de l'Institut de recherche et d'innovation (IRI) du centre Georges Pompidou et professeur à l'université de Londres, a eu pour thème « **L'Université 3.0** », l'intelligence artificielle avait été évoquée et l'idée d'« Université 4.0 » avait conclu ce colloque. Le sujet de l'intelligence artificielle a été évoquée dans ce cadre, et l'idée d'« Université 4.0 » avait conclu ce colloque.

**Aucun de ces travaux ne semblant de nature à dispenser l'OPECST d'engager une réflexion plus approfondie sur le thème de l'intelligence artificielle** selon vos rapporteurs, ceux-ci ont proposé de poursuivre leur préparation d'un rapport de l'OPECST sur l'intelligence artificielle dans un cadre et selon des modalités déterminés et approuvés par leurs collègues. En effet, en l'état de leurs investigations, aucun rapport parlementaire ou autre ne semblait dresser le bilan des opportunités et des risques relatifs à l'intelligence artificielle et, partant, répondre à l'objet d'un nouveau rapport de l'OPECST sur ce sujet, quand bien même ce rapport se voudrait exploratoire. L'Office a donc retenu le principe d'investigations approfondies dans un délai de six mois, qui ont abouti au présent rapport.

## **2. Un ciblage délibéré sur un nombre limité de problématiques et de pistes d'investigation**

Vos rapporteurs **n'ont pas souhaité, en quelques mois, faire un point complet sur les questions posées par l'intelligence artificielle** et épuiser le sujet. Ils ont donc précisé le **champ d'investigations retenu**, permettant de répondre à la saisine transmise par la commission des affaires économiques du Sénat et de contribuer à **faire connaître et partager ce qu'est aujourd'hui, et ce que peut devenir demain l'intelligence artificielle**.

---

<sup>1</sup> Cf. [http://www.craft.ai/images/posts/craft-ai-interviewed-by-cap-digital/craft\\_ai\\_article\\_cap\\_digital.pdf](http://www.craft.ai/images/posts/craft-ai-interviewed-by-cap-digital/craft_ai_article_cap_digital.pdf)

<sup>2</sup> N° 4, mars 2012.

<sup>3</sup> Cf. [http://www.agence-nationale-recherche.fr/fileadmin/user\\_upload/documents/2012/Cahier-ANR-4-Intelligence-Artificielle.pdf](http://www.agence-nationale-recherche.fr/fileadmin/user_upload/documents/2012/Cahier-ANR-4-Intelligence-Artificielle.pdf)

<sup>4</sup> Cf. <https://hal.inria.fr/ALLISTENE-CERNA/hal-01086579v1>

---

Les enjeux de ce thème d'étude sont tout autant scientifiques et technologiques que politiques, philosophiques, éthiques, juridiques, éducatifs, médicaux, militaires ou, encore, économiques

Compte tenu des délais impartis, il leur a fallu retenir seulement **certaines problématiques du champ de l'intelligence artificielle pour la conduite de leurs investigations, en ayant le souci d'optimiser la plus-value relative du rapport**. Les différents thèmes pouvant être traités ont été circonscrits et parmi eux, neuf domaines d'investigation au moins peuvent être distingués :

1-La recherche publique en intelligence artificielle et les technologies informatiques ;

2-La recherche privée en intelligence artificielle (dont la question de la place des géants de l'Internet) ;

3-Les enjeux philosophiques et éthiques de l'intelligence artificielle ;

4-Les enjeux politiques et juridiques de l'intelligence artificielle ;

5-Les enjeux éducatifs de l'intelligence artificielle ;

6-Les enjeux économiques, industriels et financiers de l'intelligence artificielle (dont les systèmes et moyens de transports) ;

7-Les usages de l'intelligence artificielle en matière de technologies médicales ;

8-Les usages de l'intelligence artificielle pour la défense et les technologies militaires ;

9-Le projet transhumaniste d'homme augmenté.

**Les aspects scientifiques et technologiques constituant le cœur de métier de l'OPECST et sa plus-value spécifique par rapport aux autres commissions et délégations parlementaires, les points 1 et 2 ont été retenus.** La place considérable prise par la recherche privée pose la question des **enjeux de pouvoir et de sécurité par rapport à la recherche publique**. Elle touche même aux **problématiques de souveraineté et d'indépendance nationale**, d'autant plus que la colonisation numérique américaine est une réalité incontestable.

**Les points 3, 4 et 5 ont aussi été retenus, car ils soulèvent des questions essentielles.** Ils couvrent d'ailleurs déjà un champ très large d'investigation. Les **enjeux philosophiques, éthiques, politiques, juridiques et éducatifs**<sup>1</sup> de l'intelligence artificielle sont en effet **majeurs et**

---

<sup>1</sup> La note « Quelles priorités éducatives pour 2017-2027 ? » de France Stratégie, parue en mai 2016, s'interroge sur les conditions de réussite du virage du numérique. Elle a conduit à un débat public le 13 juin 2016. Elle se demande comment prendre appui sur le numérique pour améliorer les pédagogies, sur quels niveaux d'enseignement faire porter en priorité les investissements en matière de numérique, s'il faut privilégier le développement préalable de contenu pédagogique numérique ou bien l'équipement des élèves, des établissements ou des enseignants et, enfin, sur quelles collectivités

---

les identifier devrait permettre de **dépasser les peurs et les inquiétudes exprimées en vue d'engager un débat public plus serein** à ce sujet.

**Les enjeux financiers, économiques et industriels (point 6) n'ont pas été écartés, mais sont mis au second plan dans la mesure où ce domaine correspond moins directement**

**Le point 8 sur les usages de l'intelligence artificielle pour la défense et les technologies militaires a été écarté** pour deux raisons : d'une part, il s'agit d'un sujet qui relève assez largement d'une régulation internationale et, d'autre part, l'accès à l'information aurait été extrêmement difficile compte tenu de la sensibilité du sujet. La confidentialité et la discrétion étant de mise en matière de recherche militaire, vos rapporteurs n'ont pas voulu s'engager dans cette voie.

Enfin, **le point 9 sur le transhumanisme est très partiellement abordé, et plus sous l'aspect des enjeux éthiques de l'intelligence artificielle**. Il n'y avait pas lieu de traiter spécifiquement du transhumanisme, mouvement controversé qui relève davantage du projet idéologique que de la réalité scientifique et qui couvre un champ différent que celui de l'intelligence artificielle. Il inclut par exemple une partie des biotechnologies, ainsi que d'autres technologies émergentes.

Vos rapporteurs ont estimé que l'angle d'entrée le plus fécond en matière d'intelligence artificielle était de **mettre l'accent sur les enjeux éthiques, car ils permettent d'aborder de manière transversale la plupart des aspects retenus de manière prioritaire au paragraphe précédent** (soit les points 1, 2, 3, 4 et 5, à savoir la recherche publique et privée ainsi que les enjeux philosophiques, politiques, juridiques et éducatifs de l'intelligence artificielle). Sous couvert d'une **focalisation sur « l'éthique de l'intelligence artificielle »**, il a été possible de poursuivre des investigations dans chacun de ces sous-domaines, en mettant en évidence les opportunités et les risques que représente l'intelligence artificielle. **Cette réflexion sur les questions éthiques a été conduite au sens large** et s'est étendue de l'éthique de la recherche en intelligence artificielle jusqu'à l'éthique des robots intelligents, en passant par la revue de l'ensemble des démarches éthiques engagées en la matière. Elle présentait aussi l'intérêt de **s'abstenir de la tentation de définir un cadre juridique contraignant**, qui aurait eu pour inconvénient de figer des règles en codifiant des préceptes moraux et, partant, de ralentir l'innovation. Vos rapporteurs, au terme de leurs investigations, ne sont toujours **pas convaincus de l'urgence du besoin d'une intervention législative ou réglementaire en matière d'intelligence artificielle, secteur très évolutif**. Pour eux, une focalisation sur l'éthique de l'intelligence artificielle permet à la fois de répondre à des préoccupations de court terme mais aussi de plus long terme. Pour les premières, vos rapporteurs retiennent un propos de Laurence Devillers, professeure à l'Université

---

*faire reposer la charge de ces financements. Une note de la fondation Terra Nova du 10 mars 2016 intitulée « L'école sous algorithme » appelle à développer la culture numérique dans le monde éducatif et à un travail de l'OPECST sur les technologies éducatives.*

---

Paris IV Sorbonne et directrice de recherche au Laboratoire d'informatique pour la mécanique et les sciences de l'ingénieur (Limsi de Saclay) : « *L'utilisation de systèmes informatiques fonctionnant à partir d'apprentissage machine, met en lumière la nécessité d'une réflexion éthique sur les limites et performances des systèmes, surtout lorsqu'ils s'adaptent en continu. Ces systèmes amènent une rupture technologique et juridique par rapport aux algorithmes classiques paramétrables.* » Sur le long terme, vos rapporteurs ont entendu l'appel que Yann LeCun a lancé lors de sa leçon inaugurale au Collège de France : « *dans quelques décennies, quand nous pourrons peut-être penser à concevoir des machines réellement intelligentes, nous devons répondre à la question de comment aligner les valeurs des machines avec les valeurs morales humaines* ». Même s'il s'agit d'une question de long terme, elle mérite d'être **posée dès aujourd'hui** selon vos rapporteurs. Ces technologies doivent en effet être **maîtrisées, utiles et faire l'objet d'usages conformes à nos valeurs humanistes.**

### C. LA MÉTHODE DE TRAVAIL

#### 1. Une méthode de travail fondée sur des auditions bilatérales et des déplacements en France et à l'étranger

Ce point est développé en annexe du présent rapport, avec une présentation des auditions bilatérales et des déplacements de vos rapporteurs en France et à l'étranger (États-Unis, Royaume-Uni, Suisse, Belgique).

Il s'agissait pour mémoire de :

- un déplacement aux **États-Unis d'Amérique** du 22 au 29 janvier 2017, pour rencontrer des spécialistes de l'intelligence artificielle, à Washington à l'Institut de technologie du Massachusetts (MIT), à la *Harvard Kennedy School of Government*, à l'Université de Washington, à l'Université de Stanford, à l'Université de Berkeley, ainsi que des représentants de Facebook (Menlo Park, Palo Alto), de Google (Mountain View), d'Apple (Cupertino) et de Salesforce (San Francisco)... ;

- trois déplacements en **Europe** : à **Genève** du 21 au 22 septembre 2016 (HBP et BBP du *Brain Mind Institute* de l'École polytechnique fédérale de Lausanne, EPFL), au **Royaume-Uni** du 13 au 16 décembre 2016 (Chambre des Communes, Royal Society, *Alan Turing Institute*, *Future of humanity Institute* de l'Université d'Oxford, CSER et LCFI de Cambridge...) et à **Bruxelles** du 8 au 9 février 2017 (laboratoire d'intelligence artificielle de l'Université libre de Bruxelles, institutions européennes...)

- un déplacement en France métropolitaine, à **Arcachon** du 26 au 30 septembre 2016, pour participer à un séminaire sur l'éthique de l'intelligence artificielle organisé par la commission de réflexion sur l'éthique de la recherche en sciences et technologies du numérique d'Allistene (CERNA).

---

## 2. L'organisation d'une journée entière d'auditions publiques au Sénat le 19 janvier 2017

Une journée entière d'auditions publiques a été conduite au Sénat le 19 janvier 2017. L'organisation de cette journée a conduit une revue spécialisée à conclure que « *les pouvoirs publics s'emparent de la question de l'intelligence artificielle* »<sup>1</sup>

Les captations de ces vidéos sont disponibles<sup>2</sup>. Vos rapporteurs relèvent que ces auditions ont constitué un **record d'audience** pour le Sénat : sur Facebook, la vidéo a été regardée en direct par 97 182 personnes, puis en différé par 8 600 personnes. Sur le site Internet du Sénat, le nombre de consultations de la vidéo s'élève à 1 839. La chaîne Public Sénat en a tiré une émission spéciale, diffusée le lendemain de la journée d'auditions<sup>3</sup>.

## 3. La consultation d'ouvrages, de rapports et d'articles parus sur le sujet

Vos rapporteurs ont enrichi leurs connaissances par la **consultation d'ouvrages, de rapports et d'articles parus sur le sujet**. Il existe peu de « manuels » d'intelligence artificielle sur le marché. Un état de la connaissance, intitulé *Panorama de l'intelligence artificielle, ses bases méthodologiques, ses développements*<sup>4</sup>, a été dirigé en 2014 par Pierre Maquis, Odile Papini et Henri Prade. Il s'agit, avec ses trois volumes, de l'un des manuels plus complets en langue française, même si d'autres ouvrages plus anciens ont pu être consultés<sup>5</sup>.

En anglais, les 1152 pages de la troisième édition du **manuel de Stuart Russell et Peter Norvig** « *Artificial Intelligence: A Modern Approach* » restent inégalées<sup>6</sup>. Vos rapporteurs ont rencontré l'un de ses deux auteurs, Stuart Russell, professeur à Berkeley et très engagé dans le débat public sur l'intelligence artificielle. Une traduction française de cet

---

<sup>1</sup> [https://www.sciencesetavenir.fr/high-tech/intelligence-artificielle/les-pouvoirs-publics-s-emparent-de-la-question-de-l-intelligence-artificielle\\_109899](https://www.sciencesetavenir.fr/high-tech/intelligence-artificielle/les-pouvoirs-publics-s-emparent-de-la-question-de-l-intelligence-artificielle_109899)

<sup>2</sup> Pour les auditions du 19 janvier 2017 matin, le lien est le suivant : [http://videos.senat.fr/video.302142\\_588070cd0f9ea.reunion-pleniere-opect](http://videos.senat.fr/video.302142_588070cd0f9ea.reunion-pleniere-opect) et pour celles de l'après-midi du 19 janvier 2017 : [http://videos.senat.fr/video.303157\\_5880bb6877f24.reunion-pleniere-opect](http://videos.senat.fr/video.303157_5880bb6877f24.reunion-pleniere-opect)

<sup>3</sup> <https://www.publicsenat.fr/emission/les-matins-du-senat/table-ronde-intelligence-artificielle-51796>

<sup>4</sup> Préfacé par Paul Braffort, le livre est édité par Cépaduès. Le volume 1 traite de la représentation des connaissances et de la formalisation des raisonnements, le volume 2 des algorithmes pour l'intelligence artificielle et le volume 3 des frontières et des applications de l'intelligence artificielle.

<sup>5</sup> Le manuel classique, mais daté, est celui de Gérard Tisseau et Jacques Pitrat, « Intelligence artificielle : problèmes et méthodes », Presses universitaires de France, 1996. Hugues Bersini a fait paraître un ouvrage de vulgarisation chez Ellipse en 2006 : « De l'intelligence humaine à l'intelligence artificielle ».

<sup>6</sup> Stuart Russell et Peter Norvig, « Artificial Intelligence: A Modern Approach », Prentice Hall, 2010. Un autre manuel plus ancien peut être cité, celui de Michael R. Genesereth et Nils J. Nilsson, « Logical Foundations of Artificial Intelligence », Los Altos Californie Morgan Kaufmann, 1987.

---

ouvrage est disponible, avec le concours de Laurent Miclet et Fabrice Popineau.

Écrit par deux experts de renommée mondiale, ce livre est une **référence incontournable en matière d'intelligence artificielle dont il présente les principaux concepts** (logique, probabilités, mathématiques discrètes et continues, perception, raisonnement, apprentissage, prise de décision et action). Il analyse l'intelligence artificielle à travers le concept d'agents intelligents. Chaque chapitre est illustré par des exemples et des activités, allant d'exercices de réflexion à des exercices de programmation, en passant par l'approfondissement des méthodes décrites, soit plus de 500 activités au total. Les auteurs exposent comment un système réussit à percevoir son environnement de manière à analyser ce qui s'y passe, et comment il transforme la perception qu'il a de son environnement en actions concrètes. Voici une liste des sujets couverts dans leur ordre d'exposition : les contributions historiques des mathématiques, de la théorie des jeux, de l'économie, de la théorie des probabilités, de la psychologie, de la linguistique et des neurosciences ; les méthodes qui permettent de prendre des décisions lors de l'établissement d'un projet, en tenant compte des étapes à venir ; les différentes manières de représenter formellement les connaissances relatives au monde qui nous entoure ainsi que le raisonnement logique fondé sur ces connaissances ; les méthodes de raisonnement qui permettent d'établir des plans et donc de proposer des actions à entreprendre ; la prise de décisions en environnement incertain, avec les réseaux bayésiens et des algorithmes tels que l'élimination de variables et les « MCMC » (Markov Chain Monte-Carlo) ; les méthodes employées pour générer les connaissances exigées par les composants de prise de décision : les algorithmes de boosting, l'algorithme « EM » (expectation-minimization), l'apprentissage à base d'exemples et les méthodes à noyaux (SVM-machines à vecteurs de support) ; les implications philosophiques et éthiques de l'intelligence artificielle. Ces technologies seront décrites plus loin.

Outre le manuel cité, la **brochure de l'Académie des technologies** intitulée « Dix Questions sur l'intelligence artificielle et la technologie » posées à l'académicien Gérard Sabah a été consultée.

Des **ouvrages scientifiques et techniques** peuvent être mentionnés, dont vos rapporteurs ont souvent rencontré les auteurs dans le cadre de leurs auditions et déplacements : Jean-Gabriel Ganascia, *Le Mythe de la singularité, L'Intelligence artificielle, Les Sciences cognitives*, Hugues Bersini, *De l'intelligence humaine à l'intelligence artificielle, Les fondements de l'informatique, Brèves réflexions d'un informaticien obtus sur la société à venir*, Laurence Devillers, *Des Robots et des hommes*, Serge Abiteboul et Gilles Dowek, *Le temps des algorithmes*, Dominique Cardon, *A quoi rêvent les algorithmes : Nos vies à l'heure des big data*, Alain Bensoussan et Jérémy Bensoussan, *Droit des robots, Comparative Handbook : robotic technologies law*, Alain Bensoussan (et autres), *En compagnie des robots*, Alain Bensoussan, *Dictionnaire politique d'internet et du numérique, Code Informatique, fichiers et*



---

libertés, *Droit de l'informatique et de la télématique*, Jean-Yves Girard et Alan Turing, *La Machine de Turing*, Nick Bostrom *Super-intelligence*, Michael Jordan, *Learning in Graphical Models (Adaptive Computation and Machine Learning)*, Daphne Koller et Nir Friedman, *Probabilistic Graphical Models : Principles and Techniques*, Jean-Claude Heudin, *Le Deep learning*, *Immortalité numérique : Intelligence artificielle et transcendance*, les 3 Lois de la robotique : *Faut-il avoir peur des robots ?*, *Robots et avatars : Le rêve de Pygmalion*, *Les créatures artificielles : des automates aux mondes virtuels*, Jacques Ferber, *Les systèmes multi-agents : vers une intelligence collective*, Michael Wooldridge, *Introduction to Multi-Agent Systems*, Alex Pentland, *Social Physics*, Jean-Pierre Changeux, *L'Homme neuronal*, Jean-Michel Besnier, Francis Brunelle et Florence Gazeau, *Un cerveau très prometteur : Conversation autour des neurosciences*, ou encore un recueil de 150 avis d'experts par John Brockman intitulé *What to think about machines that think*. Bruce Buchanan a, quant à lui, rédigé un article très éclairant<sup>1</sup> qui revient sur l'histoire de l'intelligence artificielle.

Ils ont également pu consulter des **ouvrages axés sur les enjeux économiques** : Henri Verdier et Nicolas Colin, *L'âge de la multitude: Entreprendre et gouverner après la révolution numérique*, Gilles Babinet *Transformation digitale : l'avènement des plateformes, big data, penser l'Homme et le monde autrement et L'ère numérique, un nouvel âge de l'humanité*, Thibaut Bidet-Mayer, *L'industrie du Futur à travers le monde*, Dorothee Kohler et Jean-Daniel Weisz, *L'Industrie 4.0 : Les défis de la transformation numérique du modèle industriel allemand*.

Les **questions éducatives** ont aussi été explorées par la lecture de différents livres : Joël Boissière, Simon Fau, Francesco Pedró, *Le Numérique - Une chance pour l'école*, Bruno Devauchelle, *Comment le numérique transforme les lieux de savoirs*, *Le numérique au service du bien commun et de l'accès au savoir pour tous*, Philippe Meirieu Denis Kambouchner, Bernard Stiegler, *L'école, le numérique et la société qui vient*, Franck Amadiou, André Tricot *Apprendre avec le numérique*, ainsi que des notes « *Quelles priorités éducatives pour 2017-2027 ?* » de France Stratégie et « *L'école sous algorithme* » de la fondation Terra Nova.

**Quelques essais** ont, enfin, retenu leur attention : Serge Tisseron, *Le jour où mon robot m'aimera*, Paul Dumouchel et Luisa Damiano, *Vivre avec les robots*, *Essai sur l'empathie artificielle*, Marc Dugain et Christophe Labbé, *L'homme nu : la dictature invisible du numérique*, Michel de Pracontal, *L'Homme artificiel*, *Golems, robots, clones, cyborgs*, Pierre Bellanger, *La souveraineté numérique*, Eric Sadin, *La silicolonisation du monde : l'irrésistible expansion du libéralisme numérique*, *La vie algorithmique : critique de la raison numérique et L'humanité augmentée : L'administration numérique du monde*, Luc Ferry, *La Révolution transhumaniste* et Bernard Stiegler *Dans la disruption : Comment ne pas devenir fou ?*

---

<sup>1</sup> Bruce G. Buchanan « A (Very) Brief History of Artificial Intelligence », *AI Magazine*, 2005.



---

## II. L'HISTOIRE DES TECHNOLOGIES D'INTELLIGENCE ARTIFICIELLE ET DE LEURS USAGES

### A. DES TECHNOLOGIES NÉES AU MILIEU DU XX<sup>E</sup> SIÈCLE

#### 1. La préhistoire de l'intelligence artificielle et sa présence dans les œuvres de fiction

Les paragraphes suivants font le point sur les techniques d'intelligence artificielle et utilisent notamment, outre les ouvrages cités, la brochure de l'Académie des technologies de « dix questions sur l'intelligence artificielle et la technologie » posées à l'académicien Gérard Sabah.

De nombreuses incarnations d'intelligence artificielle ont jalonné notre histoire, qu'il s'agisse de mythes ou de projets imaginés par les écrivains et les scientifiques. Comme il a été vu, Jean-Claude Heudin et Michel de Pracontal leur ont consacré des ouvrages entiers<sup>1</sup>. Bruce Buchanan a quant à lui rédigé un article sur l'histoire de l'intelligence artificielle qui revient également sur l'ensemble de ses précurseurs<sup>2</sup>. Comme le relève Jean-Gabriel Ganascia, que vos rapporteurs ont pu rencontrer à plusieurs reprises, **Homère** a décrit, dans « *L'Illiade* », des servantes en or douées de raison : « *Fabriquées par Héphaïstos, le dieu forgeron, elles ont, selon le poète, voix et force ; elles vaquent aux occupations quotidiennes à la perfection, car les immortels leur ont appris à travailler. Ce sont donc des robots, au sens étymologique de travailleurs artificiels* ». **Ovide** dans ses « *Métamorphoses* » crée la figure de Galatée, statue d'ivoire sculptée par Pygmalion et à laquelle Vénus, déesse de l'amour, accepte de donner vie. Jean-Gabriel Ganascia rappelle également qu'il existait dès l'Égypte ancienne des statues articulées, animées par la vapeur et par le feu, qui hochaient la tête et bougeaient les bras. **La Bible**, par le Psaume 139:16, a fondé le mythe du Golem, cette créature d'argile humanoïde que l'on retrouve souvent dans la tradition cabalistique juive.

En 1495, en vue de festivités organisées à Milan, **Léonard de Vinci** imagine puis construit, bien que ce dernier point reste débattu, un « **chevalier mécanique** », sorte de robot automate revêtu d'une armure médiévale. Sa structure interne en bois, avec quelques parties en métal et en cuir, était actionnée par un système de poulies et de câbles.

Avec ses « animaux-machines », **René Descartes** proposa quant à lui, dans la première moitié du XVII<sup>e</sup> siècle, de reproduire artificiellement les fonctions biologiques, y compris la communication et la locomotion. **Blaise Pascal** réfléchit à la création d'une machine à calculer. À la fin du XVII<sup>e</sup> siècle, **Leibnitz** imagine ensuite une machine à calculer capable de raisonner.

---

<sup>1</sup> Jean-Claude Heudin, *Robots et avatars : Le rêve de Pygmalion et Les créatures artificielles : des automates aux mondes virtuels*, Michel de Pracontal, *L'Homme artificiel*, Golems, robots, clones, cyborgs.

<sup>2</sup> Bruce G. Buchanan « A (Very) Brief History of Artificial Intelligence », *AI Magazine*, 2005.

Pendant le siècle des Lumières, le philosophe français **Julien de la Mettrie** anticipe le jour où les progrès de la technique permettront de créer un homme-machine tout entier, à l'âme et au corps artificiels. L'abbé Mical et Kratzenstein imaginent une **machine à parler** en 1780, bientôt construite par le baron Von Kempelen grâce à une cornemuse à tuyaux multiples, aujourd'hui propriété du « *Deutsches Museum* » de Munich.

Dès 1818, **Mary Shelley** publie son roman « *Frankenstein ou le Prométhée moderne* », dans lequel elle imagine un savant capable de créer un être artificiel, le monstre Frankenstein.

Au milieu du XIX<sup>e</sup> siècle, **George Boole** appelle à mathématiser la logique, **Charles Babbage** conçoit l'ancêtre mécanique des ordinateurs d'aujourd'hui<sup>1</sup> et l'économiste britannique **William Stanley Jevons** imagine des pianos mécaniques, capables de raisonner.

**Jules Verne**, dans son roman, *La Maison à vapeur*, paru en 1880 imagine un éléphant géant à vapeur capable de traverser l'Inde, sur terre, comme sur l'eau. Sa machine n'est cependant pas autonome.

Alors qu'**Isaac Asimov** affirmait qu'« on peut définir la science-fiction comme la branche de la littérature qui se soucie des réponses de l'être humain aux progrès de la science et de la technologie », force est de constater que l'intelligence artificielle est un thème de science-fiction particulièrement fécond pour la **littérature**, le **cinéma** et les **jeux vidéo**.

Les **ouvrages** d'Isaac Asimov lui-même, mais aussi d'Arthur C. Clarke, de Philip K. Dick ou de William Gibson l'illustrent, ainsi que le font, **au cinéma, de 1927 à 2017**, les films « *Metropolis* », « *2001, l'Odyssée de l'espace* », « *Mondwest* », « *Les Rescapés du futur* », « *Le Cerveau d'acier* », « *Génération Proteus* », « *Blade runner* », « *Tron* », « *Terminator* », « *Matrix* », « *A.I.* », « *I, Robot* », « *Iron Man* », « *Wall-E* », « *Eva* », « *The Machine* », « *Transcendance* », « *Chappie* », « *Her* », « *Ex Machina* », ou, encore, cette année, « *Ghost in the Shell* ». Les thèmes de **l'hostilité de l'intelligence artificielle** ou des **risques que cette dernière ferait courir à l'espèce humaine** sont souvent au cœur de l'intrigue de ces œuvres.

Récemment, des **séries télévisées** à succès comme « *Person of interest* », « *Emma* », « *Westworld* » ou, surtout, « *Real Humans* » et « *Humans* » ont également exploité ce sujet.

---

<sup>1</sup> En 1834, pendant le développement d'une machine à calculer, Charles Babbage imagine le premier ordinateur sous la forme d'une « machine à différences » en utilisant la lecture séquentielle des cartes du métier à tisser Jacquard afin de donner des instructions et des données à sa machine. En cela, il fut le premier à énoncer le principe d'un ordinateur.

---

## 2. Les premières étapes de formation des technologies d'intelligence artificielle au XX<sup>e</sup> siècle, la notion d'algorithme et le débat sur la définition du concept d'intelligence artificielle

L'intelligence artificielle a fêté l'année dernière son **soixantième anniversaire**, puisqu'elle est inventée en tant que discipline et concept en 1956 dans un contexte que vos rapporteurs vont présenter dans les pages suivantes. Elle repose sur l'utilisation d'**algorithmes**, dont l'histoire est bien plus ancienne que celle de leurs usages en informatique.

Le mot **algorithme** est issu de la latinisation du nom du **mathématicien Al-Khawarizmi**, dont le titre d'un des ouvrages (« *Abrégé du calcul par la restauration et la comparaison* »), écrit en arabe entre 813 et 833, et dont la seule copie est conservée à l'Université d'Oxford, visité par vos rapporteurs, est également à l'origine du mot **algèbre**. Il est le premier à proposer des méthodes précises de résolution des équations du second degré, du type «  $ax^2 + bx + c = 0$  ».

La longue histoire des algorithmes est bien décrite par Serge Abiteboul et Gilles Dowek, dans leur ouvrage *Le temps des algorithmes*. Ils ont tous les deux été auditionnés par vos rapporteurs. Ils rappellent que les algorithmes sont **utilisés depuis des milliers d'années**, qu'**Euclide a inventé en l'an 300 avant notre ère** un algorithme de calcul du plus grand diviseur commun de deux nombres entiers et que la **complexité de certains algorithmes récents est telle qu'ils peuvent être comparés à des cathédrales**. Le domaine qui étudie les algorithmes est appelé l'algorithmique.

De manière résumée, un algorithme est un ensemble de séquences d'opérations. Il s'agit, de manière plus précise et rigoureuse, d'une **suite finie et non ambiguë d'opérations ou d'instructions permettant, à l'aide d'entrées, de résoudre un problème ou d'obtenir un résultat, ces sorties étant réalisées selon un certain rendement**<sup>1</sup>. De nombreuses applications sont possibles, à commencer par l'informatique, le fonctionnement des ordinateurs, en particulier leurs systèmes d'exploitation, reposant sur des algorithmes. Les algorithmes peuvent, en effet, servir, comme le rappellent Serge Abiteboul et Gilles Dowek, à **calculer** (comme il a été vu), mais aussi à **gérer des informations** (comme le font les logiciels d'archivage par exemple), à **analyser des données** (comme le font les moteurs de recherche),

---

<sup>1</sup> Donald Knuth, pionnier de l'algorithmique moderne (*The Art of Computer Programming*), a identifié les cinq propriétés suivantes comme étant les prérequis d'un algorithme : la finitude (« Un algorithme doit toujours se terminer après un nombre fini d'étapes »), une définition précise (« Chaque étape d'un algorithme doit être définie précisément, les actions à transposer doivent être spécifiées rigoureusement et sans ambiguïté pour chaque cas »), l'existence d'entrées (« des quantités lui sont données avant qu'un algorithme ne commence. Ces entrées sont prises dans un ensemble d'objets spécifiés ») et de sorties (« des quantités ayant une relation spécifiée avec les entrées ») et un rendement (« toutes les opérations que l'algorithme doit accomplir doivent être suffisamment basiques pour pouvoir être en principe réalisées dans une durée finie par un homme utilisant un papier et un crayon »).

à **communiquer** (comme le font les protocoles utilisés pour internet par exemple), à **traiter un signal** (comme le font les appareils photo et les microphones numériques par exemple), à **commander un robot** (comme le font les systèmes d'analyse des capteurs utilisés pour les voitures autonomes par exemple), à **fabriquer des biens** (comme le font les « usines 4.0 » supervisées par des algorithmes par exemple) ou, encore, à **modéliser et simuler** (comme le font certains outils de météorologie, de sismologie, d'océanographie ou, encore, de planétologie par exemple).

L'informatique constitue un **domaine d'application privilégié** pour les algorithmes. Mais son histoire ne se confond pas avec celle de ces derniers. Il en est de même pour l'histoire de l'intelligence artificielle, bien que ces **trois histoires** soient liées. En effet, comme il sera vu plus loin, l'informatique traite plutôt de **questions résolues par des algorithmes connus**, alors que l'on applique le label d'« intelligence artificielle » à des applications permettant de résoudre des problèmes moins évidents.

Dès 1936, **Alan Turing pose les fondements théoriques de l'informatique et introduit les concepts de programme et de programmation**. Il imagine en effet, à ce moment, un modèle abstrait du fonctionnement d'un appareil doté d'une capacité élargie de calcul et de mémoire, en recourant à l'image d'un ruban infini muni d'une tête de lecture/écriture, qui sera appelé « **machine de Turing** », précurseur de l'ordinateur moderne. Puis, dans un article paru en 1950<sup>1</sup>, il explore le problème de l'intelligence artificielle et propose une expérience maintenant connue sous le nom de « **test de Turing** », qui est une tentative de définition à travers une épreuve d'un standard permettant de qualifier une machine de « consciente »<sup>2</sup>. Il fait alors le pari que « *d'ici cinquante ans, il n'y aura plus moyen de distinguer les réponses données par un homme ou un ordinateur, et ce sur n'importe quel sujet* ». Cette prophétie d'Alan Turing quant aux progrès connus en l'an 2000 ne s'est toujours pas réalisée à ce jour.

De leur côté, les mathématiciens et neurologues **Warren McCulloch et Walter Pitts** écrivent dès 1943 un article intitulé « *Un calculateur logique des idées immanentes dans l'activité nerveuse* »<sup>3</sup> dans lequel il pose l'hypothèse que les neurones avec leurs deux états, activé ou non-activé, pourraient permettre la construction d'une machine capable de procéder à des calculs logiques. Ils publièrent dès la fin des années 1950 des **travaux plus aboutis sur les réseaux de neurones artificiels**. C'est en 1957 que Frank Rosenblatt développe le perceptron, première modélisation de réseau de neurones

<sup>1</sup> « Computing Machinery and Intelligence », *Mind*, octobre 1950.

<sup>2</sup> Le test de Turing consiste à mettre en confrontation verbale un humain avec un ordinateur imitant la conversation humaine et un autre humain. Dans le cas où l'homme qui engage les conversations n'est pas capable de dire lequel de ses interlocuteurs est un ordinateur, on peut considérer que le logiciel de l'ordinateur a passé avec succès le test.

<sup>3</sup> « A Logical Calculus of Ideas Immanent in Nervous Activity », 1943, *Bulletin of Mathematical Biophysics* 5.

---

artificiels<sup>1</sup>, à partir des travaux de McCulloch et Pitts. Ces derniers publient un **article plus important que les autres en 1959**<sup>2</sup> et constituent donc un modèle simplifié de neurone biologique, communément appelé **neurone formel**. Leurs travaux démontrèrent que des réseaux de neurones formels simples pouvaient théoriquement réaliser des fonctions logiques, arithmétiques et symboliques complexes. Leur fonctionnement sera expliqué plus loin.

Trois ans plus tôt, **en 1956, John McCarthy et Marvin Minsky ont organisé une école d'été à Dartmouth** qui est considérée comme **l'acte de naissance de l'intelligence artificielle**, à la fois en tant que discipline et en tant que concept d'*artificial intelligence*.

Le concept a fait l'objet d'un débat, et il est dit *a posteriori* que le choix du mot devrait beaucoup à la **quête de visibilité de ce nouveau champ de recherche**. Parler d'**intelligence artificielle a pu apparaître comme plus séduisant que de parler des « sciences et des technologies du traitement de l'information »**. L'anthropomorphisme essentialiste<sup>3</sup> qui est exprimé par le choix du concept d'« intelligence artificielle » n'a sans doute pas contribué, selon vos rapporteurs, à apaiser les peurs suscitées par le projet prométhéen de construction d'une **machine rivalisant avec l'intelligence humaine**.

Cette conférence, soutenue par la fondation Rockefeller, par Nathan Rochester, alors directeur scientifique d'IBM, et par Claude Shannon, ingénieur, mathématicien et père des théories de l'information et de la communication, offre en effet à John McCarthy l'occasion de **convaincre les participants d'accepter l'expression « intelligence artificielle »** en tant qu'intitulé de ce domaine de recherche. La conférence affirme que *« chaque aspect de l'apprentissage ou toute autre caractéristique de l'intelligence peut être si précisément décrit qu'une machine peut être conçue pour le simuler »*. La rigueur pousse à observer que le projet n'est pas, en réalité, de construire une machine rivalisant avec l'homme mais de simuler telle ou telle tâche que l'on réserve à l'intelligence humaine.

Outre John McCarthy et Marvin Minsky, **les participants**, tels que Ray Solomonoff, Oliver Selfridge, Trenchard More, Arthur Samuel, Allen Newell et Herbert Simon, ayant posé comme conjecture que tout aspect de l'intelligence humaine peut être décrit de façon assez précise pour qu'une machine le reproduise en le simulant, **discutent ensuite des possibilités de créer des programmes d'ordinateur qui se comportent intelligemment**,

---

<sup>1</sup> « The Perceptron : A Probabilistic Model for Information Storage and Organization in the Brain ».

<sup>2</sup> « What the frog's eye tells the frog's brain » ou « ce que l'œil d'une grenouille dit à son cerveau », coécrit avec Jerome Lettvin et Humberto Maturana, 1959, *Proceedings of the Institute of Radio Engineers*.

<sup>3</sup> L'anthropomorphisme est l'attribution de caractéristiques du comportement humain ou de la morphologie humaine à d'autres entités comme des dieux, des animaux, des objets ou d'autres phénomènes. L'essentialisme est l'attribution à un être ou à un objet d'une existence propre « par essence », c'est-à-dire inhérente au sujet en question.

---

c'est-à-dire qui résolvent des problèmes dont on ne connaît pas de solution algorithmique simple.

Dans les années suivantes, soutenus par l'agence américaine pour les projets de recherche avancée de défense du ministère de la Défense (*Defense Advanced Research Projects Agency* ou DARPA), mais aussi par IBM, **les chercheurs mettent au point de nouvelles techniques informatiques** : le langage Lisp en 1958, l'un des plus anciens langages de programmation<sup>1</sup>, le premier programme démontrant des théorèmes d'où est issue la notion d'heuristique (règle empirique utile permettant de réduire les chemins possibles mais sans aboutir nécessairement à une solution), une première idée des réseaux de neurones artificiels (le perceptron, dont Marvin Minsky souligne les limites théoriques), un programme qui joue aux dames et apprend par apprentissage à jouer de mieux en mieux... **Ces découvertes rendent les pères fondateurs de l'intelligence artificielle très optimistes**<sup>2</sup>.

De même, le premier agent conversationnel (ou « *bot* ») est créé en 1966 par Joseph Weizenbaum et simule un psychothérapeute grâce à sa technologie de reconnaissance de formes. Il s'appelle « Eliza » et suscite un grand enthousiasme. Mais ses capacités restent limitées, puisqu'il est incapable de vraiment répondre aux questions posées, se contentant de continuer à faire parler son interlocuteur, dans une logique de relance.

---

<sup>1</sup> Si l'on met de côté la « machine de Turing » qui relève de l'informatique théorique, le « système A-0 » (ou « A-0 System ») est le premier compilateur (programme qui transforme un code source écrit dans un langage de programmation ou langage source en un autre langage informatique, appelé langage cible) développé en 1952 ; il est suivi notamment par le Fortran (mot valise issu de l'anglais « formula translator ») inventé dès 1954, Lisp et Algol en 1958, COBOL (acronyme de « Common Business Oriented Language ») en 1959, BASIC (acronyme de « Beginner's All-purpose Symbolic Instruction Code ») en 1964, Logo en 1967, Pascal en 1971, ou, encore, Prolog (mot valise pour Programmation logique), inventé par des chercheurs français en 1972.

<sup>2</sup> En 1958, Herbert Simon et Allen Newell déclarent ainsi que « d'ici dix ans un ordinateur sera le champion du monde des échecs » et « d'ici dix ans, un ordinateur découvrira et résoudra un nouveau théorème mathématique majeur ». En 1965, Herbert Simon assure que « des machines seront capables, d'ici vingt ans, de faire tout travail que l'homme peut faire ». En 1967, Marvin Minsky estime que « dans une génération [...] le problème de la création d'une intelligence artificielle sera en grande partie résolu », et en 1970 que « dans trois à huit ans nous aurons une machine avec l'intelligence générale d'un être humain ordinaire ».

### Capture d'écran d'un exemple de conversation avec Eliza

```

Welcome to
          EEEEE LL      IIII ZZZZZZZ  AAAAA
          EE      LL      II      ZZ  AA  AA
          EEEEE LL      II      ZZZ  AAAAAA
          EE      LL      II      ZZ  AA  AA
          EEEEE LLLLLL IIII ZZZZZZZ  AA  AA

Eliza is a mock Rogerian psychotherapist.
The original program was described by Joseph Weizenbaum in 1966.
This implementation by Norbert Landsteiner 2005.

ELIZA: Is something troubling you ?
YOU:   Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU:   They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU:   Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU:   He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU:   It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:   █

```

Source : Norbert Landsteiner <https://fr.slideshare.net/nshir233/eliza-4615>

La représentation des connaissances, le langage objet, est au cœur de l'intelligence artificielle des années 1950 et 1960, et elle est ensuite mise au service de l'informatique, avec des résultats remarquables permettant les progrès connus vers les ordinateurs modernes. Ainsi que le remarque l'académicien des sciences Gérard Sabah, **l'informatique classique traite traditionnellement de questions résolues par des algorithmes connus**, alors que **l'intelligence artificielle s'intéresse plutôt aux problèmes pour lesquels aucun algorithme satisfaisant n'existe encore**.

Le paradoxe résultant de cette définition est le suivant : **dès que le problème a été résolu par une technologie dite d'intelligence artificielle, l'activité correspondante n'est plus considérée comme une preuve d'intelligence de la machine**. Les cas connus de résolutions de problèmes d'algèbre ou de capacité à jouer à des jeux (des jeux d'échecs par exemple) illustrent ce phénomène. Nick Bostrom explique ainsi que *« beaucoup d'intelligence artificielle de pointe a filtré dans des applications générales, sans y être officiellement rattachée car dès que quelque chose devient suffisamment utile et commun, on lui retire l'étiquette d'intelligence artificielle »*.

Les progrès en matière d'intelligence artificielle étant tangibles depuis les années 1950, **les frontières de l'intelligence artificielle sont donc sans cesse repoussées et ce qui était appelé intelligence artificielle hier n'est donc plus nécessairement considéré comme tel aujourd'hui**.



---

Vos rapporteurs observent que, dès l'origine, **l'intelligence artificielle est bien une étiquette**. Ce label recouvre en réalité des technologies diverses, dont ils ont voulu retracer la richesse et la diversité dans le présent rapport.

Vos rapporteurs ont, en effet, relevé dans leurs investigations que les outils d'intelligence artificielle sont très divers, ce qui traduit la variété des formes d'intelligence en général : elles vont de **formes explicites** (systèmes experts et raisonnements logiques et symboliques) à des **formes plus implicites** (réseaux de neurones et *deep learning*).

### 3. « L'âge d'or » des approches symboliques et des raisonnements logiques dans les années 1960 a été suivi d'un premier « hiver de l'intelligence artificielle » dans les années 1970

« L'âge d'or » des approches symboliques et des raisonnements logiques se produit dans les années 1960 après la naissance de l'intelligence artificielle à Dartmouth. **Recourant à des connaissances précises, telles que des logiques diverses ou des grammaires, ces formes d'intelligence sont dites explicites.**

Il existe, ensuite, les diverses modalités de **formalisme logique**, soit sous la forme de **logique classique**, de **logique floue**, de **logique modale** ou de **logique non monotone**.

**La logique mathématique peut représenter des connaissances<sup>1</sup> et modéliser des raisonnements.** Le principe de résolution permet d'automatiser ces raisonnements : pour démontrer une propriété, on montre que son contraire entraîne une contradiction avec ce qu'on sait déjà. La seule règle utilisée est celle du « détachement » ou *modus ponens*, figure du raisonnement logique concernant l'implication (exemple : « si p implique q et si p, alors q »). Cette méthode ne s'applique qu'à des cas simples, où la combinatoire n'est pas excessive. Fondé sur le même principe, le langage Prolog (acronyme de PROgrammation LOGique, qui permet de résoudre les problèmes par raisonnement à partir de règles de logique formelle) lève ces restrictions en permettant d'aborder des problèmes plus complexes.

**Des difficultés subsistent pour traiter des connaissances vagues ou incomplètes.** Devant ces limites, des extensions théoriques ont donné lieu à des logiques non classiques permettant d'exprimer plus d'éléments que dans la logique classique. Voulant étendre les possibilités de la logique classique, les logiques multivaluées gardent les mêmes concepts de base, hormis les valeurs de vérité, qui, selon les théories, varient de trois à un nombre infini de valeurs. La **théorie des logiques floues** étend ces logiques en considérant comme valeurs de vérité le sous-ensemble réel « [0,1] ». Elles permettent de

---

<sup>1</sup> Des symboles permettent alors de représenter des faits, et des règles permettent d'en déduire de nouveaux.



---

traiter des informations incertaines (Jean viendra peut-être demain) ou imprécises (Anne et Brigitte ont à peu près le même âge).

Les **logiques modales** introduisent des notions comme la possibilité, la nécessité, l'impossibilité ou la contingence qui modulent les formules de la logique classique. La **notion de vérité devient relative** à un instant donné ou à un individu. On distingue ainsi ce qui est accidentellement vrai (contingence : Strasbourg est en France) de ce qui ne peut pas être faux (nécessité : un quadrilatère a quatre côtés). Diverses interprétations des modalités donnent lieu à des applications distinctes, dont les plus importantes sont les logiques épistémiques (savoirs, croyances), déontiques (modélisant le droit) et temporelles (passé, présent, futur).

Nos connaissances n'étant pas universelles, nous pouvons être conduits à des suppositions fausses, remises en cause à la lumière d'expériences ultérieures. Les **logiques non monotones** tiennent compte du fait que les exceptions sont exceptionnelles et formalisent les raisonnements où l'on adopte des hypothèses (tous les oiseaux volent) qui pourront être modifiées par des connaissances plus précises (mais pas les autruches). On raisonne avec des règles du type : si a est vrai et si b n'est pas incohérent avec ce qu'on sait, on peut déduire c (si Titi est un oiseau et si je ne sais pas que c'est une autruche, il vole). **On autorise ainsi la prise de décision malgré une information incomplète** : des suppositions plausibles permettent certaines déductions ; si, à la lumière d'informations ultérieures, ces suppositions se révèlent fausses, on remettra en question les déductions précédentes (non-monotonie).

S'agissant des **grammaires**, le traitement automatique des langues est un des grands domaines de l'intelligence artificielle, qui vise l'application de ses techniques aux langues humaines. Très pluridisciplinaire, il collabore avec la linguistique, la logique, la psychologie et l'anthropologie. Les travaux en traitement automatique des langues ont donné lieu à la constitution de divers ensembles de données numériques (dictionnaires de langue, de traduction, de noms propres, de conjugaison, de synonymes ; grammaires sous diverses formes ; données sémantiques), ainsi qu'à divers logiciels (analyseurs et générateurs morphologiques ou syntaxiques, gestionnaires de dialogue...). Du point de vue conceptuel, ces travaux ont produit des théories grammaticales plus compatibles avec les questions d'informatisation, des théories formelles pour la représentation du sens des mots, des phrases, des textes et des dialogues, ainsi que des techniques informatiques spécifiques pour le traitement de ces éléments sur ordinateur.

John McCarthy a inventé le langage de programmation « LISP » dès 1958<sup>1</sup>, c'est un mot valise formé à partir de l'anglais *list processing* ou traitement de listes. De grands espoirs sont alors placés dans la compréhension du langage naturel, dans la vision artificielle, mais au final

---

<sup>1</sup> « Fonctions Récursives d'expressions symboliques et leur évaluation par une Machine » ou « Recursive Functions of Symbolic Expressions and Their Computation by Machine », *Communications of the ACM*, Avril 1960.

les résultats sont décevants, largement en raison des limitations de puissance du matériel disponible mais aussi des limites des technologies. Ainsi le perceptron, dans lequel Frank Rosenblatt plaçait tant d'espérance est rapidement critiqué. Le livre *Perceptrons* de Marvin Minsky et Seymour Papert, paru en 1969, démontre les limites des réseaux de neurones artificiels de l'époque<sup>1</sup>.

Après cet **âge d'or, qui court de 1956 au début des années 1970**, les **financements sont revus à la baisse**, suite à **différents rapports assez critiques** : les prédictions exagérément optimistes des débuts ne se réalisent pas et les techniques ne fonctionnent que dans des cas simples. À l'évidence, les difficultés fondamentales de l'intelligence artificielle furent alors largement sous-estimées en particulier la question de savoir comment donner des connaissances de sens commun à une machine. Les recherches se recentrent alors sur la programmation logique, les formalismes de représentation des connaissances et sur les processus qui les utilisent au mieux.

En dépit de cette réorientation, qui témoigne d'**une certaine cyclicité des investissements en intelligence artificielle selon une boucle « espoirs-déceptions »**, Marvin Minsky et ses équipes du MIT (Massachusetts Institute of Technology) développent divers systèmes (Sir, Baseball, Student..) qui relancent les recherches sur la compréhension automatique des langues.

#### **4. Un enthousiasme renouvelé dans les années 1980 autour des systèmes experts, de leurs usages et de l'ingénierie des connaissances précède un second « hiver de l'intelligence artificielle » dans les années 1990**

Au cours des années 1980, **de nouveaux financements publics** sont ouverts avec le projet japonais dit de « cinquième génération », le programme britannique Alvey, le programme européen Esprit et le soutien renouvelé de la DARPA aux États-Unis. Les approches sémantiques, en lien avec les sciences cognitives, sont alors en plein essor, en lien avec la représentation des connaissances mais surtout avec les systèmes experts et l'ingénierie des connaissances. Leurs usages dans le monde économique sont des signes de cette vitalité.

Il s'agit tout d'abord des **systèmes experts**, appelés aussi systèmes à base de connaissances. Un système expert est un logiciel qui vise à reproduire les raisonnements d'un expert, dans un domaine particulier. La connaissance est décrite sous la forme générale de règles :

---

<sup>1</sup> La critique principale concerne l'incapacité du perceptron à résoudre les problèmes non linéairement séparables, tels que le problème du « X OR » (« OU exclusif »). Il s'en suivra alors, face à la déception, une période noire d'une vingtaine d'années pour les réseaux de neurones artificiels.

---

« SI *Condition* (s) » → « ALORS *Action* (s) »

Ces systèmes analysent une représentation de la situation pour voir quelles règles sont pertinentes, résolvent les éventuels conflits si plusieurs règles s'appliquent et exécutent les actions indiquées en modifiant la situation en conséquence. Ces systèmes sont efficaces dans des domaines restreints, mais deviennent difficiles à gérer quand ils doivent manipuler beaucoup de règles ou dans des domaines ouverts.

Destiné au diagnostic des maladies infectieuses du sang sur la base d'un ensemble de règles déclaratives (si tels faits - alors effectuer telles actions), le **premier système expert dit « MYCIN » est créé en 1974 et se diffuse dans les années 1980**. Il s'agit alors d'extraire des connaissances à partir du savoir des experts humains.

Les succès de cette approche restent relatifs, car elle ne fonctionne bien que dans des domaines restreints et spécialisés. L'incapacité de l'étendre à des problèmes plus vastes renforce alors le désintérêt pour l'intelligence artificielle.

Après ce court regain d'intérêt, la recherche subit à nouveau un déclin des investissements. L'enthousiasme renouvelé dans les années 1980 autour des systèmes experts, de leurs usages et de l'ingénierie des connaissances précède donc un **second « hiver de l'intelligence artificielle » dans les années 1990**.

Pour autant, des découvertes scientifiques sont faites dans la période. Après la renaissance de l'intérêt pour les **réseaux de neurones artificiels** avec de nouveaux modèles théoriques de calculs, les années 1990 voient se développer la **programmation génétique** ainsi que les **systèmes multi-agents** ou **l'intelligence artificielle distribuée**. La nécessité de méta-connaissances<sup>1</sup> émerge également.

Les usages des systèmes experts et de l'ingénierie des connaissances persistent jusqu'à aujourd'hui ainsi que l'a expliqué Alain Berger, directeur général d'Ardans, dans son intervention lors de la journée « Entreprises françaises et intelligence artificielle » organisée par le MEDEF et l'AFIA le 23 janvier 2017. Il a ainsi rappelé qu'il reste **essentiel de faire coopérer et inter-opérer les connaissances et les données** ; en cela, le développement d'outils précis revêt une importance capitale pour faire parvenir cette intelligence vers l'utilisateur, l'humain demeurant par son expertise la clé de validation de la connaissance. Depuis 1956, de nombreux progrès ont été accomplis, à l'instar du développement des systèmes experts, de la production et du recueil de volumes importants de données mais également de solutions coopératives. Au fil des siècles, le terme « connaissance » a évolué ; cependant, l'attachement à la compréhension d'une vérité et à sa construction a demeuré. Ce terme pourrait aujourd'hui être défini comme le

---

<sup>1</sup> Il s'agit de connaissances à propos des connaissances elles-mêmes.

---

fait de comprendre, de connaître les propriétés, les caractéristiques et les traits spécifiques d'une chose. Selon Alain Berger, l'ingénierie de la connaissance s'articule donc autour d'une approche de type cognitiviste, qui postule que la pensée est un processus de traitement de l'information. **Le cognitivisme dans l'ingénierie de la connaissance consiste ainsi à coupler représentation et computation.** La connaissance, d'un point de vue technique, vise à être structurée efficacement pour l'expert comme pour l'utilisateur ; d'un point de vue stratégique, il est essentiel de rendre explicites les savoirs tacites, de capitaliser les expériences singulières et de capitaliser les connaissances pour les préserver, les exploiter, les enrichir et les amplifier. En ce sens, l'approche cognitiviste consiste en trois clés : **une structuration d'une intelligence humaine, une justification du contenu par la validation d'un expert, et l'interopérabilité avec d'autres systèmes.** L'ingénierie de la connaissance fait donc figure, pour Alain Berger, de véritable tremplin de l'innovation, c'est **une compétence clé pour l'organisation, en ce qu'elle permet la maîtrise de ses savoirs et la performance de ses systèmes.** Il peut exister, dans le cadre de sujets exploratoires un besoin de modéliser des phénomènes, des interactions, des acteurs qui permettront de construire des scénarii et de construire de nouvelles connaissances. L'ingénierie de la connaissance a pour points forts :

- la formation des acteurs ;
- l'amélioration des compétences des acteurs d'un service ;
- la résolution de problèmes ;
- la pérennisation de l'expertise, qui est liée à un homme, à une technologie ou à un projet, la pérennisation des connaissances d'une technologie ou d'un projet.

### **5. Les autres domaines et technologies d'intelligence artificielle : robotique, systèmes multi-agents, machines à vecteur de support (SVM), réseaux bayésiens, apprentissage machine dont apprentissage par renforcement, programmation par contraintes, raisonnements à partir de cas, ontologies, logiques de description, algorithmes génétiques...**

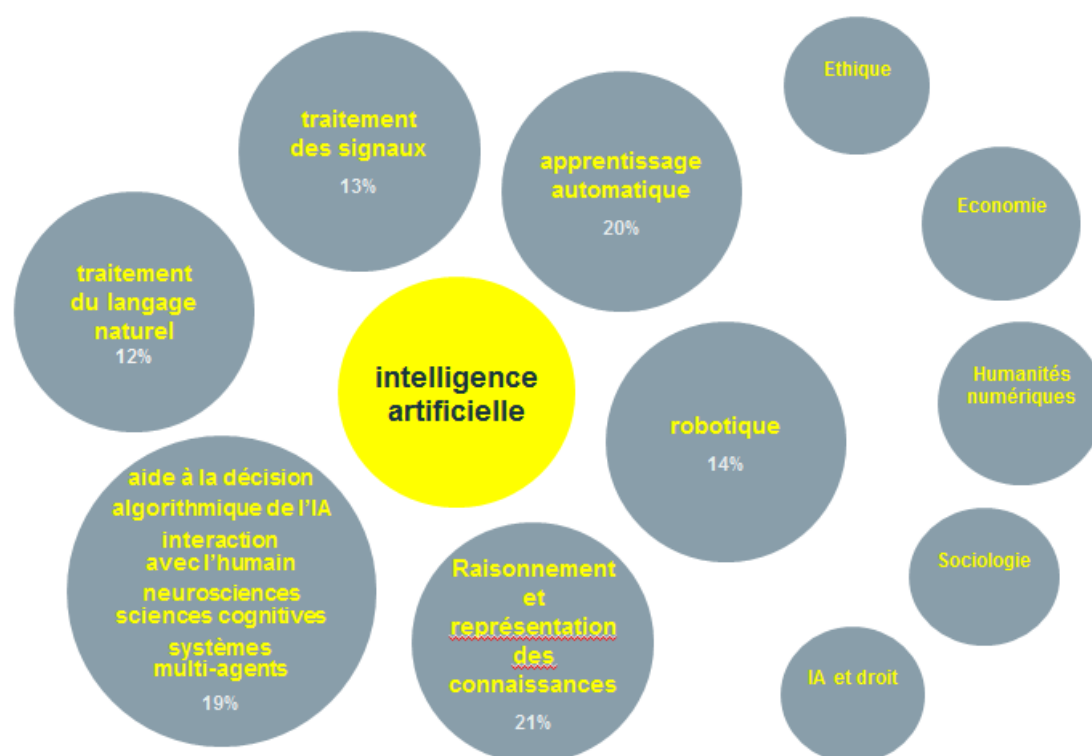
De très nombreux autres domaines et technologies d'intelligence artificielle peuvent être ajoutés à ceux déjà mentionnés précédemment. Certains vont être abordés au présent chapitre sans que cette liste ne soit en rien exhaustive : il ne s'agit que de **quelques exemples visant à illustrer la variété et la richesse qui se cache derrière le label d'intelligence artificielle.**

Le tableau académique des domaines de l'intelligence artificielle (IJCAI) retient **cinq domaines** : traitement du langage naturel, vision (ou

traitement du signal), apprentissage automatique, systèmes multi-agents, robotique.

Mais les technologies d'intelligence artificielle sont quasi-innombrables, surtout que les chercheurs, tels des artisans, hybrident des solutions inédites au cas par cas, en fonction d'un tour de main souvent très personnel. Il s'agit d'une caractéristique propre à la recherche en intelligence artificielle, souvent peu connue à l'extérieur du cercle des spécialistes et à laquelle ont été sensibilisés vos rapporteurs.

### Les domaines de l'intelligence artificielle



Source : Gouvernement (les pourcentages indiquent la répartition estimée des chercheurs français entre les différents domaines de l'intelligence artificielle)

Le premier domaine qui peut être remarqué est celui de la **robotique**, qui a toujours entretenu des liens très étroits avec celui de l'intelligence artificielle. Pour Raja Chatila, directeur de l'Institut des systèmes intelligents et de robotique (ISIR), la problématique de l'intelligence artificielle telle que posée par Alan Turing<sup>1</sup> était de savoir si les ordinateurs pouvaient être capables de « pensée » (*Can Machines Think?*) et il l'a traduite par la question de l'imitation de l'homme. Or les fondateurs de « l'intelligence artificielle » dans les années 1950 ont orienté la problématique vers celle de l'« intelligence », ou de « mécanismes de haut niveau ».

<sup>1</sup> A. M. Turing (1950) Computing Machinery and Intelligence. *Mind* 49: 433-460.

---

Cette manière de poser la question néglige un constat pourtant simple : le cerveau humain a évolué vers ce qu'il est en développant des capacités de perception, d'interprétation, d'apprentissage et de communication en vue d'une action plus efficace. Or la problématique de la robotique pose l'ensemble de ces questions. Le robot-machine est soumis à la complexité du monde réel dans lequel il évolue et dont il doit respecter sa dynamique. La notion d'intelligence doit alors être posée de manière à rendre compte globalement des processus sensori-moteurs, perceptuels et décisionnels permettant l'interaction en temps réel avec le monde en tenant compte des contraintes d'incomplétude et d'incertitude de perception ou d'action. C'est le sens de la définition de la robotique donnée par Mike Brady (Oxford) dans les années 1980 : « *La robotique est le lien intelligent entre la perception et l'action.* » Dans ce sens on peut dire que le robot est le paradigme de l'intelligence artificielle « encorporee », c'est à dire une intelligence matérialisée dans un environnement qu'elle découvre et dans lequel elle agit.

Il est nécessaire, selon Raja Chatila, d'adopter une vision d'ensemble du robot, en tant que **système** intégrant ses différentes capacités (perception/interprétation, mouvement/action, raisonnement/planification, apprentissage, interaction) et permettant à la fois la réactivité et la prise de décision sur le long terme. Ces fonctions doivent être intégrées de manière cohérente dans une architecture de contrôle globale (architecture cognitive) ; leur étude de manière séparée risque d'aboutir à des solutions inappropriées.

De nombreuses avancées ont été réalisées dans chacune des fonctions fondamentales du robot : **perception, action, apprentissage**. Dans les années 1985-2000, la problématique de la localisation et de la cartographie simultanées a connu un développement formidable qui a permis de bien en cerner les fondements et de produire des systèmes efficaces, le point faible important restant le manque d'interprétations plus sémantiques de l'environnement et des objets qui le composent.

Un autre domaine est celui des **systèmes multi-agents** ou **l'intelligence artificielle distribuée**, inspirée des comportements sociaux de certaines familles d'insectes, permettant la mise en œuvre de systèmes qui s'auto-organisent. Vos rapporteurs ont expérimenté ce système lors d'une rencontre avec les chercheurs du laboratoire de robotique et d'intelligence artificielle de l'Université Libre de Bruxelles dirigé par le professeur Hugues Bersini. Une colonie de petits robots peu intelligents travaillent ensemble, développent une coopération puis des stratégies plus complexes que ce que ne leur permet leur intelligence individuelle et démontrent ainsi la pertinence de l'intelligence collective. Les travaux de Michael Wooldridge sur les systèmes multi-agents, chercheur également rencontré par vos rapporteurs, illustrent les résultats de ces méthodes d'intelligence artificielle distribuée. **Les systèmes multi-agents et l'intelligence artificielle distribuée renvoient donc à des formes collectives de décision.** Le développement de systèmes de plus en plus complexes implique l'utilisation de connaissances expertes, hétérogènes, plus ou moins indépendantes les unes des autres.

---

Différents experts n'aboutissant pas toujours au même résultat, il faut confronter leurs décisions pour prendre une décision. Les architectures classiques (des modules qui s'enchaînent dans un ordre prédéfini) ont alors été remises en cause au profit d'architectures multi-agents. **Alors qu'un agent est un logiciel autonome percevant son environnement et agissant dessus, un système multi-agent est constitué d'un ensemble de tels agents, partageant des ressources communes et communiquant entre eux.**

On trouve principalement des agents peu complexes, n'utilisant ni buts ni plans (et qui sont généralement en grand nombre) ou des agents disposant de buts, de plans, de croyances et de connaissances (ces agents plus élaborés sont souvent peu nombreux). **Le point crucial de tels systèmes concerne la coordination entre les agents.** Pour ce faire, différents modes de communication entre ces agents sont possibles :

- soit chaque agent analyse les données contenues dans une zone commune et, s'il en trouve qu'il peut utiliser, il les traite et écrit de nouvelles données à utiliser par d'autres agents ;

- soit l'agent concerné quand il rencontre un problème envoie un message à d'autres agents afin de trouver qui peut l'aider à le résoudre. Le système adapte ainsi de manière dynamique son comportement à la situation à traiter.

La **programmation génétique**<sup>1</sup> donne lieu à l'élaboration d'algorithmes génétiques ou algorithmes évolutifs, qui imitent la façon dont la vie biologique a évolué sur terre. En effet, il est loisible d'interpréter le monde d'aujourd'hui comme une succession de stratégies gagnantes. Les espèces qui ont survécu à la sélection naturelle sont autant d'exemples de réussite. La nature a, par tâtonnement, créé un grand nombre de combinaisons de codes génétiques qu'elle a ensuite sélectionnés dans la mesure où ils fonctionnent, survivent et parviennent à dominer leur environnement. **Les algorithmes génétiques appliquent les mécanismes fondamentaux de l'évolution et de la sélection naturelle à des problèmes d'optimisation.** On code les caractéristiques des objets manipulés et on définit une fonction qui évalue la valeur attribuée à chaque objet. On fait évoluer une population initiale en créant de nouveaux objets à partir des anciens et en permettant diverses mutations. **La sélection permet d'éliminer les objets les moins efficaces.** Ce type de processus donne de bons résultats dans divers domaines, la difficulté résidant dans le choix du codage (c'est-à-dire les paramètres pertinents des objets considérés) et les types de mutations autorisées.

Les **réseaux bayésiens**, qui se situent à l'intersection de l'informatique et des statistiques<sup>2</sup>, donnent de bons résultats parmi les

---

<sup>1</sup> Les mécanismes d'évolution et de sélection génétique sont appliqués aux programmes informatiques.

<sup>2</sup> Les modèles de régression linéaire sont aussi utilisés comme méthode d'apprentissage supervisé pour prédire une variable quantitative. Ils peuvent aider à prédire un phénomène ou chercher à l'expliquer. L'inventeur de la notion en 1886, Francis Galton, mettait en évidence, dans un article

---

technologies d'intelligence artificielle. Un réseau bayésien est un outil mathématique de modélisation graphique probabiliste et d'analyse de données. La modélisation est graphique en ce qu'elle représente les variables aléatoires sous la forme d'un graphe orienté<sup>1</sup>. Judea Pearl, prix Turing<sup>2</sup> 2011 pour « *ses contributions fondamentales à l'intelligence artificielle par le développement de l'analyse probabiliste et du raisonnement causal* », est l'un des inventeurs de ces modèles. Ils sont particulièrement adaptés à la prise en compte de l'incertitude et peuvent être décrits manuellement par des experts ou produits automatiquement par apprentissage. Un réseau bayésien permet de représenter la connaissance acquise (modèle de représentation des connaissances) ou de découvrir la connaissance dans un contexte par l'analyse de données (c'est une machine à calculer les probabilités conditionnelles), afin de mener des opérations de prise de décisions, de diagnostic, de simulation et de contrôle d'un système. L'intérêt particulier des réseaux bayésiens est de tenir compte simultanément de connaissances *a priori* d'experts (dans le graphe) et de l'expérience contenue dans les données, ce qui est très pertinent pour l'aide à la décision.

Les réseaux bayésiens sont donc souvent utilisés pour des **solutions décisionnelles** qui correspondent souvent aux défis lancés aux technologies d'intelligence artificielle.

Michael Jordan, professeur à l'Université de Berkeley, que vos rapporteurs ont rencontré, identifie quatre pistes sur lesquelles il fait particulièrement travailler ses équipes, qui s'inscrivent dans le prolongement des modèles graphiques probabilistes et en particulier des réseaux bayésiens, dont il est une des figures avec Judea Pearl et Daphne Koller. Il décrit ainsi quatre axes de recherche pertinents : **les variables latentes, les modèles topiques, les modèles de causalité et les séries temporelles**.

Deux autres outils d'intelligence artificielle sont rappelés ici. La **programmation par contraintes**, qui se rapproche d'un raisonnement humain, et les **raisonnements à partir de cas**, qui se fondent sur la notion d'analogie.

Dans certains problèmes, on connaît les valeurs possibles que peuvent prendre certaines des variables - on parle alors de contraintes. Résoudre le problème consiste alors à affecter à chaque variable une valeur satisfaisant ces contraintes. L'évolution de Prolog système, évoqué plus haut,

---

*fondateur, un phénomène de « régression vers la moyenne » en analysant la taille des fils en fonction de la taille de leurs pères. Avant cela, Carl Friedrich Gauss avait démontré, dès 1821, un théorème relatif à l'estimateur linéaire selon la méthode des moindres carrés, connu aujourd'hui sous le nom de « théorème de Gauss-Markov », car redécouvert et complété en 1900 par Andrei Markov. Ce dernier a ainsi mis en évidence les processus aléatoires dans le calcul des probabilités. Ces aléas, ou « chaînes de Markov » sont les fondements théoriques du calcul stochastique.*

<sup>1</sup> Pour un domaine donné, on décrit les relations causales entre variables d'intérêt par un graphe. Dans ce graphe, les relations de cause à effet entre les variables ne sont pas déterministes, mais probabilisées. Ainsi, l'observation d'une cause ou de plusieurs causes n'entraîne pas systématiquement l'effet ou les effets qui en dépendent, mais modifie seulement la probabilité de les observer.

<sup>2</sup> Ce prix est la plus haute distinction en informatique.



---

a été fondée sur cet aspect. **Cette technique de programmation par contraintes**, d'origine française, **permet des raisonnements locaux, en simplifiant le problème global en sous problèmes partiels, puis une propagation des contraintes sur l'ensemble du problème.** Elle est largement utilisée en biologie moléculaire, en conception de produits industriels, en planification de production, en gestion du trafic dans les villes et les aéroports. Ses limites sont patentées dans les problèmes dynamiques (où les contraintes varient dans le temps) ou dans les problèmes « sur-contraints » (dans les cas où il n'existe pas de solution qui vérifie toutes les contraintes). Le fait de savoir quelles contraintes négliger reste un problème ouvert.

**Le raisonnement à partir de cas se fonde sur des analogies entre des expériences passées et un problème actuel.** On mémorise un certain nombre de situations spécifiques dans un domaine donné (les « cas ») et, face à un nouveau problème, on essaye de trouver le ou les cas les plus proches, puis on transpose les solutions déjà rencontrées au nouveau problème. C'est typiquement le **raisonnement utilisé par la justice pour adapter la jurisprudence à une nouvelle situation.** Deux étapes sont nécessaires pour ce type de raisonnement : l'indexation, qui sert à trouver les cas pertinents pour le problème actuel, et l'adaptation, pour modifier un ou plusieurs cas et les rendre applicables au problème actuel. Les métriques permettant l'indexation calculent une mesure de similarité entre cas. La difficulté essentielle de ce type de raisonnement étant de trouver le chemin qui va de la solution du cas connu au problème en cours, sachant que les métriques les plus élaborées tentent d'en tenir compte.

**Le mot algorithme est issu, comme il a été vu, de la latinisation du nom du mathématicien Al-Khwarizmi** et correspond à une **suite finie et non ambiguë d'opérations ou d'instructions permettant de résoudre un problème ou d'obtenir un résultat.** La difficulté liée aux algorithmes classiques réside dans le fait que l'ensemble de tous les comportements possibles compte tenu de toutes les entrées possibles devient rapidement trop complexe à décrire. Cette **explosion combinatoire justifie de confier à des programmes le soin d'ajuster un modèle adaptatif permettant de simplifier cette complexité** et de l'utiliser de manière opérationnelle en prenant en compte l'évolution de la base des informations pour lesquelles les comportements en réponse ont été validés. C'est ce que l'on appelle l'**apprentissage automatique** ou *machine learning*, qui permet donc d'apprendre et d'améliorer le système d'analyse ou de réponse. En ce sens, on peut dire que **ces types particuliers d'algorithmes apprennent.**

Un apprentissage est dit « **supervisé** » lorsque le réseau est forcé à **converger vers un état final précis**, en même temps qu'un motif lui est présenté.

À l'inverse, lors d'un **apprentissage « non-supervisé »**, le réseau est laissé **libre de converger vers n'importe quel état final** lorsqu'un motif ou un élément lui est présenté.

---

Entre ces deux extrêmes, l'apprentissage automatique ou *machine learning* peut-être **semi-supervisé ou partiellement supervisé**.

L'apprentissage automatique peut lui-même reposer sur plusieurs méthodes : **l'apprentissage par renforcement, l'apprentissage par transfert**, ou, encore, **l'apprentissage profond**, qui sera vu plus tard.

L'apprentissage par renforcement conduit l'algorithme à apprendre, à partir d'expériences ou d'observations, un **comportement optimal** ou stratégie, selon une **logique itérative de recherche de récompenses**, un peu comme dans le cas du dressage d'un animal. L'action de l'algorithme sur un environnement donné produit une valeur de retour qui guide son apprentissage dans la mesure où l'algorithme cherche dans ce cadre d'apprentissage par renforcement à optimiser sa fonction de récompense quantitative au cours des expériences. Par exemple, l'algorithme de « Q-learning », qui optimise les actions accessibles sans même avoir de connaissance initiale de l'environnement par une comparaison de récompenses probables, est un cas d'apprentissage par renforcement.

L'apprentissage par transfert peut être vu comme la capacité d'un système à **reconnaître à partir de tâches antérieures** apprises des connaissances et des compétences, puis à **appliquer ces dernières sur de nouvelles tâches** partageant des similitudes.

L'apprentissage supervisé permet, de plus, des méthodes prédictives utiles en reconnaissance de formes, selon plusieurs approches : les **arbres de décision**, les **réseaux bayésiens** ou, encore, les **machines à vecteurs de support**.

Avec les **arbres de décision**<sup>1</sup>, les algorithmes d'apprentissage supervisés peuvent calculer automatiquement à partir de bases de données en sélectionnant automatiquement les variables discriminantes à partir de données non-structurées et potentiellement volumineuses. Ils peuvent ainsi permettre **d'extraire des règles logiques de cause à effet** (des déterminismes) **qui n'apparaissent pas initialement dans les données brutes**.

Avec les **réseaux bayésiens**, qui ont été mentionnés plus hauts, l'apprentissage automatique peut être utilisé de deux façons : pour estimer la structure d'un réseau, ou pour estimer les tables de probabilités d'un réseau, dans les deux cas à partir de données. L'intérêt particulier des réseaux bayésiens est de **tenir compte simultanément de connaissances a priori d'experts et de l'expérience contenue dans les données**.

Les **machines à vecteurs de support** (en anglais *support vector machines* ou SVM), parfois appelées séparateurs à vaste marge sont des techniques d'apprentissage supervisé reposant sur les notions de marge maximale et de fonction noyau, destinées à **résoudre des problèmes de**

---

<sup>1</sup> Un arbre de décision est un outil d'aide à la décision représentant un ensemble de choix sous la forme graphique d'un arbre.

---

**discrimination et de régression.** Il s'agit de classifieurs linéaires<sup>1</sup> dont les excellentes capacités de généralisation leur ont permis d'être l'une des technologies dominantes en intelligence artificielle dans les années 1990 et 2000.

La **méthode Monte-Carlo**<sup>2</sup> et la **méthode du recuit simulé**<sup>3</sup>, techniques plus anciennes, sont d'autres méthodes dont le but est de trouver une solution optimale pour un problème donné et qui peuvent se combiner avec les technologies d'apprentissage automatique.

Les **réseaux de neurones artificiels** prennent en compte l'apprentissage de manière dite « implicite » ou, en tout état de cause, plus implicite que l'ensemble des méthodes qui viennent d'être présentées.

Un **réseau de neurones artificiels** est constitué d'un ensemble d'**éléments interconnectés, chacun ayant des entrées et des sorties numériques**. Le comportement d'un neurone artificiel dépend de la somme pondérée de ses valeurs d'entrée. **Si cette somme dépasse un certain seuil, la sortie prend une valeur positive, sinon elle reste nulle.** Un réseau peut comporter une couche d'entrée (les données), une de sortie (les résultats), et une ou plusieurs couches intermédiaires, avec ou sans boucles.

Le principe de fonctionnement consiste, dans une première phase, à présenter en entrée les valeurs correspondant à de nombreux exemples, et en sortie les valeurs respectives des résultats souhaités. **Cet apprentissage permet d'ajuster les poids synaptiques afin que les correspondances entre les entrées et les sorties soient les meilleures possible.** Après un nombre statistiquement pertinent d'exemples, l'apprentissage (implicite) est terminé et le réseau peut être utilisé, dans une seconde phase, pour la reconnaissance. Comme il produit toujours une sortie, même pour des entrées non rencontrées auparavant, **il a le plus souvent une bonne capacité de généralisation, qui dépend du corpus d'apprentissage.**

Il s'agit donc de **combinaison de nombreuses fonctions simples pour former des fonctions complexes et d'apprendre les liens entre ces fonctions simples à partir d'exemples étiquetés.** L'analogie avec le fonctionnement du cerveau humain repose sur le fait que les fonctions simples rappellent le rôle joué par les neurones, tandis que les connexions rappellent les synapses. **Il ne s'agit en aucun cas de réseaux de neurones de synthèse, ce n'est qu'une**

---

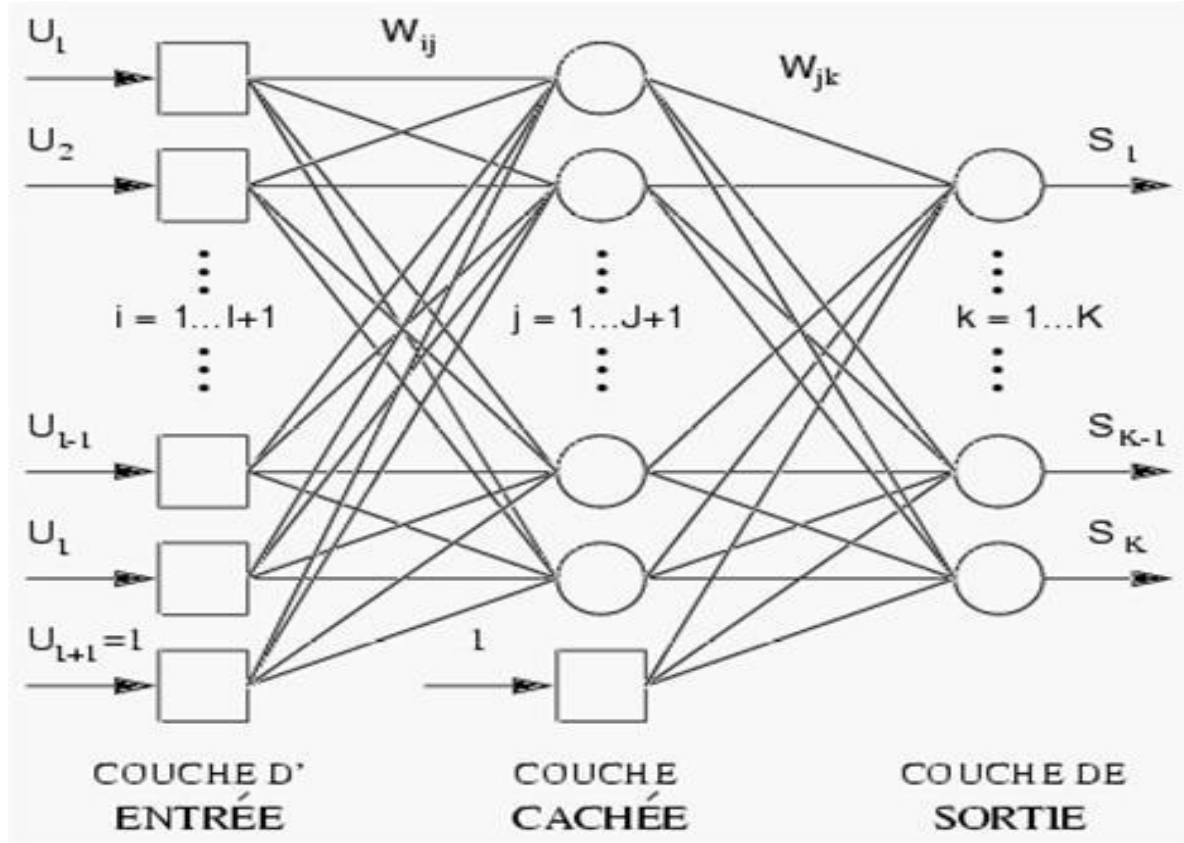
<sup>1</sup> Algorithmes de classement statistique, les classifieurs permettent de classer dans des groupes des échantillons qui ont des propriétés similaires, mesurées par des observations. Un classifieur linéaire en est un type particulier, qui calcule la décision par combinaison linéaire des échantillons.

<sup>2</sup> Il s'agit de méthodes algorithmiques visant à calculer une valeur numérique approchée en utilisant des techniques probabilistes ou aléatoires. Nicholas Metropolis a utilisé le nom de méthode Monte-Carlo en 1947 en faisant allusion aux jeux de hasard pratiqués au casino de Monte-Carlo.

<sup>3</sup> Adapté par des chercheurs d'IBM en 1983, le recuit simulé est une méthode empirique ou méta-heuristique cherchant à optimiser les chances de découvertes des extrêmes d'une fonction. Elle est inspirée d'un processus traditionnel utilisé en métallurgie, qui consiste à alterner des cycles de refroidissement lent et de réchauffage dans le but de minimiser l'énergie du matériau. Elle a été, S. Kirkpatrick, C.D. Gelatt et M.P. Vecchi, et indépendamment par V. Černý en 1985.

image, sans doute malheureuse selon vos rapporteurs car elle entretient une forme de **confusion**, en lien avec la notion d'**intelligence artificielle**.

Schéma d'un réseau de neurones artificiels



Source : Académie des technologies

D'autres technologies peuvent être citées comme la recherche dans les espaces d'états, la planification (très efficace au jeu d'échecs), les ontologies, les logiques de description... Les domaines de l'intelligence artificielle, comme le traitement du langage naturel ou la vision artificielle utilisent plusieurs des technologies disponibles, qui, comme il sera vu plus loin, peuvent de plus **se combiner entre elles**.

Dans la période récente, un système d'intelligence artificielle nommé *Libratus*, créé par le professeur Tuomas Sandholm et son doctorant Noam Brown, tous deux chercheurs de l'université Carnegie Mellon de Pittsburgh, a affronté et **battu en janvier 2017 quatre joueurs de poker professionnels** dans un casino de Pennsylvanie, au cours d'une partie de poker<sup>1</sup> de 120 000 mains successives sur 20 jours, intitulée « *Cerveau contre Intelligence Artificielle : on monte la mise* » (« *Brains Vs. Artificial Intelligence : Upping the Ante* »). Sa victoire sans appel, avec un gain de 1,8 million de dollars (contre des pertes pour tous les autres joueurs), marque à son tour **l'histoire des**

<sup>1</sup> Il s'agit de parties de poker « Texas Hold'em », en face à face ou heads-up et sans limite de mise ou no limit.

---

**progrès des systèmes d'intelligence artificielle**, surtout que **le poker requiert une forme de raisonnement particulièrement difficile à imiter pour une machine**. *Libratus* a utilisé les capacités de calcul du superordinateur de l'université Carnegie Mellon et combiné des algorithmes de Public Chance Sampling (PCS, à ce stade non traduit en français et qui signifie « Échantillonnage de hasard public »), variante de la « réduction des regrets contrefactuels »<sup>1</sup> (Counterfactual Regret Minimization ou CFR), avec la méthode d'Oskari Tammelin introduite en 2014, permettant l'optimisation des résultats dans un contexte d'informations imparfaites<sup>2</sup>. Un article collectif paru dans la revue « *Science* » en 2015 présentait déjà les évolutions théoriques nécessaires à cette victoire<sup>3</sup>.

En effet, un joueur de poker ne connaît pas les ressources (cartes) et les stratégies (sincérité ou pas) de son adversaire et doit donc agir sans informations certaines et sans écarter la possibilité d'un bluff. La réflexion de la machine au poker doit donc **prendre en compte des données incomplètes ou dissimulées ce qui distingue le poker d'autres jeux comme le go ou les échecs**, dans lesquels l'intelligence artificielle avait déjà démontré sa supériorité sur l'homme. Le poker fait intervenir les notions de hasard, de piège et de bluff, alors que les jeux où dominaient l'intelligence artificielle jusqu'en janvier 2017 étaient fondés sur des stratégies relevant de l'analyse combinatoire : les deux adversaires s'y affrontaient en continu en visualisant l'ensemble du jeu et des pions.

Cette victoire de *Libratus* en 2017 n'a reposé que sur des duels, la machine jouant contre un seul joueur à la fois. La prochaine étape pour les développeurs sera d'assurer la victoire d'une intelligence artificielle dans des parties à plusieurs. Tuomas Sandholm et Noam Brown travaillent à ce nouveau projet.

De manière caricaturale, on pourrait résumer les technologies d'intelligence artificielle à un champ de recherche où s'opposent deux grands types d'approches : les **approches symboliques** et les **approches connexionnistes**. Comme il a été vu à travers la description des nombreuses technologies développées la réalité est souvent plus complexe que cela puisqu'il existe une multitude de technologies, qui de plus peuvent se conjuguer. Parmi les approches connexionnistes, voire parmi toutes les familles d'approches en intelligence artificielle, l'apprentissage profond ou « *deep learning* » est devenu dominant au cours des dernières décennies, en particulier au cours des quatre dernières années.

---

<sup>1</sup> Cf. <https://www.quora.com/What-is-an-intuitive-explanation-of-counterfactual-regret-minimization>

<sup>2</sup> Cf. <https://arxiv.org/abs/1407.5042>

<sup>3</sup> Cf. <http://science.sciencemag.org/content/347/6218/145>

---

## B. L'ACCÉLÉRATION RÉCENTE DE L'USAGE DES TECHNOLOGIES D'INTELLIGENCE ARTIFICIELLE GRÂCE AUX PROGRÈS EN APPRENTISSAGE AUTOMATIQUE (« MACHINE LEARNING »)

### 1. Les découvertes en apprentissage profond (« deep learning ») remontent surtout aux années 1980, par un recours aux « réseaux de neurones artificiels » imaginés dès les années 1940

Cette technologie d'apprentissage profond ou « *deep learning* » rencontre un **succès particulièrement remarquable dans les années 2010**, pourtant elle est ancienne. Son essor doit beaucoup à l'émergence récente de données massives (« *big data* ») et à l'accélération de la vitesse de calcul des processeurs, mais son histoire remonte aux années 1940, ce que vos rapporteurs ont évoqué mais jugent nécessaire de rappeler de façon plus détaillée ici : comme il a été vu les « **réseaux de neurones artificiels** » sont **imaginés dès les années 1940**<sup>1</sup> et aboutissent au **Perceptron** à la fin des années 1950. En 1957, au laboratoire d'aéronautique de l'université Cornell, Frank Rosenblatt invente ce dernier à partir des travaux de McCulloch et Pitts, ce qui constitue la **première modélisation d'un réseau de neurones artificiels, dans sa forme la plus simple, à savoir un classifieur linéaire**<sup>2</sup>. Marvin Minsky ayant pointé comme il a été vu les défauts de ce système, des **perceptrons multi-couches** ont ensuite été proposés en 1986, parallèlement, par David Rumelhart et Yann LeCun.

Les réseaux de neurones artificiels peuvent être à **apprentissage supervisé ou non** (ils sont le plus souvent supervisés, comme dans le cas du Perceptron), **avec ou sans rétropropagation** (*back propagation*).

Outre les réseaux multi-couches, d'importantes découvertes en apprentissage profond (« *deep learning* ») remontent aux années 1980, telles que la **rétropropagation du gradient**. Les pionniers de ces découvertes sont Paul Werbos, David Parker et le français Yann LeCun, rencontré plusieurs fois par vos rapporteurs. David Rumelhart, Geoffrey Hinton et Ronald Williams théorisent cette découverte en 1986 dans un fameux article intitulé *Learning representations by back-propagating errors*<sup>3</sup>. L'idée générale de la rétropropagation consiste à rétropropager l'erreur commise par un neurone à ses synapses et aux neurones qui y sont reliés.

Le principe de rétropropagation du gradient fonde les **méthodes d'optimisation utilisées dans les réseaux de neurones multicouches**, comme les perceptrons multicouches. Il s'agit en effet de **faire converger l'algorithme de manière itérative vers une configuration optimisée des**

---

<sup>1</sup> Les mathématiciens et neurologues Warren McCulloch et Walter Pitts posent dès 1943 l'hypothèse que les neurones avec leurs deux états, activé ou non-activé, pourrait permettre la construction d'une machine capable de procéder à des calculs logiques. Ils publièrent dès la fin des années 1950 des travaux plus aboutis sur les réseaux de neurones artificiels.

<sup>2</sup> « The Perceptron : A Probabilistic Model for Information Storage and Organization in the Brain ».

<sup>3</sup> <http://www.nature.com/nature/journal/v323/n6088/abs/323533a0.html>

---

**poids synaptiques.** L'algorithme étant itératif, la correction s'applique autant de fois que nécessaire pour obtenir une bonne prédiction. Une vigilance est requise face aux problèmes de surapprentissage liés à un mauvais dimensionnement du réseau ou un apprentissage trop poussé.

La correction des erreurs peut aussi se faire selon d'autres méthodes, en particulier le **calcul de la dérivée seconde**.

**L'apprentissage profond** ou *deep learning* regroupe donc des **méthodes plus récentes d'apprentissage automatique**, ou *machine learning* dont elles sont une sous-catégorie.

Ces méthodes, parfois qualifiées de **révolution dans le domaine de l'intelligence artificielle**, ont pour spécificité d'utiliser **des modèles de données issus d'architectures articulées en différentes couches d'unité de traitement non-linéaire, qui sont autant de niveaux d'abstraction des données**. La façon de rétropropager l'erreur au sein de plusieurs couches cachées permet de généraliser plus efficacement, ce qui permet des représentations de plus haut niveau et une capacité à traiter des données plus complexes.

Selon Yann LeCun « *Les cerveaux humain et animal sont "profonds", dans le sens où chaque action est le résultat d'une longue chaîne de communications synaptiques (de nombreuses couches de traitement). Nous recherchons des algorithmes d'apprentissage correspondants à ces "architectures profondes". Nous pensons que comprendre l'apprentissage profond ne nous servira pas uniquement à construire des machines plus intelligentes, mais nous aidera également à mieux comprendre l'intelligence humaine et ses mécanismes d'apprentissages* ».

L'apprentissage profond a récemment fait une incursion considérable en **robotique**, contestant la place dominante de l'apprentissage bayésien, à la fois pour la perception et pour la synthèse d'actions. Mais, comme le remarque Raja Chatila, la perception en robotique nécessite une *interaction* du robot avec son environnement et non une simple *observation* de celui-ci. L'apprentissage par renforcement est un apprentissage souvent non supervisé qui permet au robot de découvrir à la fois les effets de ses actions, caractérisés par une « récompense » obtenue comme conséquence de l'action, et l'incertitude de ses actions qui n'ont pas toujours les mêmes effets. Le lien entre perception et apprentissage - en particulier avec l'apprentissage par renforcement - est essentiel pour extraire la notion d'*affordance* qui rend compte des propriétés des objets en ce qu'elles représentent pour l'agent, et qui associe les représentations perceptuelles aux capacités d'action. C'est cela qui sert de base au robot pour exprimer le sens du monde qui l'entoure.

## **2. L'apprentissage profond connaît un essor inédit dans les années 2010 avec l'émergence de données massives (« big data ») et l'accélération de la vitesse de calcul des processeurs**

En *deep learning*, toute chose étant égale par ailleurs, **plus les données sont massives, variées et rapides** (ce sont les trois V du *big data* : un



---

volume croissant de données, issues d'une large variété de sources et qui circulent à une vitesse élevée), **plus les algorithmes donnent de bons résultats.**

Avec la vitesse de calcul des processeurs, dont il sera question plus loin, les données massives (« *big data* ») ont conduit dans la période récente à **d'importantes améliorations dans l'efficacité des algorithmes.** Mais ces données massives n'ont pas été par elles-mêmes le seul facteur de progression des algorithmes dans les années 2000 et 2010, la réalisation de **grandes bases de données labellisées** a souvent constitué un préalable<sup>1</sup>.

En 1998, MNIST a fait figure de pionnier en utilisant les images des données postales manuscrites de la poste américaine. Deux autres bases de données labellisées ont plus récemment permis aux développeurs d'entraîner, de faire progresser et de comparer leurs algorithmes. En 2009, l'Institut canadien de recherche avancée, basé à Toronto a créé les bases de données CIFAR-10 et CIFAR-100 du nom de l'acronyme anglais de l'Institut, *Canadian Institute for Advanced Research* ou CIFAR. La distinction entre les deux bases de données, vient du nombre de classes utilisé pour l'apprentissage : 10 classes de données pour CIFAR-10 ou 100 pour CIFAR-100. En 2010, le projet **ImageNet** a été lancé aux États-Unis<sup>2</sup> avec l'idée d'organiser un concours annuel sur les programmes mis en place pour traiter la base de données éponyme. Ce concours s'intitule *ImageNet Large Scale Visual Recognition Challenge* (ILSVRC) et rassemble plus de 50 organisations participant chaque année (universités, centres de recherche, entreprises...). En 2016, la base de données avait annoté un total de **dix millions d'images disponibles sur Internet.**

Les **données sont donc essentielles car l'apprentissage des algorithmes repose sur celles-ci.** L'acquisition de données annotées représente un enjeu stratégique pour les États et un enjeu industriel pour les entreprises.

Ces dernières, telles Google ou Facebook, donnent d'ailleurs assez largement **accès à leurs logiciels en *open-source*,** mais - à ce stade - pas à leurs données. La réflexion sur *l'open-source* est importante mais doit aller jusqu'à **poser la question de l'accès aux données.** Pour Stéphane Mallat, professeur à l'École normale supérieure et rencontré par vos rapporteurs, pour donner des résultats satisfaisants, « *ces algorithmes (de deep learning) doivent tout d'abord être alimentés par des quantités de données gargantuesques. C'est pour cela que DeepMind, le projet de Google, possède aujourd'hui une telle longueur d'avance* ». Les grandes firmes américaines disposent en effet de données personnelles massives, qu'elles peuvent utiliser librement dans leurs projets de recherche internes. Mais l'étendue des corpus de données ne

---

<sup>1</sup> Une liste de ces nombreuses bases de données peut être consultée ici : [https://en.wikipedia.org/wiki/List\\_of\\_datasets\\_for\\_machine\\_learning\\_research](https://en.wikipedia.org/wiki/List_of_datasets_for_machine_learning_research)

<sup>2</sup> Avec des chercheurs tels que Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg ou Fei-Fei Li. Site : <http://image-net.org/>



---

fait pas tout : « pour les applications médicales, il ne suffit pas d'avoir à disposition un grand nombre de mesures par patients : encore faut-il qu'elles portent sur beaucoup de personnes différentes. Sinon, la règle construite par l'algorithme fonctionnera peut-être très bien pour une personne donnée... mais sera difficilement généralisable à toute la population. La médecine serait le champ de recherche le plus propre à bénéficier des big data... mais c'est celui qui est le plus entravé par les problématiques de confidentialité des données ». Volume, variété et vélocité sont bien les trois V complémentaires du big data selon la formule consacrée déjà rappelée par vos rapporteurs.

Les **techniques d'apprentissage automatique**, ou *machine learning*, se renforcent au cours des quinze dernières années, surtout au cours des cinq dernières années en bénéficiant du concours de données massives ou *big data*. Sans que d'importantes nouveautés théoriques n'aient émergées, à l'exception de l'apprentissage profond ou *deep learning*, les outils de l'intelligence artificielle se sont largement diffusés, aussi bien dans la vie quotidienne que dans des applications industrielles ou militaires. Il convient de relever que nous ne disposons d'aucune explication théorique des raisons pour lesquelles les réseaux de neurones fonctionnent, c'est-à-dire donnent, dans un certain nombre de domaines, d'excellents résultats.

En apprentissage profond, qui repose donc sur des réseaux de neurones profonds (*deep neural networks*), **on peut distinguer les technologies** selon la **manière particulière d'organiser les neurones en réseau** : les réseaux peuvent être **en couches**, tel le perceptron, les perceptrons multicouches et les architectures profondes (plusieurs dizaines ou centaines de couches), dans lesquels chaque neurone d'une couche est connecté à tous les neurones des couches précédentes et suivantes (c'est la structure la plus fréquente), les réseaux **totalelement interconnectés**, dans lesquels tous les neurones sont connectés les uns aux autres (cas rare des « réseaux de Hopfield » et des « machines de Boltzmann »), les **réseaux neuronaux récurrents** et les **réseaux neuronaux à convolution**.

Ces deux dernières technologies, imaginées à la fin des années 1980 et au début des années 1990, font l'objet d'investigations particulièrement poussées et d'applications de plus en plus riches depuis trois ans.

Les **réseaux neuronaux récurrents** (RNR ou *recurrent neural networks-RNN* en anglais) permettent de **prendre en compte le contexte** et de relever le défi de traiter des séquences avec des réseaux de neurones (il existe, au moins, un cycle dans la structure du réseau). Au sein de ces RNR, on relève les architectures MARNN (pour *Memory-Augmented Recurrent Neural Networks* ou réseaux neuronaux récurrents à mémoire augmentée), les architectures LSTM (pour *Long Short Term Memory*), les architectures BLSTM (pour *Bidirectional Long Short Term Memory*), les architectures BPTT (pour *BackProp Through Time*), les architectures RTRL (pour *Real Time Recurrent Learning*) et les architectures combinées, avec par exemple des modèles de

---

Markov<sup>1</sup> à états cachés (MMC, ou *Hidden Markov Model*, HMM en anglais). Ces RNR, notamment les LSTM et les MARNN, forment un chantier de recherche prioritaire pour les chercheurs de Google (DeepMind en particulier), Baidu, Apple, Microsoft et Facebook. Leur utilisation pour la traduction, la production de légendes pour les images et les systèmes de dialogues vise à répondre à la question de la capacité à apprendre des tâches qui impliquent non seulement d'apprendre à se représenter le monde, mais aussi à se remémorer, à raisonner, à prédire et à planifier. L'apprentissage par renforcement recourt de plus en plus à ces RNR, notamment en combinaison avec des algorithmes génétiques qui permettent de mieux les entraîner. Les LSTM peuvent aussi faire l'objet d'améliorations avec les SVM.

Les **réseaux neuronaux à convolution** (RNC) appelés aussi réseaux de neurones profonds convolutifs (*convolutional deep neural networks* ou CNN) sont **inspirés des processus biologiques du cortex visuel des animaux**. En effet, les neurones de cette région du cerveau sont arrangés de sorte à ce qu'ils correspondent à des régions qui se chevauchent lors du pavage du champ visuel, le motif de connexion entre ces réseaux de neurones artificiels à convolution repose sur un procédé similaire. Les réseaux neuronaux sont ici soumis à un mécanisme de poids synaptiques partagés, qui offre l'intérêt d'une plus grande capacité de généralisation pour moins de paramètres. Destinées en priorité à traiter les images, et trouvant leurs principales applications en reconnaissance d'images et de vidéos, leurs applications sont et seront de plus en plus diversifiées, du traitement du langage naturel aux systèmes de recommandation.

L'essor de l'intelligence artificielle avec le *deep learning* est facilité **par la croissance exponentielle des avancées technologiques matérielles dans ce secteur**, en particulier les vitesses de calcul des processeurs, appelée aussi « loi de Moore ».

La « loi de Moore » est une **conjecture**, et donc en réalité une supposition, **concernant l'évolution de la puissance de calcul des ordinateurs** et, plus généralement, la complexité du matériel informatique. En 1965, Gordon Moore, ingénieur chez Fairchild Semiconductor, un des trois fondateurs d'Intel, constate que depuis 1959 la complexité des semi-conducteurs d'entrée de gamme a doublé tous les ans à coût constant.

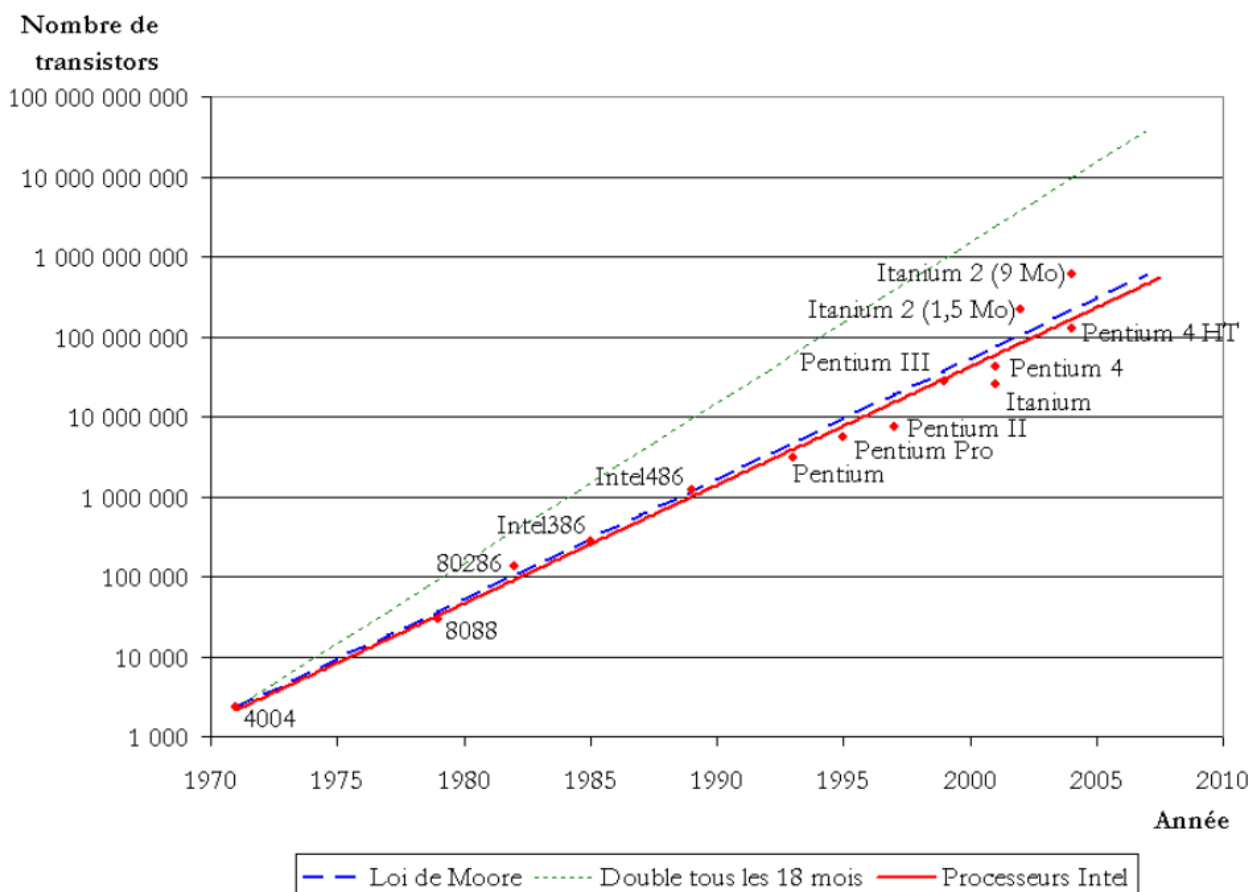
Il s'agit donc d'une **loi relative au développement exponentiel des capacités de traitement de l'information en vertu d'un doublement constaté pour le même coût, depuis une quarantaine d'année, du nombre de transistors des microprocesseurs sur une puce de silicium**. L'observation

---

<sup>1</sup> En mathématiques, un processus de Markov (la théorie des probabilités parle plutôt de « processus de décision markovien ») est une chaîne stochastique possédant la propriété de Markov qui réside dans le fait que la prédiction du futur à partir du présent n'est pas rendue plus précise par des éléments d'information concernant le passé. L'agent prend dans ce cas des décisions avec un résultat aléatoire de ses actions. Claude Shannon s'en est inspiré en 1948 pour fonder sa théorie de l'information et l'algorithme de classement par popularité de Google repose notamment sur un modèle de ce type. L'apprentissage par renforcement permet de résoudre le problème des processus markoviens.

empirique, ainsi que l'illustre le graphique suivant, démontre que ce doublement a lieu tous les dix-huit mois.

### La loi de Moore rapportée à l'évolution réelle du nombre de transistors dans les microprocesseurs



Source : contributeur « QcRef87 », licence de documentation libre

Cette conjecture, connue sous le nom de loi de Moore, est le fondement sur lequel reposent certains spécialistes pour fixer l'avènement de l'intelligence artificielle forte en 2030. Or, vos rapporteurs rappellent que **la conjecture n'a trait qu'aux capacité de calcul et de stockage de données informatiques**, elle n'est donc **pas de nature à garantir ou à permettre une prévision de la future date de naissance d'une intelligence artificielle égale à celle de l'homme, contrairement à ce que certains font valoir.**

Ils notent, par ailleurs, que le *deep learning* a largement profité des **processeurs graphiques** dédiés (GPU), souvent issus des exigences des joueurs de jeux vidéo. À la différence des processeurs principaux traditionnels (CPU) aux fréquences d'horloge élevées, les GPU possèdent de nombreux cœurs (unités de calcul), composants parfaitement adaptés aux **traitements parallélisables de données de grande dimension.**

L'intelligence artificielle, dont certaines des technologies recourent à des analyses qui multiplient les matrices et les convolutions (cas des RNC et

des RNR) a donc profité, au cours des dernières années, de ces processeurs graphiques plus efficaces.

La loi de Moore devrait, à technologie égale, atteindre les **limites des capacités des puces en silicium**. Il semble en effet **difficile d'écrire à terme sur des surfaces plus petites que la taille de l'atome**. L'avenir de l'accélération des vitesses des processeurs pourrait donc dépendre des innovations futures **en informatique quantique** ou, encore, des inventions en matière de **processeurs fondés sur l'optique**. La *start-up* française *LightOn*, créée par Igor Carron, co-fondateur du *Paris Machine Learning Meetup*, poursuit ses recherches dans ce sens.

À ce stade du rapport, il semble utile de récapituler par une chronologie les principales étapes et découvertes en intelligence artificielle.

### Chronologie des principales étapes et découvertes en intelligence artificielle

330 avant J.-C. : Invention par Euclide de l'algorithme de calcul du plus grand diviseur commun de deux nombres entiers ;

833 : Le mathématicien Al-Khawarizmi, dont les travaux fondent l'algèbre, invente des méthodes précises de résolution des équations du second degré, qui seront appelées algorithmes. Son nom latinisé est à l'origine du mot « algorithme » ;

1694 : Leibniz construit la première machine à calculer ;

1780 : Invention d'une machine à parler par l'abbé Mical et Christian Gottlieb Kratzenstein ;

1834 : Invention du premier « ordinateur » sous la forme d'une « machine à différences » inspirée par les machines à tisser par Charles Babbage ;

1847 et 1854 : Premières mathématisations de la logique par Georges Boole ;

1869 : Création de pianos mécaniques capables de raisonner par William Stanley Jevons ;

1936 : Formulation des fondements théoriques de l'informatique par Alan Turing (son appareil sera plus tard appelé « machine de Turing ») par l'introduction des concepts de programme et de programmation ;

1943 : Premier article sur le potentiel des réseaux de neurones artificiels par Warren McCulloch et Walter Pitts ;

1950 : Invention du « test de Turing » en vue d'évaluer l'intelligence d'un ordinateur par rapport à celle d'un être humain ;

1956 : Invention, en tant que discipline et en tant que concept, de l'intelligence artificielle lors de la conférence de Dartmouth par John McCarthy et Marvin Minsky ;

1957 : Invention du perceptron, première utilisation de réseaux de neurones artificiels modélisés, par Frank Rosenblatt ;

1958 : Invention du langage de programmation Lisp ;

1965 : Formulation par Gordon Moore de la loi qui porte son nom concernant le doublement de la vitesse de calcul des ordinateurs tous les 18 mois à coût constant ;

1966 : Invention par Joseph Weizenbaum du premier agent conversationnel « Eliza » ;

1974 : Invention du premier système expert, dit « Mycin » ;

1986 : Invention des perceptrons multicouches par Yann LeCun et David Rumelhart et de la rétropropagation du gradient par David Rumelhart, Geoffrey Hinton et Ronald Williams ;

1997 : Victoire du système Deep Blue aux échecs face à Garry Kasparov ;

Années 2000 et 2010 : Conjugaison efficace des technologies de *deep learning* avec l'émergence des données massives et l'accélération marquée de la vitesse de calcul des processeurs ;

Années 2010 : Les réseaux neuronaux récurrents (RNR) et les réseaux neuronaux à convolutions (RNC), imaginés dès la fin des années 1980, font l'objet d'usages particulièrement remarquables ;

2011 : Victoire du système *Watson* au jeu télévisé *Jeopardy* en 2011 ;

2016 : Victoire du système *AlphaGo* au jeu de Go face au champion Lee Sedol.

2017 : Victoire du système *Libratus* au cours d'une partie de poker de 20 jours face à quatre joueurs professionnels

Source : OPECST

### 3. Les technologies d'intelligence artificielle conduisent d'ores et déjà à des applications dans de nombreux secteurs








Les applications sectorielles présentes ou futures sont d'envergure considérable, que l'on pense par exemple **aux transports, à l'aéronautique, à l'énergie, à l'environnement, à l'agriculture, au commerce<sup>1</sup>, à la finance, à la défense, à la sécurité, à la sécurité informatique, à la communication, à l'éducation, aux loisirs, à la santé, à la dépendance ou au handicap.**

Il s'agit d'autant de jalons d'applications sectorielles. Car en réalité, derrière le concept d'intelligence artificielle, ce sont **des technologies très variées qui donnent lieu à des applications spécifiques pour des tâches toujours très spécialisées.**

---

<sup>1</sup> Le « yield management », qui consiste à faire varier les prix en vue de l'optimisation du remplissage (transport aérien et ferroviaire ou hôtellerie) et/ou du chiffre d'affaires, est déjà connu de chacun de nous. En 2013, à titre d'exemple, Amazon changeait ses prix en moyenne plus de 2,5 millions de fois par jour.

## Applications des technologies d'intelligence artificielle en France

FILIÈRES INDUSTRIELLES ET PRIORITÉS IA	ENTREPRISES
<b>Aéronautique, Espace</b> Développement de systèmes autonomes (avions, imagerie satellitaire), et leur circulation dans l'espace aérien et en orbite Automatisation de la production et de la maintenance des plateformes et systèmes. Conception de produits complexes assistée pour les bureaux d'études avec aide à la décision.	 <b>AIRBUS</b> : Avions et hélicoptères à la demande. Airbus Smarter Fleet (avec IBM). Production: Robots humanoïdes (projet FUTURASSY avec Kawada). Inspection automatisée. Détection d'images satellitaires (avec Google, projet OneAtlas).  <b>THALES</b> : Cybersécurité. Circulation aérienne intelligente.  <b>SAFRAN</b> : Aide à la décision  <b>DASSAULT</b> : Systèmes de pilotage autonome pour avion de combat.
<b>Banque, Assurance</b> Recherche et analyse des clients approfondie (assistance conseillers). Systèmes anti-fraude	<b>Crédit mutuel</b> : partenariat IBM Watson. <b>BNPP</b> : investissement dans Smartly.ai (Chatbots). Partenariat avec Critéo. <b>SG</b> : assistants conseillers. <b>Axa</b> : création d'un fonds pour la recherche finance des chercheurs qui travaillent sur l'IA et sur la confiance dans le Big Data
<b>Energie, Environnement</b> Surveillance de sites industriels. Villes intelligentes: exploitation des données clients avec capteurs intelligents (smart grid et IoT)	<b>eDF</b> : développement de réseaux intelligents (smart grids). <b>Veolia</b> : récolte et tri de déchets intelligente (avec Huawei)
<b>Distribution, luxe, tourisme (B2C)</b> Aide à la recherche produit/service et à la décision des clients Placement produit multi-canal optimisés.	<b>L'Oréal</b> : investissement en systèmes prédictifs et relation client. <b>Accor</b> : intégration moteurs IA, pour CRM et marketing avancé. <b>Publicis</b> : système Cyc pour aide à la décision.
<b>Santé</b> Développement de traitements plus efficaces. Exploitation de données santé clients adaptées.	<b>Sanofi</b> : co-entreprise Onduo (avec Google) lutte contre le diabète. Traitement du cancer avec IBM (Watson). Accès à une base de données de 118 malades avec la FDA (US). <b>Dassault Systèmes</b> avec Ipsen : Optimisation processus de R&D (consortium BioIntelligence)
<b>Transports</b> Analyse clients. Développement de systèmes autonomes.	 <b>Alstom</b> : investissement navettes autonomes EasyMile  <b>Renault/Valéo</b> : développement de systèmes de conduite autonome. Investissement dans des capteurs intelligents.  <b>PSA</b> : prototype de voiture autonome Picasso C4.
<b>Télécoms</b> Gestion et optimisation du réseau. Développement et traitement automatique de la relation client. Interfaces client (chatbots, box en connexion avec l'environnement).	<b>Orange</b> : apprentissage automatique appliqué à la relation clients et centres d'appel. <b>Bouygues</b> : monétisation des données et systèmes anti-fraude. <b>Nokia</b> : augmentation des services de localisation, réseaux programmés et automatisés.

Source : Gouvernement

Les applications dans le secteur financier, en particulier **les banques et les assurances** sont nombreuses. La **moitié du volume des transactions financières et 90 % des ordres résultent de l'activité d'algorithmes**. Du tiers des échanges boursiers en Europe en 2010, ce taux a dépassé les 90 % depuis 2012. Le sujet du « *high frequency trading* » (HFT) ou trading à haute fréquence (THF) constitue un questionnement en soi, sur lequel la commission des Finances du Sénat a déjà commencé à travailler<sup>1</sup>. Depuis 2012, IBM obtient un brevet l'autorisant à procéder à une estimation de la volatilité des transactions à haute fréquence.

### L'intelligence artificielle et les technologies financières (« Fintech »)

Le renouveau de l'intelligence artificielle est permis par une accélération spectaculaire des investissements, notamment de la part des grands acteurs industriels et du capital-risque, ainsi que par les progrès conséquents des performances d'intelligence artificielle

<sup>1</sup> Cf. Par exemple son rapport n° 369 (2011-2012) sur la proposition de résolution sur la régulation des marchés financiers. Cette proposition, devenue une résolution du Sénat le 21 février 2012, estime « nécessaire de renforcer l'encadrement des pratiques mettant en péril l'intégrité des marchés financiers et notamment les transactions sur base d'algorithmes (trading algorithmique ou trading haute fréquence) », cf. <https://www.senat.fr/leg/tas11-079.html>



visibles, notamment, dans le développement de la reconnaissance d'images, de la parole et de la traduction.

Les technologies d'intelligence artificielle sont souvent produites à l'extérieur des entreprises : la compétence primordiale que doivent acquérir les entreprises est d'intégrer le flux permanent de ces technologies quand elles n'en sont pas directement productrices. Plus précisément, les technologies financières (*Fintech*) permettent par exemple d'utiliser l'intelligence artificielle pour des applications d'interaction avec les clients, de tri dans la proposition de contrats et de détection de fraude dans le traitement des demandes.

L'intelligence artificielle développée par de nombreuses entreprises émergentes est tournée vers le client. Ces systèmes d'intelligence artificielle visent à toucher toutes les difficultés relationnelles que peuvent avoir les entreprises. Les technologies modernes de *chatbot* sont mises au service du client. Les algorithmes peuvent être connus, car souvent en *Open source*, cependant, c'est le savoir-faire des ingénieurs et des développeurs qui leur donnent leur complexité.

Source : Intervention de M. Yves Caseau, animateur du groupe de travail sur l'IA de l'Académie des Technologies, lors de la journée « Entreprises françaises et intelligence artificielle » organisée par le MEDEF et l'AFIA le 23 janvier 2017

## Les « Fintechs » et l'IA

**Exemples**

Lemonade  
wealthfront  
CAPE ANALYTICS  
Praedicat  
Shift technology

**Complexité de l'IA**

- Outils open-source ou services GAFIM
- Algorithmes connus de *data science*
- Savoir-faire : recettes d'intégration

**Caractéristiques**

- Focus sur l'histoire racontée au client
- Automatisation pour rendre du temps utile au client
- Assistance pour faciliter l'usage

**Points forts à imiter**

- Capacité d'apprentissage à partir de la donnée client, en cycle itératif (ne pas sous-estimer un départ lent)
- Utilisation des piles logicielles modernes (cloud), prêt pour le passage à l'échelle

Source : groupe de travail sur l'IA de l'Académie des Technologies animé par Yves Caseau

Concernant le secteur **automobile**, le logiciel n'a, pendant longtemps, rien représenté dans la valeur d'un véhicule, il s'agissait il y a dix ans de 3 %, aujourd'hui il s'agit de 10 % et il pourrait s'agir de l'ordre de 15



ou 30 % demain. Tesla annonce qu'il produira 500 000 voitures autonomes d'ici 2020. Les **constructeurs français** se sont eux-aussi engagés dans la voiture autonome, à l'image de **PSA** qui a présenté son prototype « Peugeot Instinct » au salon de Genève en mars 2017. **Renault** accompagne ce développement d'une démarche éthique intéressante sur la protection des données selon une approche *Privacy by design* (respect de la vie privée dès le stade de la conception) et a élaboré un pack de conformité avec la CNIL. L'entreprise a ainsi joué un rôle moteur dans la charte de constructeurs élaborée au sein de l'association des constructeurs européens d'automobiles. On peut relever qu'Uber dispose d'un centre de recherche à Pittsburgh, fondé en partenariat avec l'université Carnegie-Mellon, dédié aux véhicules sans conducteurs. Quatre prototypes y sont testés par Uber depuis septembre 2016. Les quinze employés de Geometrics Intelligence, racheté par Uber, qui sont des universitaires et des scientifiques, se trouvent désormais affectés à ces recherches.

### L'intelligence artificielle, levier de progrès pour l'industrie automobile

L'industrie automobile est mobilisée pour réussir trois révolutions, chacune d'elles suffisant pour transformer en profondeur l'industrie automobile : **la voiture électrique** (ce qui impliquera des réseaux électriques intelligents ou *smart grids*), **la voiture connectée** (l'intelligence artificielle permet ici de gérer et d'exploiter les données, ou de mettre à disposition des consommateurs des assistants virtuels) **et la voiture autonome**. Ces trois révolutions sont concomitantes et doivent être gérées par l'industrie. Elles nécessitent que l'industrie automobile maîtrise des technologies qui ne la concernaient pas jusqu'il y a peu.

Pour être davantage autonome, la voiture développe son intelligence. L'aide à la conduite et l'autonomisation de la conduite d'un véhicule suivent trois étapes :

- Percevoir l'environnement grâce à des systèmes de caméra, de radar et de technologies à ultra-son.
- Analyser et décider, le véhicule étant doté d'un calculateur et du traitement de l'image et du son. L'objectif est de fusionner l'image et le son captés par le radar et de permettre au système de réagir à une situation de manière adéquate.
- Prévenir et/ou agir dans l'utilisation de la direction, des freins et du moteur, cette utilisation étant couplée avec une interaction homme-machine.

Ces étapes sont celles qui guident la conduite humaine.

Percevoir l'environnement est complexe et dépend de la photographie renvoyée par ce que perçoit une caméra automobile. Il faut assurer un équilibre entre la performance du système et la robustesse de ce système. D'un point de vue de l'analyse d'image, la perception est complexe pour un système autonome.

Il existe une véritable rupture entre une aide à la conduite dans laquelle le conducteur reste maître et une voiture autonome où la conduite est déléguée. Cette rupture technologique implique la fabrication de très nombreux scénarios. L'intelligence artificielle permet de gérer des scénarios très diversifiés et d'entrer dans une phase d'apprentissage.

L'intelligence artificielle contribuera à la transformation de l'industrie automobile. L'ensemble des constructeurs investissent dans l'intelligence artificielle, soit en créant leurs propres laboratoires de recherche, soit en établissant des contrats avec des laboratoires existants. L'intégration de l'intelligence artificielle dans l'industrie automobile

impliquera un changement des modèles d'affaires des entreprises du secteur. **Le caractère crucial de l'intelligence artificielle pour le secteur automobile réside dans le fait que sa maîtrise permettra de développer des applications d'un bout à l'autre de la chaîne de production**, de la conception des logiciels et des véhicules jusqu'au service après-vente. La maîtrise de l'intelligence artificielle représente donc un enjeu essentiel pour les constructeurs automobiles.

*Source : intervention de Patrick Bastard, directeur de l'ingénierie et des technologies électroniques chez Renault, lors de la journée « Entreprises françaises et intelligence artificielle » organisée par le MEDEF et l'AFIA le 23 janvier 2017*

En lien avec l'intelligence artificielle, la robotique va connaître un essor inédit. Selon certaines prévisions récentes, le marché de la robotique devrait ainsi atteindre **35 millions de robots vendus d'ici 2018**<sup>1</sup>, sachant qu'environ 1,5 million sont aujourd'hui en fonctionnement. Le syndicat de la robotique (SYROBO) établit des prévisions du même ordre, puisqu'il estime que 31 millions de robots - industriels et personnels - pourraient être vendus dans le monde entre 2014 et 2017. La croissance mondiale de ce marché serait d'environ 10 % par an en moyenne sur dix ans à partir de 2016, selon le Boston Consulting Group (BCG), au lieu de 2 % par an jusqu'en 2014. La seule **robotique de service** représente à elle seule un énorme gisement de croissance : le marché est estimé à environ 100 milliards d'euros à l'horizon 2020 par la Commission européenne, soit une multiplication par 30 en dix ans. La **France se place numéro 3 mondial dans la recherche fondamentale en robotique** derrière les USA et le Japon, ce qui témoigne d'un avantage comparatif à consolider. **Nao, Pepper ou Romeo** ont été conçus par Aldebaran, entreprise française basée à Issy-les-Moulineaux rachetée en 2012 par SoftBank Robotics, désormais leader mondial de la robotique humanoïde. **Buddy** est un autre robot de service créé par l'entreprise française Bluefrog et dont la commercialisation en 2017 a été annoncée l'année dernière pour moins de 1 000 euros. La société française Robosoft a annoncé en 2016 le lancement de la seconde version de son robot **Kompaï**, conçu pour assister les personnes âgées au quotidien. D'autres robots existent sur le marché mondial et peuvent être mentionnés comme **Paro, Jibo, Asimo, Amazon Echo, Otto, Floka...**

Les capacités des **agents conversationnels** dit *chatbots* ou même *bots* sont, il est vrai, encore limitées, mais ces derniers vont rendre **de plus en plus de services à leurs utilisateurs**. Le cabinet d'études Forrester estime, d'ailleurs, que les *bots* ne sont pas encore à la hauteur des attentes des usagers. Laurence Devillers explique ainsi qu'ils n'ont pas de mémoire, qu'ils se contentent de suivre des scénarios de questions-réponses et qu'ils ne savent pas répondre aux utilisateurs qui se plaignent de dépression ou de maladies physiques, mais pour vos rapporteurs la réalité est en fait plus nuancée puisqu'ils ont pu constater que les *bots* font déjà un excellent travail en matière de **prévention du suicide** ainsi que l'a expliqué à vos rapporteurs Alex Acero, directeur du projet Siri chez Apple. Siri reçoit par exemple 5 000

<sup>1</sup> Cf. [www.worldrobotics.org/uploads/tx\\_zeifr/Executive\\_Summary\\_WR\\_2015.pdf](http://www.worldrobotics.org/uploads/tx_zeifr/Executive_Summary_WR_2015.pdf)

---

propos suicidaires par jour qu'il traite en rassurant le propriétaire du téléphone et en orientant vers des services spécialisés. Cortana de Microsoft gère ce type de conversations avec la même efficacité. Vos rapporteurs ont pu tester en situation ces usages bénéfiques pour la société des agents conversationnels.

En matière d'**éducation**, les perspectives pour l'intelligence artificielle sont riches mais les applications restent encore rares.

Dans le secteur des **loisirs**, tels que les jeux vidéos ou le cinéma, l'intelligence artificielle est utilisée assez massivement. Il peut par exemple s'agir de simuler des foules grâce à des systèmes multi-agents, comme dans les trilogies *Le Seigneur des anneaux* et *Le Hobbit*.

Les secteurs de **l'énergie** et de **l'environnement** commencent à recourir à des solutions fondées sur l'intelligence artificielle. Les **compteurs intelligents** sont une des pistes visibles de cette évolution en cours, sur laquelle travaille l'OPECST.

Citons encore **l'agriculture**, où les possibilités sont également nombreuses. Les applications concernent la gestion des exploitations mais vont bien au-delà. Il peut être donné l'exemple d'une entreprise qui a ainsi mis au point un système de conteneurs intelligents, en partenariat avec le *MIT Media Lab*. À l'intérieur, l'intelligence artificielle contrôle la lumière, l'humidité, la température, mais aussi les nutriments apportés aux plantes, supervisant leur croissance en temps réel. Cela permet d'améliorer l'efficacité du cycle de production sans avoir recours aux OGM.

Dans le secteur de la **défense**, les drones autonomes sont de plus en plus utilisés, à l'image de l'expérience des armées australiennes, israéliennes ou, encore, saoudiennes. Ce n'est pas le secteur où le développement de l'intelligence artificielle est le plus souhaitable.

**La sécurité** est d'ores et déjà améliorée avec une intelligence artificielle qui peut détecter les situations anormales (par exemple sur les flux vidéo des caméras de surveillance) et alerter les services compétents.

Dans **l'assistance au diagnostic** ou dans les **services de maintenance prédictive** dans l'industrie et l'électroménager, l'intelligence artificielle optimise et détecte les défaillances en amont.

En matière de **sécurité informatique**, ce sont les fraudes et les cyberattaques qui peuvent être prévues et gérées de manière plus efficace. **La cybersécurité** peut être révolutionnée par l'intelligence artificielle.

Les usages de l'intelligence artificielle en matière de **technologies médicales**, de gestion de la **dépendance** ou de **handicap** seront **considérables** mais ils n'ont pas été au cœur du travail de vos rapporteurs, notamment car le sujet a déjà été traité à plusieurs reprises par l'OPECST et qu'il continuera à l'être, autour de rapports spécifiques. Le dernier exemple remonte à 2015 avec un rapport consacré au thème « *Le numérique au service*

---

de la santé »<sup>1</sup>. Il est certain que l'intelligence artificielle est et sera de plus en plus utile à la médecine, notamment et y compris à court terme en matière de diagnostic et de dépistage des maladies. Les systèmes recourant à l'IA font de plus en plus souvent aussi bien, voire mieux, que les médecins dans le dépistage du cancer<sup>2</sup>.

Quelques cas peuvent être mentionnés. En **génomique**, par exemple pour la validation et la critique des thérapies, outre le *deep learning*, on peut utiliser des systèmes non logiques ou partiellement logiques, des réseaux bayésiens, des systèmes de règles et d'arbres de décision, des systèmes experts... En matière de **prédiction du repliage de protéines** ou de **segmentation des IRM du cerveau** en vue d'identifier certaines zones, là aussi des projets de recherche sont menés avec l'utilisation de l'apprentissage automatique (*machine learning*) et de systèmes multi-agents collaboratifs pour découvrir les règles qui régulent la géométrie spatiale de structures complexes (exemple d'un projet associant l'Université de Grenoble, l'Inria et l'INSERM). L'utilisation de systèmes multi-agents est également possible pour analyser les courbes de réponse d'assistants respiratoires et détecter les anomalies.

Le Dr Lionel Jouffe, président de Bayesia, s'est spécialisé sur **l'utilisation des réseaux bayésiens pour l'aide à la décision médicale**. Il s'agit d'une modélisation des connaissances par apprentissage automatique à partir des données. Le réseau bayésien peut en effet être utilisé pour des applications de différentes natures. Il rappelle qu'un réseau bayésien peut ainsi permettre à une entreprise de calculer la probabilité qu'un client soit intéressé par un produit, à une banque ou un *business angel* de calculer la probabilité de faillite d'une *start-up*, tout comme il peut être appliqué à la thérapie endovasculaire. Ainsi, les applications diverses d'un réseau bayésien lui permettent d'être utilisé pour de nombreuses tâches, comme l'analyse de leviers d'optimisation, le calcul de scores, l'analyse de défauts par l'optimisation de processus, l'analyse opérationnelle, le diagnostic et le dépannage ou encore l'analyse de risques et la maintenance préventive.

L'**exemple de Watson** est également instructif, pour le secteur médical et l'aide au diagnostic mais même au-delà. Dévoilé au grand public par IBM en 2011, ce système a affronté, avec succès, des candidats humains au jeu télévisé américain « Jeopardy ! ». En 1996 et 1997, IBM avait déjà prouvé les capacités de son superordinateur Deep Blue en organisant des parties d'échecs contre Garry Kasparov. Nicolas Sekkaki, responsable d'IBM France, assure que sa société est aujourd'hui engagée dans **une dizaine de projets faisant appel à Watson sur notre territoire**, mais les retours d'expérience dignes de ce nom sur le sujet restent encore peu nombreux. Le Crédit Mutuel teste avec IBM l'utilisation de l'intelligence artificielle et des

---

<sup>1</sup> Rapport Sénat n° 465 (2014-2015) de Mme Catherine PROCACCIA, sénateur et M. Gérard BAPT, député.

<sup>2</sup> Un exemple récent concerne le dépistage du cancer du sein : <http://www.numerama.com/sciences/176579-une-ia-sait-detecter-le-cancer-du-sein-presque-aussi-bien-quun-medecin.html>

---

technologies cognitives depuis juin 2015 et a intégré certaines technologies dans la gestion de sa **relation client** depuis 2016. Watson est ainsi utilisé pour l'assistance des conseillers dans le traitement des courriels d'une part, et sur les produits d'assurance et d'épargne d'autre part. Une assistance informatisée qui vise à optimiser la productivité du conseiller et améliorer la pertinence des réponses fournies aux clients finaux. Pour l'instant, il ne s'agirait pas de laisser l'intelligence artificielle interagir directement avec le client.

Les progrès dans les domaines de l'intelligence artificielle et de la robotique, en matière de **vision par ordinateur**, de **traitement automatique du langage naturel**, de **reconnaissance automatique de la parole**, ou, encore de **bio-informatique**, à travers par exemple l'étude de l'ADN, sont encore toute une série de perspectives d'**applications fécondes**.

Vos rapporteurs relèvent **une part d'effet de mode dans l'écosystème entrepreneurial**, visible dans le recours à certains concepts, tels que l'intelligence artificielle, le *big data*, le *cloud*, l'*IoT* (Internet des objets), le *blockchain*. Pour le journaliste Olivier Ezratty, le stéréotype de la *start-up* en intelligence artificielle serait, de manière caricaturale, une « solution d'agent conversationnel en *cloud* faisant du *big data* sur des données issues de l'*IoT* en sécurisant les transactions via des *blockchains* ».

Les entreprises de l'intelligence artificielle se diversifient, se reconfigurent et s'absorbent les unes les autres, et d'autres entreprises, issues d'autres secteurs, parfois plus traditionnels, tentent de les rejoindre dans une **course propre à l'économie des plateformes**, que vos rapporteurs décriront plus loin.

#### **4. Par leurs combinaisons en évolution constante, ces technologies offrent un immense potentiel et ouvrent un espace d'opportunités transversal inédit**

Le **potentiel de ces technologies est immense** et ouvre de manière transversale un **espace d'opportunités inédit** : nos économies peuvent en bénéficier car les champs d'application sont et seront de plus en plus nombreux. Ces technologies sont non seulement en évolution constante, mais leurs **combinaisons ouvrent de nouvelles perspectives**.

Avec **l'explosion des données massives ou *big data* et l'augmentation des vitesses de calcul** (vue plus haut avec la loi de Moore), **ces techniques d'intelligence artificielle deviennent de plus en plus puissantes et efficaces**, grâce aux combinaisons de compétences et de technologies en particulier.

**Les combinaisons et les hybridations** entre technologies mises au point par Google Deep Mind vont dans ce sens, en utilisant tant des outils traditionnels comme la méthode Monte-Carlo que des systèmes plus récents comme l'apprentissage profond. L'entreprise fait figure de structure à l'avant-garde de la recherche mondiale en intelligence artificielle. Les

---

combinaisons de technologies d'intelligence artificielle ouvrent un champ de recherche fécond et elle en a fait sa spécialité. Le programme AlphaGo a ainsi appris à jouer au jeu de Go par une **méthode de *deep learning*** couplée à un **apprentissage par renforcement** et à une optimisation selon la **méthode Monte-Carlo**.

**Dans les faits**, et comme l'illustrent les cas déjà évoqués de Prolog<sup>1</sup>, des réseaux de neurones profonds<sup>2</sup> ou du programme AlphaGo<sup>3</sup>, **l'intelligence artificielle combine très souvent plusieurs techniques**.

De plus en plus, **les outils d'intelligence artificielle sont systématiquement utilisés conjointement**.

Par exemple, les **systèmes experts** sont utilisés avec le **raisonnement par analogie**, éventuellement dans le cadre de **systèmes multi-agents**.

De même, les **SVM** et **l'apprentissage par renforcement** se combinent très efficacement avec **l'apprentissage profond des réseaux de neurones**<sup>4</sup>. Ce dernier, le *deep learning*, peut aussi s'enrichir de logiques floues ou d'algorithmes génétiques et trouve de nombreuses applications dans le domaine de la reconnaissance de formes (lecture de caractères, reconnaissance de signatures, de visages, vérification de billets de banque), du contrôle de processus et de prédiction.

Selon Stéphane Mallat, professeur à l'École normale supérieure, le *deep learning* représente en tout cas « *une rupture non seulement technologique, mais aussi scientifique : c'est un changement de paradigme pour la science* ». Traditionnellement, les modèles sont construits par les chercheurs eux-mêmes à partir de données d'observation, en n'utilisant guère plus de dix variables alors que « *les algorithmes d'apprentissage sélectionnent seuls le modèle optimal pour décrire un phénomène à partir d'une masse de données* » et avec une complexité inatteignable pour nos cerveaux humains, puisque cela peut représenter jusqu'à **plusieurs millions de variables**. Alors que le principe de base de la méthode scientifique réside dans le fait que les modèles ou les théories sont classiquement construits par les chercheurs à partir des observations, le *deep learning* change la donne en assistant l'expertise scientifique dans la construction des modèles. Stéphane Mallat remarque également que la physique fondamentale et la médecine (vision, audition) voient converger leurs modèles algorithmiques.

Denis Girou, directeur de l'Institut du développement et des ressources en informatique scientifique au CNRS estime que « *la science a pu construire des modèles de plus en plus complexes grâce à l'augmentation de la*

---

<sup>1</sup> Les raisonnements formels de Prolog ont été enrichis de la méthode de programmation par contraintes.

<sup>2</sup> Les réseaux de neurones artificiels sont dans ce cas couplés aux méthodes d'apprentissage profond.

<sup>3</sup> Ce programme a appris à jouer au jeu de Go en combinant apprentissage profond et apprentissage par renforcement.

<sup>4</sup> L'efficacité est avérée pour le traitement automatique du langage naturel, la reconnaissance automatique de la parole, la reconnaissance audio, la bio-informatique ou, encore, la vision par ordinateur.

---

*puissance de calcul des outils informatiques, au point que la simulation numérique est désormais considérée comme le troisième pilier de la science après la théorie et l'expérience* ». En sciences du climat par exemple, l'approche traditionnelle qui consiste à injecter les mesures issues de capteurs en tant que conditions initiales des simulations s'est enrichie : les approches *big data* avec le *machine learning* et l'analyse statistique des données ouvrent ainsi une nouvelle voie : « *ce qu'on appelle « climate analytics » a permis aux climatologues de découvrir, grâce au travail de statisticiens, de nouvelles informations dans leurs données* ». Il s'agit d'outils sur lesquels s'appuie notamment le Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC) dans ses prédictions sur le réchauffement climatique.

Vos rapporteurs appellent à la vigilance face à **l'illusion du « jamais vu »**, il faut en effet relativiser la nouveauté de l'aide apportée par l'intelligence artificielle, la découverte d'autres outils complexes ayant jalonné l'histoire des civilisations humaines. Dans un texte intitulé « *L'ordinateur et l'intelligence* »<sup>1</sup>, l'économiste Michel Volle rappelle ainsi que « *des machines remplacent nos jambes (bateau, bicyclette, automobile, avion), des prothèses assistent nos sens (lunettes, appareils acoustiques, téléphones, télévision). L'élevage et l'agriculture pratiquent la manipulation génétique, depuis le néolithique, par la sélection des espèces. La bionique, l'intelligence artificielle ne font que s'ajouter aujourd'hui au catalogue des prothèses qui assistent nos activités physiques ou mentales* ».

Toutefois, quand bien même **l'illusion du jamais vu doit être dénoncée**, il convient d'**éviter aussi l'écueil du toujours ainsi**. L'intelligence artificielle représente une série d'outils à l'autonomie croissante, qui offre de nouvelles opportunités et qui pose de nombreuses questions.

## **5. L'apprentissage automatique reste encore largement supervisé et fait face au défi de l'apprentissage non-supervisé**

**Selon Yann LeCun, le défi scientifique auquel les chercheurs doivent s'atteler**, au-delà de la redécouverte de ces deux techniques, **c'est celui de l'apprentissage non supervisé** alors que l'apprentissage machine reste le plus souvent supervisé : on apprend aux ordinateurs à reconnaître l'image d'une voiture en leur montrant des milliers d'images et en les corrigeant quand ils font des erreurs. Or les humains découvrent le monde de façon non supervisée.

Dans sa leçon inaugurale au Collège de France, Yann LeCun estime ainsi que « *tant que le problème de l'apprentissage non-supervisé ne sera pas résolu, nous n'aurons pas de machines vraiment intelligentes. C'est une question fondamentale scientifique et mathématique, pas une question de technologie. Résoudre ce problème pourra prendre de nombreuses années ou plusieurs décennies. À la vérité, nous n'en savons rien* ». Selon lui, cette technologie, qui peut

---

<sup>1</sup> Cf. <http://www.volle.com/ulb/021116/textes/intelligence.htm>



---

prendre la forme de l'apprentissage prédictif, devrait permettre aux machines d'acquiescer ce que l'on appelle le sens commun.

L'apprentissage non-supervisé permettra de **faire progresser les algorithmes sans le coût lié à l'étiquetage et à la supervision humaine de l'apprentissage**. C'est à l'évidence un **défi scientifique**, mais vos rapporteurs notent qu'il n'est pas sûr que l'on y parvienne et que les moyens mis en œuvre doivent rester proportionnés, surtout qu'il n'est absolument pas sûr que l'on puisse parvenir à l'apprentissage non-supervisé.

Ils relèvent que les **travaux en robotique développementale et sociale** de Pierre-Yves Oudeyer, responsable du laboratoire Flowers de l'Inria, sont particulièrement féconds : il s'agit, en faisant appel à de nouvelles disciplines connexes (neurosciences et psychologie développementale), de concevoir des robots capables d'apprendre des choses nouvelles sur le long terme sans l'intervention d'un ingénieur, en combinant curiosité artificielle et interactions sociales avec des humains, selon la maxime « *humaniser les machines plutôt que machiniser les hommes* » et en visant à reproduire les comportements d'apprentissage des enfants. Vos rapporteurs ont noté l'influence des **travaux de Jean Piaget** dans ces recherches.

En conclusion de cette partie, vos rapporteurs prennent acte des **limites des technologies actuelles d'intelligence artificielle** et font valoir que l'intelligence artificielle, qui agit sur la base de ce qu'elle sait, devra relever le défi d'agir sans savoir, puisque comme l'affirmait le biologiste, psychologue et épistémologue Jean Piaget « *L'intelligence, ça n'est pas ce que l'on sait, mais ce que l'on fait quand on ne sait pas* ».

### III. CARACTÉRISTIQUES GÉNÉRALES DE LA RECHERCHE EN INTELLIGENCE ARTIFICIELLE ET ORGANISATION NATIONALE EN LA MATIÈRE

#### A. LES CARACTÉRISTIQUES DE LA RECHERCHE EN INTELLIGENCE ARTIFICIELLE

##### 1. La place prépondérante de la recherche privée, dominée par les entreprises américaines et, potentiellement, chinoises

La **place prépondérante de la recherche privée** a été **particulièrement ressentie par vos rapporteurs, y compris sur le plan de la recherche fondamentale**. Cette recherche est, de plus, **dominée par les entreprises américaines** et pourrait, potentiellement, être **dominée demain par les entreprises chinoises**.

Ainsi plusieurs enseignants-chercheurs, souvent les plus brillants, ont été recrutés par ces grandes entreprises : **Yann LeCun** (Facebook),



---

**Andrew Ng** (Baidu, ex-Google), **Geoffrey Hinton** (Google), **Fei Fei Li** (Google), **Rob Fergus** (Facebook), **Nando de Freitas** (Google), même **Yoshua Bengio** qui a longtemps refusé d'être recruté dans le secteur privé a créé la plateforme « Element AI » avec un investissement de Microsoft et a rejoint Intel à la fin de l'année 2016 au Conseil d'administration de Intel Nervana AI.

En 2010, **Demis Hassabis** a inventé, avec Mustafa Suleyman et Shane Legg, une *start-up* londonienne qui se donnait pour objectif de faire ce qui se fait de mieux en intelligence artificielle, de « résoudre l'intelligence ». Il a rapidement vu sa structure, Deepmind, être rachetée pour 628 millions de dollars par Google (en 2014).

Vos rapporteurs ont identifié plusieurs cas de chercheurs français recrutés par ces entreprises américaines, dites « GAFA » ou « GAFAMI », même s'il serait plus juste de parler des « GAFAMITIS »<sup>1</sup>, ou par leurs équivalentes chinoises dites « BATX »<sup>2</sup>. Il est à noter que certains de ces chercheurs ont refusé de rencontrer vos rapporteurs.

Laurent Massoulié, directeur du Centre de recherche commun INRIA-Microsoft explique que « *les frontières public/privé sont de plus en plus perméables* ». L'enjeu affiché est de réunir le meilleur des deux mondes et de favoriser la mobilité des chercheurs, mais il faut observer qu'il concourt aussi à la concentration des compétences au sein des entreprises privées américaines.

Comme il a été vu les technologies d'apprentissage machine tel que le *deep learning* recourent à des méthodes plus ou moins statistiques qui nécessitent des données massives pour être efficaces, or ces entreprises disposent d'un avantage comparatif difficile à rattraper : des jeux de données massives.

Les entreprises américaines font de plus en plus de recherche fondamentale en intelligence artificielle, ouvrent une partie de leurs résultats et communiquent sur leurs travaux et leur fonctionnement. C'est en particulier visible pour Google, Facebook, IBM ou Baidu. Même Apple, longtemps très discrète, a décidé, en 2016, d'offrir à ses chercheurs la possibilité de publier dans des revues scientifiques<sup>3</sup>. Six chercheurs de l'entreprise ont ainsi publié un article sur l'apprentissage machine à partir d'images de synthèse.

---

<sup>1</sup> Google, Apple, Facebook, Amazon, Microsoft, IBM, Twitter, Intel et Salesforce. Ces entreprises américaines représentent la pointe de la recherche et des applications de l'IA.

<sup>2</sup> L'expression désigne les géants chinois du numérique : Baidu, Alibaba, Tencent et Xiaomi.

<sup>3</sup> Cette évolution doit beaucoup à Russ Salakhutdinov, directeur de l'intelligence artificielle chez Apple et professeur à l'université Carnegie-Mellon, rencontré par vos rapporteurs. Cette communication de plus en plus scientifique d'Apple est visible aussi dans cet article qui met notamment à l'honneur Alex Acero, directeur du projet Siri, également rencontré par vos rapporteurs : <https://backchannel.com/an-exclusive-look-at-how-ai-and-machine-learning-work-at-apple-8dbfb131932b#.6oxrqyd5g>

---

De plus, certains **outils d'intelligence artificielle** des entreprises IBM, Microsoft, Google et Facebook ont été **rendus publics en « open-source »** en vue d'encourager la constitution de communautés de développeurs.

Ces entreprises entretiennent ainsi un formidable vivier d'experts et de chercheurs, aux États-Unis mais aussi dans le reste du monde, qui leur permet de perfectionner leurs algorithmes à moindre coût.

Les entreprises américaines dominent donc pour l'heure, et certains parlent de « silicolonisation » du monde ou de l'Europe, mais les **entreprises chinoises**, qui se sont longtemps contentées d'exploiter et de répliquer des technologies élaborées à l'extérieur de leurs frontières peuvent potentiellement monter en puissance. La recherche chinoise indique d'ailleurs cette tendance. La Chine a ainsi pris la tête des publications en *deep learning* depuis trois ans.

Même si **les progrès visibles reposent encore sur des architectures conçues initialement par des scientifiques occidentaux**, les atouts chinois sont réels, comme l'a indiqué à vos rapporteurs le service scientifique de l'Ambassade de France en Chine : elle dispose en effet *« des deux supercalculateurs les plus puissants du monde, d'un marché intérieur très important et friand des avancées potentielles du secteur, d'une collusion féconde entre État, instituts de recherche, universités, géants de l'internet et de l'informatique, start-ups »*. Ainsi, le 13<sup>ème</sup> plan quinquennal chinois comprend une liste de 15 « nouveaux grands projets – innovation 2030 » qui structurent les priorités scientifiques du pays et correspondent chacun à des investissements de plusieurs milliards d'euros. Parmi ces 15 projets, on en trouve quatre qui sont dédiés indirectement à l'IA, pour un montant de **100 milliards de yuans en trois ans** : un projet de « Recherche sur le cerveau » et des projets d'ingénierie intitulés « Mega données », « Réseaux intelligents » et « Fabrication intelligente et robotique ».

Ce **plan** a été décidé par les autorités chinoises dans le but de **dynamiser la recherche en IA en Chine et de relever le défi de la concurrence avec les États-Unis**. Les investissements seront orientés sur la robotique, les assistants personnels (domotique), les voitures autonomes et les objets connectés.

Les enjeux sont nationaux et les financements suivent. Mais **les entreprises privées sont aussi des moteurs puissants du secteur**.

On voit dans la période récente **les entreprises chinoises monter en puissance, avec les « BATX »**<sup>1</sup>.

L'entreprise **Baidu** a développé le **principal moteur de recherche chinois**, site le plus consulté en Chine et 5<sup>ème</sup> site le plus consulté au niveau mondial, indexant près d'un milliard de pages, 100 millions d'images et 10 millions de fichiers multimédia. Elle communique beaucoup sur le sujet

---

<sup>1</sup> Baidu, Alibaba, Tencent et Xiaomi.

---

de l'IA, y consacre une part conséquente de sa recherche (*Institute of deep learning, Big data lab, ...*) et, comme les géants américains, dispose d'un flux de données permettant d'envisager des applications dans de nombreux domaines.

**Le recrutement du chercheur de Stanford Andrew Ng par Baidu en 2014 en tant que responsable de l'intelligence artificielle alors qu'il était le responsable de l'intelligence artificielle chez Google** est emblématique.

De même, en 2017, **Baidu débauche Qi Lu**, au poste de numéro 2, alors qu'il était **auparavant vice-président chez Microsoft** et directeur des projets Bing, Skype et Microsoft Office et, auparavant, directeur de la recherche de Yahoo.

L'entreprise considère, comme ses concurrents, que **l'IA est son principal défi** comme solution clé pour des applications en vision, parole, traitement du langage naturel et sa compréhension, génération de prédictions et de recommandations, publicité ciblée, planification et prise de décision en robotique, la conduite autonome, pilotage de drones,... Elle travaille en étroite relation avec de nombreuses universités et *start-ups*.

**Les résultats algorithmiques de Baidu et de son *Institute of Deep Learning* sont impressionnants et du meilleur niveau mondial**, malgré son existence récente. Il n'est pas évident selon le service scientifique de l'Ambassade de France en Chine d'évaluer la maîtrise théorique des ingénieurs mais l'entreprise montre une incontestable efficacité pour implémenter rapidement les dernières innovations du secteur.

Le système de reconnaissance d'image de **Baidu a ainsi battu celui de Google depuis 2015**. Son logiciel a décrit 100 000 images avec une précision de 95,42 % contre 95,20 % pour celui de Google. Baidu représente à lui seul plus de 80 % du marché chinois de la recherche en ligne (contre 9 % pour Google). La capacité de leurs algorithmes à retrouver une image dans une base de données de 10 milliards d'images est de moins d'une seconde. Ils ont aussi de bonnes performances sur les *benchmarks* ICDAR, où ils se placent 1<sup>er</sup> sur 5 des 8 évaluations conduites parmi 4 tâches.

Sur FDDB<sup>1</sup> (*Face Detection Data Set and Benchmark*) et sur la base de données de visages LFW, ils progressent vite : 8 % d'erreurs en décembre 2015, 2,3 % en septembre 2016 et bientôt 1 %. L'entreprise annonce aussi la meilleure précision sur la collection de *benchmarks* KITTI<sup>2</sup> orientés pour la conduite de voitures autonomes. Baidu développe aussi des applications de reconnaissance d'image pour la plateforme de services *Baidu Nuomi* : une application permet par exemple de reconnaître le restaurant (et le plat) en prenant une photo de nourriture dans un restaurant.

Les autres géants chinois du net, comme **Alibaba** (distribution) ou **Tencent** (réseaux sociaux), tirent eux aussi dans la même direction : développement et diffusion grand public d'applications plus ou moins

---

<sup>1</sup> Voir <http://vis-www.cs.umass.edu/fddb/>

<sup>2</sup> Voir <http://www.cvlibs.net/datasets/kitti/>

---

convaincantes mais manifestement exploitant des techniques d'IA un peu évoluées, même s'ils semblent moins présents dans la recherche « fondamentale ».

Créé le 6 avril 2010, Xiaomi a rapidement rejoint le club des géants chinois, en devenant l'un des plus gros constructeurs mondiaux dans les TIC (smartphones, tablettes, mais aussi bracelets connectés, télévisions intelligentes, équipements pour les maisons connectées, caméras miniatures...) sans même devoir s'installer en Occident, devenant le quatrième membre des BATX. Seul Uber représente un dynamisme comparable. **Le recrutement en 2013 de Hugo Barra par Xiaomi** en tant que vice-président est significatif. D'abord chercheur du laboratoire d'intelligence artificielle et d'informatique du MIT, laboratoire visité par vos rapporteurs, il est ensuite devenu responsable du développement du système d'exploitation Android chez Google entre 2010 et 2013. En janvier 2017, il a rejoint Facebook, en tant que vice-président de la division Oculus en charge de la réalité virtuelle. Vos rapporteurs ont été reçus au siège de Facebook à Menlo Park par les responsables de la recherche en IA de Facebook et par différents responsables de sa division Oculus le 26 janvier 2017, au moment de ce transfert de Hugo Barra de Xiaomi vers Facebook.

**Huawei** (telecom, téléphones), entreprise qui accorde une grande importance à la recherche fondamentale et dont la recherche s'internationalise rapidement (pôle mathématique implanté en France il y a deux ans) a mis en avant début janvier 2017 un concept de téléphone intelligent, dont on ne peut encore savoir s'il ira effectivement plus loin que ceux de ses concurrents.

Plus globalement concernant les données, les **pôles universitaires peuvent aussi compter sur le soutien des industriels** comme, par exemple, *National Grid, China Mobile, China Unicom, Shanghai Meteorological Bureau* ou *Environmental Monitoring Center...*

La Chine prend aussi position dans les technologies de **cartographie numérique**. Deux entreprises nationales, le chinois NavInfo et la société de services Internet Tencent, associées au fonds singapourien GIC, ont acquis ensemble une participation de 10 % dans le groupe HERE, contrôlé par BMW, Daimler et Volkswagen, qui ont réduit leur part en proportion. Dans ce cadre, HERE crée une filiale avec NavInfo pour étendre son offre à la Chine. Et Tencent utilisera de son côté les prestations de cartographie et de localisation de la société pour ses propres produits et services.

## 2. Une recherche essentiellement masculine

Vos rapporteurs dressent le constat évident que **la recherche en intelligence artificielle et en robotique est conduite par des hommes. Cette situation d'extrême masculinisation n'est pas souhaitable.**

Selon Mady Delvaux, la domination masculine dans ce secteur serait de nature à créer des biais, elle dénombre de l'ordre de **90 % de programmeurs et de développeurs**.

### 3. Une interdisciplinarité indispensable mais encore insuffisante

En matière d'intelligence artificielle, **l'interdisciplinarité est particulièrement requise**. En effet, **il s'agit à la fois d'un secteur de recherche en informatique et d'un champ de réflexion bien plus large, qui mobilise des connaissances provenant de nombreuses disciplines**.

L'interdisciplinarité est donc indispensable en intelligence artificielle. Cette prise en compte du critère de l'interdisciplinarité est essentielle mais, elle-même sous-domaine de l'informatique, l'intelligence artificielle demeure éclatée en une **cinquantaine de sous-domaines de recherche** identifiés.

#### L'éclatement de la discipline en une cinquantaine de domaines

IA et SHS	Représentation des connaissances	apprentissage automatique	traitement du langage naturel	traitement des signaux	robotique*	neurosciences, sciences cognitives	algorithmique de l'IA	aide à la décision	systèmes multi-agents	interaction avec l'humain
Ethique	Bases de connaissances	Apprentissage supervisé / non-supervisé / séquentiel et par renforcement	Analyse syntaxique Lexiques Discours  (Interaction, Connaissances et Langage Naturel)	Parole Vision	Conception Perception	Compréhension et stimulation du cerveau et du système nerveux	Programmation logique et ASP		Coordination Multi-Agents (Planification multi-agents, Décision multi-agents)	Interaction avancée, apprentissage humain (EIAH)
Droit	Extraction et nettoyage de connaissances			Reconnaissance d'objets	Décision	Sciences cognitives	Déduction, preuve		Résolution Distribuée de Problèmes	
Economie	Inférence	Optimisation	Reconnaissance de la parole et traduction automatique	Reconnaissance d'activités	Interactions avec les robots		Théories SAT		Apprentissage multi-agents	
Sociologie	Web sémantique	Méthodes bayésiennes		Recherche dans des banques d'images et de vidéos	Flottes de robots		Raisonnement causal, temporel, incertain		Ingénierie Multi-Agents (Langages, plateformes, méthodologies)	
Humanités numériques	Ontologies	Réseaux de neurones ou neuronaux		Reconstruction 3D et spatio-temporelle	Apprentissage des robots		Programmation par contraintes		Simulation Multi-Agents (intéresse aussi les SHS)	
		Méthodes à noyau		Suivi d'objets et analyse des mouvements	Cognition pour la robotique et les systèmes					
		Apprentissage profond		Localisation d'objets						
		Fouille de données		Asservissement visuel						
		Analyse de données massives					Planification et ordonnancement			

Source : Gouvernement

Non seulement l'interdisciplinarité reste **insuffisamment mise en œuvre** mais l'intelligence artificielle elle-même souffre de ces découpages internes, qui tendent à cloisonner les recherches.

Vos rapporteurs ont ainsi relevé que **les recherches en intelligence artificielle empruntent et devront, de plus en plus, emprunter à diverses autres disciplines et en agréger les connaissances et les démarches**. Elles peuvent même aller jusqu'à s'inscrire dans des espaces disciplinaires différents (mathématiques, physique, biologie...).

**Issue des mathématiques, de la logique et de l'informatique**, l'intelligence artificielle fait, depuis des décennies, de plus en plus appel à la **psychologie, à la linguistique, à la neurobiologie, à la neuropsychologie et au design**. Dans la période plus récente, elle s'ancre encore davantage dans

---

les **sciences cognitives**, et mobilise les outils de la **génétique** et des Sciences de l'homme et de la Société (SHS), en particulier de la **sociologie**.

L'intelligence artificielle se nourrit de plus en plus des recherches issues des mathématiques, des statistiques, de la physique, de la biologie, en particulier pour ses **méthodes et ses champs d'application**. Elle doit également, et de plus en plus, s'alimenter auprès des SHS, plutôt sur les **usages mais aussi sur les questionnements éthiques en matière de conception**.

Les **humanités numériques** sont un exemple de démarches à l'intersection de l'intelligence artificielle et des SHS. Comme l'a expliqué Jean-Gabriel Ganascia à vos rapporteurs, les disciplines relevant des humanités, à savoir dont l'objet d'étude porte sur les œuvres humaines, comme l'histoire, l'archéologie, la littérature, etc., tirent, depuis environ quinze ans, avantage de la numérisation des sources et de l'utilisation des techniques d'intelligence artificielle et d'apprentissage machine pour concevoir de nouveaux opérateurs d'interprétation. On désigne aujourd'hui ce courant scientifique situé à la frontière de l'informatique et des « sciences de la culture » sous le vocable d'humanités numériques. Cela fait l'objet de nombreuses recherches dans tous les pays du monde, en particulier aux États-Unis, dans les États européens, dont la France. On peut, par exemple, citer l'école polytechnique fédérale de Lausanne en Suisse, visitée par vos rapporteurs, ou le Labex OBVIL entre la Sorbonne et l'UPMC.

Par ailleurs, les **liens entre intelligence artificielle et robotique** ont longtemps été très étroits mais dans le dernier quart du XX<sup>e</sup> siècle une **distance s'est créée sous le double effet de la spécialisation de la robotique industrielle sur des automates pas ou peu autonomes et du moindre coût de la recherche sur des systèmes logiciels d'intelligence artificielle** (le coût relatif d'un robot intelligent étant plus élevé). D'après Raja Chatila, cette distance devrait maintenant se réduire car les robots deviennent de plus en plus des intelligences artificielles incorporées (*embodied*), la formule étant plus pertinente qu'incorporées car elles s'incarnent dans des corps physiques plus qu'elles n'y sont placées. Au-delà du traitement du langage naturel, les problèmes d'interaction et d'action conjointe homme-robot posent des questions d'interdisciplinarité selon Raja Chatila : la mise en œuvre de « prises de perspective » est une problématique fondamentale pour permettre une interaction efficace et naturelle entre l'humain et le robot. Ce sujet demande un développement qui associe des recherches en robotique et en Sciences de l'Homme et de la Société (SHS), en particulier, en sociologie, philosophie, psychologie, linguistique... Le **rôle des émotions dans l'interaction** est à explorer, bien au-delà de travaux actuels qui se contentent de classer *a priori* des expressions faciales ou de produire des expressions d'émotions artificielles par le robot. L'expression d'émotions par un robot pose des **questionnements scientifiques et éthiques sur l'authenticité de ces émotions et sur l'anthropomorphisation** qui peut en résulter, sujet sur lequel s'est spécialisée Laurence Devillers, professeure à l'Université Paris IV Sorbonne et directrice de recherche au Laboratoire d'informatique pour la

---

mécanique et les sciences de l'ingénieur (Limsi de Saclay) et qui plaide elle aussi pour une plus grande interdisciplinarité, totalement nécessaire selon elle.

La recherche en robotique pose en effet des questions proches des sciences cognitives, des neurosciences et de plusieurs domaines des SHS, comme la sociologie, la psychologie et la philosophie. Des programmes interdisciplinaires sont probablement le bon moyen d'aborder les différentes facettes des questions fondamentales posées par l'intelligence artificielle et la robotique.

En plus des aspects technologiques, les chercheurs appellent à **plus de transversalité** en vue de croiser leurs approches, indispensable au progrès dans ce domaine où se télescopent les disciplines. Selon Bertrand Braunschweig, directeur du centre de l'Inria de Saclay, « nous avons besoin de doubles, voire de triples-compétences ».

**Comprendre l'intelligence reste une des grandes questions scientifiques de notre temps, au-delà même des progrès et des limites de l'intelligence artificielle** et, comme le souligne Yann LeCun, « aucune organisation, si puissante soit-elle, ne peut résoudre ce problème en isolation. La conception de machines intelligentes nécessitera la collaboration ouverte de la communauté de la recherche entière ».

L'interdisciplinarité doit, en outre, se conjuguer avec une **meilleure prise en compte du long terme**, qui représente un autre défi en matière de recherche scientifique<sup>1</sup>, notamment en matière d'intelligence artificielle. Pour Jean-Gabriel Ganascia, cela résulte notamment des modes de financement (logiques de court terme des projets, donc pas de pérennité et pas de vision de long terme, pas de grandes ambitions possibles), il note que les LabEx, ou Laboratoires d'excellence, lauréats d'appels à projets lancés dans le cadre du Programme investissements d'avenir (Grand emprunt) ont moins présenté ce défaut. Selon l'académicien des sciences Gérard Sabah, c'est la structure de la recherche française qui gênerait la recherche en intelligence artificielle parce que, selon lui, à part les chercheurs du CNRS souvent envahis par les tâches administratives, seuls les doctorants pourraient se consacrer à plein temps à la recherche en intelligence artificielle qui demande beaucoup d'investissement en temps d'expérimentation et en retours d'expériences : « La contrainte d'une thèse en trois ans maximum et le fait qu'après leur soutenance ils se voient généralement obligés de continuer ailleurs (au mieux comme maître de conférences, chargé de nombreuses tâches autres que de recherche pure) font que des projets à long terme ont du mal à se développer. À mon sens, c'est une des raisons pour lesquelles la recherche fondamentale en intelligence artificielle a peu avancé. Comme elle est également proche des sciences cognitives et fait donc appel à diverses disciplines, la difficulté de mener des carrières interdisciplinaires dans le cadre

---

<sup>1</sup> Pierre-Gilles de Gènes estime ainsi que « l'avenir de la recherche est dorénavant aux mains de gens qui ne voient que leur intérêt à trois ans (...). Le mot d'ordre semble être de supprimer tous les investissements à moyen et long terme, en postes de chercheurs comme en moyens de laboratoires. Or une bonne recherche suppose un horizon à dix ans sinon trente ans et les moyens adéquats ».



---

*académique est également un frein à son développement. Or, la fréquentation d'autres disciplines demande une culture à la fois très vaste et très profonde, qui n'est pas toujours dispensée par les formations actuelles : la plupart des systèmes d'enseignement européens fonctionnent selon des structures anciennes, qui ont conduit à la spécialisation ».*

Pour votre rapporteure Dominique Gillot le point évoqué d'une surcharge administrative des chercheurs du CNRS constitue une appréciation controversée : loin d'aller de soi, elle devrait être évaluée et approfondie.

Par ailleurs, et avec un point de vue plus interdisciplinaire, il sera aussi possible de supposer que l'intelligence artificielle permettra de **mieux modéliser le fonctionnement du cerveau et les relations entre le cerveau et la conscience, ainsi que la mémoire.**

**Ces découvertes, qui profiteront aux neurosciences, rétroagiront très probablement avec la recherche en intelligence artificielle dans un cercle vertueux.**

Les projets de recherche *Humain Brain Project* (HBP) et *Blue Brain Project* (BBP) que vos rapporteurs sont allés découvrir *in vivo* en Suisse devraient aller dans ce sens.

Jean-Pierre Changeux, partie prenante des projets, leur en a donné la garantie. L'équipe du *Blue Brain Project* de l'École polytechnique fédérale de Lausanne (EPFL) se donne pour objectif depuis mai 2005 de créer un cerveau synthétique par rétro-ingénierie.

Ce projet pluridisciplinaire permet aussi d'étudier l'architecture et les principes fonctionnels du cerveau. Il a été suivi, en 2013, du HBP financé par l'Union européenne et certains résultats sont déjà tangibles : en octobre 2015, l'équipe est parvenue à assurer la **reproduction virtuelle de microcircuits de neurones de rat.**

### **Présentation du Human Brain project (HBP)**

Le *Human Brain project* (HBP) est un projet scientifique lancé en 2013 qui s'appuie sur le BBP et qui vise à simuler le fonctionnement du cerveau humain grâce à un superordinateur d'ici à 2024. Il a été choisi pour être l'un des deux « *FET Flagships* » de l'Union européenne<sup>1</sup> et son coût total est estimé à 1,19 milliard d'euros. Il s'agit ainsi de mieux comprendre le cerveau et ses mécanismes de base, d'appliquer ces connaissances dans le domaine médical et de contribuer au progrès de l'informatique (et en IA). Les résultats obtenus permettraient par exemple de développer de nouvelles thérapies médicales plus efficaces sur les maladies neurologiques : en effet le projet vise à créer une nouvelle plate-forme informatique

---

<sup>1</sup> Il s'agit des « *Initiatives-phare des Technologies Futures et Émergentes* » de l'Union européenne, soutenues financièrement à hauteur d'un milliard d'euros chacune sur dix ans, dont la moitié est versée par le budget de l'UE (l'autre projet porte sur le graphène).



médicale pour tester des modèles de maladies, améliorer le diagnostic et accélérer le développement de nouvelles thérapies.

S'agissant des progrès en informatique et en IA, l'objectif du projet est de tirer parti d'une meilleure compréhension du fonctionnement du cerveau pour le développement de technologies de l'information et de la communication plus performantes s'inspirant des mécanismes du cerveau humain. Les bénéfices espérés sont une meilleure efficacité énergétique, une fiabilité améliorée et la programmation de systèmes informatiques complexes.

Le projet est mené par une équipe coordonnée par Henry Markram, un neuroscientifique de l'École polytechnique fédérale de Lausanne (EPFL) qui, parallèlement, animait déjà le projet Blue Brain; et codirigé par le physicien Karlheinz Meier de l'université de Heidelberg et le médecin Richard Frackowiak du Centre hospitalier universitaire vaudois et l'université de Lausanne, en collaboration avec plus de 90 instituts de recherche européens et internationaux répartis dans 22 pays différents. Il rassemble au total des milliers de chercheurs.

Le projet a été contesté en 2014 et 2015<sup>1</sup>, ce qui a conduit à réorienter en partie le projet, en accordant notamment plus d'importance aux neurosciences cognitives.

En mai 2015, les ingénieurs du HBP montrent les premières simulations en vue de la réalisation d'une « souris virtuelle » en plaçant un modèle informatique simplifié du cerveau d'une souris dans un corps virtuel soumis à des stimulations. Ces résultats n'ont cependant pas fait cesser les critiques contre le HBP.

La première phase du HBP a débuté fin 2013 et a duré 2 ans et demi. Un premier bilan est donc à dresser.

Pour simuler le fonctionnement d'un cerveau humain, la puissance calculatoire nécessaire est estimée à un exaflops<sup>2</sup>. Or un superordinateur atteignant l'exaflops sera difficile à atteindre avant 2020. Steve Furber de l'université de Manchester souligne que les neuroscientifiques ne savent toujours pas avec certitude quels détails biologiques sont essentiels au traitement cérébral de l'information, et en particulier ceux qu'on peut s'abstenir de prendre en compte dans une simulation visant à simplifier ce processus.

Les perspectives d'application sont grandes, mais la perspective d'un **homme non seulement réparé mais augmenté qui pourrait se dessiner soulève d'importantes questions éthiques.**

Au sein des projets de grande envergure, le plus souvent menés aux États-Unis, en Chine et au Japon (le cas du HBP de l'EPFL peut s'apparenter à cette famille de grands projets pluridisciplinaires), le caractère

---

*1 En juillet 2014, une lettre ouverte signée de 130 scientifiques est adressée à la Commission européenne. Elle critique les orientations prises par la direction du HBP et appelle l'UE à prendre des mesures pour réorienter le projet. L'objet de cette lettre est surtout la gouvernance du HBP, mais aussi le manque de réalisme du projet et son coût important. En 2015, Yann LeCun a critiqué l'idée qu'une IA pourrait émerger simplement d'un ordinateur, aussi puissant soit-il, en utilisant quelques algorithmes simples d'apprentissage. Il pense que les progrès en matière d'IA viendront plutôt de l'apprentissage machine non supervisé.*

*2 10<sup>18</sup> soit un milliard de milliards, ou un trillion de flops (un flop est une unité de mesure de la vitesse d'un système informatique).*

---

pluridisciplinaire est évident : au-delà du matériel informatique et des logiciels, cœur historique de l'intelligence artificielle, d'autres savoir-faire et domaines de connaissances sont largement mobilisés : la robotique, surtout parce que les mouvements et les perceptions sont essentiels pour produire de l'intelligence, les sciences de la vie, à l'image de la physiologie, des neurosciences et de la génétique, mais aussi les sciences humaines, avec la psychologie, la linguistique et la sociologie.

Pour vos rapporteurs, il existe un **débat sur la façon de réussir cette interdisciplinarité dans le cadre existant**, contraint par un académisme d'un autre temps. La **remise en cause du cadre existant** pourrait être bienvenue. La place que le CNRS fait, par exemple, à l'intelligence artificielle dans son organisation semble être, pour le moins, insuffisante.

#### **4. Une recherche soumise à une contrainte d'acceptabilité sociale assez forte sous l'effet de représentations catastrophistes de l'intelligence artificielle**

La recherche en intelligence artificielle est soumise à une **contrainte d'acceptabilité sociale assez forte**, notamment sous l'effet de représentations catastrophistes, comme en témoignent les sondages d'opinion.

Lors du lancement de France IA, le Gouvernement a rappelé que **65 % des Français interrogés se disent inquiets du développement de l'intelligence artificielle** alors que, comparativement, 36 % des Britanniques et 22 % des Américains expriment la même crainte.

Selon un autre sondage Odoxa, réalisé en mai 2016 pour Stratégies et Microsoft, qui posait la question « L'intelligence artificielle, une chance ou une crainte ? », **les Français sont aussi nombreux à voir dans cette forme d'algorithmes une opportunité (49 %) qu'un motif de peur (50 %)**.

Un troisième sondage, réalisé par Orange et « 01.net », réalisé, lui, en ligne parvient à des résultats plus rassurants : **40 % des sondés disent avoir peur** de l'intelligence artificielle et **60 % disent ne pas en avoir peur** (le fait qu'il s'agisse d'un sondage en ligne auprès des internautes introduit forcément plus de biais, les réponses étant issues de personnes déjà familières des TIC).

Les chiffres annoncés par le Gouvernement en janvier 2017 sont en fait les résultats d'une **enquête de l'IFOP** sur le sujet. Ces résultats traduisent, sur un plan statistique, la **prise de conscience assez élevée de l'essor de l'intelligence artificielle et de ses opportunités**.

Cet enthousiasme majoritaire doit être **nuancé par les craintes - elles aussi majoritaires - que cette technologie suscite dans le même temps**. Une majorité de sondés (67 %) voit en effet tout autant l'intérêt de l'intelligence artificielle pour améliorer le bien-être individuel et collectif, qu'il s'inquiète (à 65 %) de l'autonomie croissante des machines.

On se retrouve donc face à une **logique d'appréciation contrastée** qui n'est sans doute pas sans lien avec l'accent mis sur les risques dans la culture populaire (cinéma de science-fiction en particulier) et dans le débat public.

### Enquête d'opinion conduite en 2016 sur l'intelligence artificielle

	Plutôt d'accord (%)	Plutôt pas d'accord (%)	TOTAL (%)
• L'intelligence artificielle est amenée à prendre un essor considérable avec le Big Data .....	69	31	100
• Le Big Data fera l'objet d'une utilisation très importante à long terme par les pouvoirs publics et les entreprises (profilage des individus, surveillance) .....	68	32	100
• Le Big Data présente des avantages à court terme pour la santé et le bien-être des individus (meilleure prévention des maladies et des risques, traitements plus adaptés, découvertes scientifiques, etc.).....	67	33	100
• L'intelligence artificielle caractérisée par l'autonomie croissante des machines (comme les drones armés ou la voiture Google) vous inquiète .....	65	35	100

Source : IFOP

D'après Bernard Stiegler, directeur de l'Institut de recherche et d'innovation (IRI) du centre Georges Pompidou et professeur à l'université de Londres, l'enseignement de ce sondage est le suivant : « *la population est consciente dans son ensemble de l'importance des enjeux liés à la nouvelle intelligence artificielle qui émerge (...) de ses promesses potentielles, mais aussi et surtout de ses dangers. C'est assez rassurant : quand on sait que des personnalités aussi bien informées que Stephen Hawking ou Bill Gates ont elles-mêmes manifesté leur très grande préoccupation avec des dizaines de scientifiques de grand renom face à ce qui se met en place, il est heureux de constater que les personnes interrogées reflètent une conscience de la dimension pharmacologique du numérique, c'est à dire le fait qu'il constitue autant un remède qu'un poison – et que dans l'immédiat, le coût toxique semble s'imposer plutôt que les dimensions curatives* ».

Vos rapporteurs, s'ils y sont sensibles, ne partagent pas l'idée de Bernard Stiegler selon laquelle les coûts de l'intelligence artificielle seraient, à ce stade, supérieurs à ses gains pour nos sociétés. Rien ne permet de le démontrer et la vigilance qui se met en œuvre sur un plan national et international devrait en être garante.

Un dernier **sondage, mené par l'IFOP pour la CNIL** en janvier 2017, peut être cité<sup>1</sup>. Il est consacré au cas plus général des algorithmes et montre que **la majorité des personnes interrogées (64 %) considèrent que les algorithmes représentent plutôt une menace**, en raison de l'accumulation de données personnelles sur les choix, les goûts et les comportements. Cette perception varie fortement en fonction de l'âge : pour les 18-24 ans, les algorithmes représentent d'abord une opportunité (à 51 %).

### Les Français et les algorithmes : notoriété et perception

D'après un sondage mené par l'IFOP pour la CNIL en janvier 2017, les algorithmes sont présents dans l'esprit des Français mais de façon assez confuse. Si 83 % des Français ont déjà entendu parler des algorithmes, ils sont plus de la moitié à ne pas savoir précisément de quoi il s'agit (52 %). Leur présence est déjà jugée massive dans la vie de tous les jours par 80 % des Français qui considèrent, à 65 %, que cette dynamique va encore s'accroître dans les années qui viennent.

Concernant l'opinion sur les algorithmes, une courte majorité (53 %) estime qu'ils sont plutôt sources d'erreur contre 47 % qui pensent qu'ils sont fiables. Mais, la confiance s'élève à mesure que le niveau de connaissance sur les algorithmes progresse. Un effort de pédagogie et de transparence peut donc contribuer à renforcer la confiance.

Sous un angle marketing, 57 % des Français pensent que les algorithmes limitent l'étendue des choix proposés. Chez les plus jeunes, la tendance s'inverse, puisque 53 % des moins de 35 ans et 56 % des 18-24 ans mettent plutôt en avant le fait que les algorithmes proposent plus de choix.

Enfin, c'est sous l'angle de la perception citoyenne que l'opinion est la plus tranchée en fonction de l'âge. Si les 2/3 des Français (64 %) considèrent que les algorithmes représentent plutôt une menace en raison de l'accumulation de données personnelles sur les choix, les goûts et les comportements, les 18-24 ans inversent cette tendance nettement affirmée puisque 51 % estiment au contraire que les algorithmes représentent une opportunité.

Source : CNIL

Vos rapporteurs, qui ont la mémoire des quarante dernières années, n'oublient pas que la popularisation d'Internet dans les années 1990 s'est elle aussi accompagnée, comme auparavant avec la télématique, de l'expression de craintes : **peur de l'affaiblissement des relations sociales**<sup>2</sup>, des **pannes**

<sup>1</sup> Cf. [https://www.cnil.fr/sites/default/files/atoms/files/presentation\\_ifop\\_-\\_presentation.pdf](https://www.cnil.fr/sites/default/files/atoms/files/presentation_ifop_-_presentation.pdf)

<sup>2</sup> Il existe, depuis une trentaine d'années et surtout depuis dix ans, une peur de la perte d'une sociabilité authentique et de son remplacement par une sociabilité virtuelle, moins durable et plus égocentrique. Les travaux du sociologue Dominique Cardon, spécialiste de l'usage de l'IA, des TIC et des comportements sur Internet et les réseaux sociaux, démontrent plutôt une corrélation en sens inverse : les jeunes adeptes des réseaux sociaux enrichissent leur vie sociale physique plus qu'ils ne l'appauvrissent. En la matière, le virtuel ne se substitue donc pas au réel. Rodolphe Gélén et Olivier Guilhem, respectivement directeur scientifique et directeur juridique d'Aldebaran puis de Softbank robotics, ne croient pas que l'intelligence artificielle puisse être un fossoyeur des relations sociales et

---

**informatiques, du piratage, de l'escroquerie commerciale, d'un nivellement intellectuel par le bas<sup>1</sup> ou, encore, de l'exploitation non contrôlée et non autorisée des données personnelles...** Certaines de ces peurs se sont avérées en partie légitimes et vos rapporteurs ne reprennent pas le refrain d'un scientisme optimiste et naïf face aux mutations issues des technologies numériques. Ils ne souscrivent pas, pour autant, à une application stérilisante du principe de précaution qui empêcherait la recherche en intelligence artificielle : le présent rapport le démontre, notamment dans les propositions qui seront développées plus loin.

L'actualité, au-delà de la victoire médiatisée à juste titre d'AlphaGo de DeepMind déjà évoquée et des résultats de plus en plus significatifs obtenus en intelligence artificielle, pourrait sembler mettre **l'accent sur les risques liés à l'intelligence artificielle** :

- le **premier accident mortel** lors d'un trajet en voiture autonome, a eu lieu en Floride dans un véhicule Tesla le 7 mai 2016. Le rapport d'expertise de l'agence fédérale américaine de sécurité routière (NHTSA) a dédouané Tesla et a invoqué des facteurs humains, dans la mesure où « le conducteur de la Tesla, censé garder ses mains à tout moment sur le volant, a eu 7 secondes pour voir le semi-remorque en travers de la route ». Il s'agissait en mai 2016 du premier décès sur plus de 209 millions de kilomètres parcourus par des voitures Tesla avec le pilote automatique activé, or parmi tous les véhicules circulant aux États-Unis, il y a un décès tous les 152 millions de kilomètres ;

- le **premier accident responsable** de la voiture autonome construite par Google, la « Google Car », a eu lieu le 14 février 2016 sur une route de Mountain View ;

- « **Tay** », un avatar algorithmique d'intelligence artificielle créé par Microsoft dans le but de conduire des conversations sur Twitter, est devenu **raciste** en miroir de ses interlocuteurs<sup>2</sup> quelques heures après son activation le 23 mars 2016 ;

---

*rappellent que « lorsque nos enfants jouaient des heures, seuls, sur leur console de jeux ou leur ordinateur, nous redoutions qu'ils deviennent des êtres désociabilisés n'interagissant plus qu'avec leur machine. Ils ont aujourd'hui beau jeu de nous dire que, grâce aux réseaux sociaux et aux jeux en réseau, ils sont bien plus en relation avec le monde que nous, perdus dans nos livres en papier pendant des heures sans parler à personne ! ».*

<sup>1</sup> « Est-ce que Google nous rend idiots ? » *s'interrogeait Nicholas Carr en juin 2008 dans la revue The Atlantic (traduction du nom anglais de l'article « Is Google making us stupid ? »).* La consommation de plus en plus rapide et superficielle d'informations au travers des outils numériques est patente et remet en question nos capacités intellectuelles traditionnelles, nos relations sociales et nos habitudes de travail. Une étude canadienne publiée en février 2016 démontre que les réseaux sociaux encourageraient une pensée rapide et superficielle pouvant, à terme, entraîner une superficialité cognitive et morale (cf. Logan Annisette et Kathryn Lafreniere, « Social media, texting, and personality : a test of the shallowing hypothesis » *Department of Psychology, University of Windsor*).

<sup>2</sup> Cette évolution de Tay s'est faite au contact d'internautes cherchant délibérément à tester les limites du système en le faisant déraiper d'un point de vue politique et moral.

- une association française contre l'intelligence artificielle (AFCIA) a déposé ses statuts le 18 juillet 2015<sup>1</sup> et son président Cédric Sauviat a été rencontré par vos rapporteurs, qui ont jugé ses analyses excessives voire infondées, mais révélatrices d'un certain climat d'angoisse puisque la France serait le seul pays où une telle association existerait<sup>2</sup>. L'AFCIA juge « *illégitime et dangereuse la recherche scientifique visant à créer des organismes à intelligence artificielle supra-humaine* » et considère que le seul moyen « *d'éviter un avenir funeste pour l'humanité est d'obtenir l'interdiction légale de la recherche en intelligence artificielle à l'échelle mondiale* ». Se définissant comme association de lobbying, elle vise à **obtenir cette interdiction auprès des pouvoirs publics** ;

- et plusieurs interventions médiatiques, pétitions et lettres ouvertes, ont cherché en 2015 à interpeler l'opinion à propos des risques qui seraient inhérents à l'intelligence artificielle<sup>3</sup>.

Des figures médiatiques ont par ailleurs tenu des discours catastrophistes. Ainsi, Stephen Hawking, professeur de mathématiques connu pour ses contributions dans les domaines de la cosmologie et la gravité quantique, a déclaré l'année dernière dans une interview à la BBC que « *les formes d'intelligences que nous avons déjà se sont montrées très utiles. Mais je pense que le développement d'une intelligence artificielle complète pourrait mettre fin à la race humaine. Les humains, limités par une lente évolution biologique, ne pourraient pas rivaliser et seraient dépassés* ».

Bill Gates, le fondateur de Microsoft, s'est aussi inquiété en 2015 des progrès de la super-intelligence : « *dans quelques décennies, l'intelligence sera suffisamment puissante pour poser des problèmes* ».

Jacques Attali s'est, à son tour, expliqué sur les priorités souhaitables en matière d'intelligence artificielle et s'est prononcé à la fin de l'année 2016 pour un moratoire sur les technologies d'intelligence artificielle, ce qui a surpris vos rapporteurs.

## 5. Une recherche en intelligence artificielle qui s'accompagne de plus en plus d'interrogations et de démarches éthiques

Vos rapporteurs ont observé une **multiplication des initiatives visant la prise en compte de principes éthiques dans la recherche et les**

<sup>1</sup> La lecture de son site, qui utilise en fond d'écran des images extraites des films Metropolis et Matrix, est édifiante : <http://afcia-association.fr/>

<sup>2</sup> Ce constat est renforcé par l'existence d'autres associations technophobes, inspiré par John Zerzan, Jacques Ellul, Ivan Illich ou Georges Bernanos et parfois par le néo-luddisme, à l'instar du collectif « Pièces et main d'œuvre » (souvent abrégé en PMO) : <http://www.piecesetmaindoeuvre.com/>

<sup>3</sup> À l'image en janvier 2015 de la lettre d'avertissement sur les dangers potentiels de l'intelligence artificielle signée par 700 personnalités (le plus souvent des scientifiques et des chefs d'entreprise, rejoints par plus de 5000 signataires en un an) et, en juillet 2015 pour l'ouverture de la Conférence internationale sur l'intelligence artificielle qui s'est tenue à Buenos Aires, de la lettre signée par plus de mille personnalités, demandant l'interdiction des robots tueurs, lettres évoquées dès l'ouverture de l'introduction du présent rapport.



---

**usages de l'intelligence artificielle.** Cela vaut pour la recherche publique, comme pour la recherche privée, en Europe comme en Amérique. Il s'agit d'une caractéristique qui singularise la recherche en intelligence artificielle.

Ils décrivent les détails de cette réalité de plus en plus tangible plus loin dans le rapport, dans la partie consacrée à la prise en compte grandissante des enjeux éthiques.

## **B. TABLEAU DE LA RECHERCHE FRANÇAISE EN INTELLIGENCE ARTIFICIELLE**

### **1. De nombreux organismes publics intervenant dans la recherche en intelligence artificielle**

Notre pays dispose, en matière de recherche en intelligence artificielle, d'**importants atouts à faire valoir**, riche de la compétence de ses enseignants, de ses chercheurs et de ses étudiants, même si la communauté française de l'intelligence artificielle est encore insuffisamment organisée, connue et visible. **La reconnaissance internationale des travaux des chercheurs français doit beaucoup à nos universités, au CNRS, à nos grandes écoles mais aussi plus spécifiquement à deux organismes : le Commissariat à l'énergie atomique et aux énergies alternatives (CEA) et ses nombreux centres de recherche (à l'image de l'institut Carnot « CEA-List » spécialisé dans les systèmes numériques intelligents pour l'industrie), et l'Institut national de recherche en informatique et en automatique (INRIA), créé dès janvier 1967 dans le cadre du Plan Calcul sous le statut d'établissement public à caractère scientifique et technologique.**

Les centres de recherche de l'**Institut Mines-Télécom**, qui regroupe les écoles des mines et les écoles des télécommunications françaises, méritent aussi d'être signalés, notamment le centre de recherche en informatique (CRI), le centre de Bio-informatique (CBIO) et le centre de Robotique (CAOR).

Au total, l'INRIA, le CNRS, le CEA, différentes universités et grandes écoles, par exemple l'ENS et Mines-télécom, sont les **principaux organismes de recherche publique en intelligence artificielle et produisent des travaux à visibilité internationale.**

Cette excellence est reconnue comme ont pu le constater vos rapporteurs lors de leurs déplacements. Le bon niveau des étudiants et des enseignants est également souvent cité.

## Les principaux organismes de recherche publique en intelligence artificielle (à visibilité internationale)



### Examples of Research centers



Source : ISAI/Paul Strachman

L'**excellence de l'école mathématique française** contribue à nos succès, avec un nombre de 13 médaillés Fields, soit une place de numéro 2, juste derrière les États-Unis avec 14 médaillés Fields. L'ENS, avec ses 11 médailles, figure en tête de la liste des institutions au niveau mondial. Cinq de nos universités figurent dans le top 30 des universités de mathématiques et une dans le top 5. Nous disposons de plus de 200 écoles d'ingénieurs qui forment chaque année 38 000 nouveaux diplômés. Plusieurs de ces écoles disposent de cursus ou de formations en intelligence artificielle ou en robotique.

L'histoire montre que la recherche française en matière d'intelligence artificielle a toujours été **assez forte** et s'est placée à une place enviable par rapport à ses concurrents en recherche fondamentale, même si elle court le **risque d'un décrochage face aux pays les plus avancés** dans la course mondiale en intelligence artificielle, États-Unis, Chine et Royaume-Uni en tête.

**Au tournant des années 1970 et 1980, la recherche en intelligence artificielle en France a connu une certaine accélération**, avec notamment le groupe de recherche parisien animé par Claude-François Picard appelé « GR 22 ». Jacques Pitrat, Jean-Louis Laurière, Jean-François Perrot et Jean-Charles Pomerol y ont, par exemple, travaillé. En 1987, il a pris le nom de Laboratoire FORMes et Intelligence Artificielle (LAFORIA) puis a rejoint l'Institut Blaise Pascal (IBP) en 1989 avant de fusionner en 1997 au sein du Laboratoire d'Informatique de l'Université Paris 6 dit « LIP6 ». D'autres laboratoires investis dans la recherche en intelligence artificielle ont également marqué les années 1980-1990 en France<sup>1</sup>. Beaucoup sont rappelés sous leur nom actuel au paragraphe suivant.

<sup>1</sup> Le Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMSI) à Orsay, l'Institut de recherche en informatique et en automatique (IRIA) et en particulier le Laboria (autour



---

**Un inventaire des douze principaux laboratoires du CNRS en matière d'intelligence artificielle** ayant été dressé par Gérard Sabah<sup>1</sup>, vos rapporteurs ont souhaité les rappeler ici (le nombre de chercheurs par laboratoire n'a pas été actualisé et a pu connaître des variations depuis cet inventaire) :

- le groupe de recherche en informatique, image, automatique et instrumentation (GREYC), spécialisé dans le traitement automatique des langues, la sémantique, et la fouille de données, basé à Caen, compte 20 permanents ;

- l'Institut de recherche en informatique de Toulouse (IRIT) est spécialisé dans la communication, les agents intelligents, l'ingénierie des connaissances, l'aide au handicap, et le traitement automatique des langues. Il compte 40 collaborateurs ;

- le laboratoire d'architecture et d'analyse des systèmes (LAAS), situé à Toulouse, dédié au logiciel, à la communication et à la robotique, compte 40 permanents ;

- le laboratoire d'analyse et modélisation de systèmes pour l'aide à la décision (LAMSADE) établi en région parisienne, spécialisé dans les agents intelligents et modèles coopératifs ainsi que dans la gestion des connaissances, est composé de 15 chercheurs ;

- le laboratoire des langues, textes, traitements informatique et cognition (LATTICE) basé en région parisienne, est spécialisé dans le traitement automatique des langues avec 5 permanents ;

- le laboratoire d'informatique fondamentale (LIF), établi à Marseille, est spécialisé dans l'apprentissage et le traitement automatique des langues et regroupe 17 agents ;

- le laboratoire d'informatique de Grenoble (LIG) est spécialisé dans l'environnement pour apprentissage, la traduction automatique, la réalité virtuelle et les agents intelligents, il est composé de 35 permanents ;

- le laboratoire d'informatique pour la mécanique et les sciences de l'ingénieur (LIMSI) basé à Orsay et spécialisé en agents communicants, le traitement automatique des langues et de la parole, ainsi qu'en réalité virtuelle, regroupe 15 chercheurs ;

- le laboratoire d'informatique de Paris-Nord (LIPN) est spécialisé dans l'apprentissage, la logique, le calcul, le raisonnement, la représentation des connaissances et le traitement automatique des langues. On y dénombre 25 collaborateurs ;

- le laboratoire d'informatique, de robotique et de microélectronique de Montpellier (LIRMM) est spécialisé dans l'apprentissage, les contraintes, la représentation des connaissances, les systèmes multi-agents, le traitement

---

*de Gérard Huet) à Rocquencourt, le GIA (autour d'Alain Colmerauer) à Marseille, le CRIN à Nancy ou, encore, le CERFIA à Toulouse.*

<sup>1</sup> *Cet inventaire est présenté dans la brochure de l'Académie des Sciences déjà citée.*

---

automatique des langues, la visualisation, le *web* sémantique. Il est composé de 17 permanents ;

- le laboratoire lorrain de recherche en informatique et ses applications (LORIA) situé à Nancy, est spécialisé dans la communication multimodale, la représentation et la gestion des connaissances, la reconnaissance de l'écriture, le traitement automatique des langues et de la parole. Il regroupe 30 agents ;

- le laboratoire de recherche en informatique (LRI) établi en région parisienne, est spécialisé dans l'apprentissage et l'optimisation, les systèmes d'inférence. Il compte 24 permanents ;

- le laboratoire « techniques de l'ingénierie médicale et de la complexité » (TIMC) basé à Grenoble, est spécialisé dans l'apprentissage, la sémantique, la gestion et le traitement des connaissances. Il est composé de 12 collaborateurs.

## 2. Quelques exemples de centres, de laboratoires et de projets de recherche

Les *start-ups* profitent également des atouts des centres de recherche français en intelligence artificielle. Ainsi, **Heuritech** s'appuie sur les travaux de recherche de deux laboratoires publics le LIP6 (CNRS) et l'ISIR de l'UPMC (Paris VI) pour proposer sa solution logicielle Hakken d'analyse sémantique, de tagging et classement automatiques de textes, images et vidéos<sup>1</sup>, en s'appuyant sur des technologies de *machine learning* et en particulier de *deep learning*. Les avancées en matière de robotique en France ont permis à des innovations d'essaimer, en particulier l'entreprise Aldebaran. La *start-up* **Angus.AI**, créée par d'anciens ingénieurs d'Aldebaran ayant développé la partie logicielle des robots Nao et Pepper, a ainsi développé une solution logicielle embarquée dans les robots leur apportant des fonctions de base de reconnaissance vocale et faciale et de détection d'obstacles, qui sont fournies sous la forme d'un kit de développement et d'interfaces de programmation applicative (souvent désignée par le terme *API* pour *Application Programming Interface*), en recourant largement à des solutions *open source*. Cette entreprise est déjà sous contrat avec la SNCF.

Le **Laboratoire d'informatique pour la mécanique et les sciences de l'ingénieur** (LIMSI-CNRS), situé à Orsay, sur le campus de l'Université Paris-Sud, au sein de l'Université Paris-Saclay, mène des recherches sur deux grands thèmes : la mécanique et l'énergétique, d'une part, et la communication homme-machine d'autre part. Les recherches en interaction homme-machine portent sur : l'analyse, la compréhension et la modélisation des interactions entre humains et systèmes artificiels dans des contextes et

---

<sup>1</sup> Ils proposent aussi HeuritechDIP qui permet d'améliorer sa connaissance des clients et d'anticiper leurs besoins, évidemment, surtout dans les applications de commerce en ligne.

---

selon des modalités les plus variées, les interactions haptiques, tangibles, gestuelles et ambiantes, la psychologie des interactions affectives non-verbales et collectives chez l'humain ainsi que sur la conception d'interfaces homme-machine les faisant intervenir, ou, enfin, sur les dispositifs de réalité virtuelle et augmentée.

Le **Laboratoire lorrain de recherche en informatique et ses applications** (LORIA-CNRS et Université de Lorraine), basé à Nancy est un autre exemple de centre qui mérite d'être cité. Un grand nombre de projets de recherche fondamentale en intelligence artificielle référencés sur leur site font appel aux technologies de l'intelligence artificielle, même s'ils ne sont pas forcément labellisés intelligence artificielle, *machine learning* ou réseaux neuronaux. C'est ainsi le cas du projet Orpailleur<sup>1</sup> mené à Nancy en lien avec l'INRIA et dédié à la représentation des connaissances et au raisonnement. L'équipe travaille sur l'extraction de données dans les bases de connaissances non structurées, et notamment dans le domaine de la santé, le même que celui qui est investi par IBM Watson et de nombreuses *start-ups*.

### 3. Une reconnaissance internationale de la recherche française et qui s'accompagne d'un phénomène de rachat de *start-ups* et de fuite des cerveaux lié aux conditions attractives offertes à l'étranger

Outre le départ de Yann LeCun, pour l'Université de New York puis pour Facebook, vos rapporteurs s'inquiètent d'un phénomène de **rachat de *start-ups* et de fuite des cerveaux** lié aux conditions attractives offertes à l'étranger. Lors de son audition Stéphane Mallat a fait valoir que depuis plusieurs années **la quasi-totalité des étudiants issus des masters spécialisés de l'ENS** quittaient la France aussitôt effectuée leur formation.

---

<sup>1</sup> Le projet vise la découverte de connaissances dans les bases de données (KDD pour Knowledge Discovery in Databases) et utilise l'ingénierie des connaissances (KE pour Knowledge Engineering). Le processus de KDD consiste à traiter de grands volumes de données pour y découvrir des motifs qui sont signifiants et réutilisables. En considérant les motifs comme des pépites d'or et les bases de données comme des régions à explorer, le processus de KDD peut être comparé à la recherche d'or. Cette analogie explique le nom de l'équipe, où l'orpailleur désigne le chercheur d'or. Le processus de KDD est interactif et itératif et s'appuie sur trois opérations principales : la préparation des données, la fouille de données et l'interprétation des motifs extraits. Les connaissances du domaine peuvent être prises en compte pour guider et améliorer le processus de KDD, conduisant à la découverte de connaissances guidée par les connaissances du domaine (KDDK pour Knowledge Discovery guided by Domain Knowledge). Les motifs découverts peuvent être représentés comme des éléments de connaissances en utilisant un langage de représentation des connaissances et servir en résolution de problèmes. La découverte de connaissances et l'ingénierie des connaissances sont deux processus complémentaires qui servent de support aux recherches menées dans l'EPI Orpailleur. Les domaines d'application traités par l'EPI Orpailleur sont liés aux sciences de la vie et comprennent l'agronomie, la biologie, la chimie et la médecine. La cuisine, la culture et l'héritage culturel, la sécurité et la qualité des réseaux de télécommunications sont aussi des domaines d'intérêt pour l'équipe. <https://www.inria.fr/equipes/orpailleur>

---

La reconnaissance des talents français est donc certaine. Mais cela conduit à un pillage de nos talents qui résulte de conditions attractives, à commencer par les salaires, qui conduisent à une aspiration des jeunes diplômés français spécialisés par les entreprises, le plus souvent des entreprises américaines. Le mirage de la Silicon Valley fait sans doute rêver beaucoup de jeunes esprits brillants, mais cette caractéristique n'est pas propre à la France.

Il faut permettre à ces jeunes génies, qui sont autant d'entrepreneurs en devenir, de **disposer d'opportunités en France et permettre aux start-ups de se développer sans être rachetées par les géants américains, chinois ou japonais** dès qu'elles présentent un profil viable. Le cas d'Aldebaran, racheté par le japonais SoftBank, illustre le fait que nos talents ne sont pas chassés que par les firmes américaines du numérique mais aussi par les géants chinois ou japonais.

#### **4. Une communauté française de l'intelligence artificielle encore insuffisamment organisée et visible**

La communauté française de l'intelligence artificielle se constitue surtout **en-dehors des institutions, à travers les meetups**. Le principal d'entre eux, le « *Paris Machine learning Meetup* »<sup>1</sup>, auquel ont participé vos rapporteurs, représentent 5 205 membres au 1<sup>er</sup> mars 2017. Ses animateurs Igor Carron, Franck Bardol, Frédéric Dembak et Isabelle Guyon jouent maintenant un rôle clé dans la communauté française de l'intelligence artificielle. Les réunions sont en général mensuelles et réunissent des centaines de participants. Il peut, par exemple, s'agir d'échanges entre développeurs utilisant TensorFlow, la technologie de Google mise en *Open source*.

D'autres *meetups* peuvent être cités : « *Paris.AI Meetup* », « *Big Data Paris Meetup* », « *Deep Learning Meetup Paris* », « *Big Data et Machine learning Paris Meetup* ».

D'après les chiffres provisoires du Gouvernement dans le cadre de la cartographie réalisée pour France IA il existerait **230 équipes de recherche publique réparties sur toute la France, dont 50 en région parisienne, ce qui représenterait 5 300 chercheurs en France**, ce nombre de chercheurs travaillant sur l'intelligence artificielle pouvant varier de façon significative selon qu'on y inclut ou pas la recherche SHS portant sur l'intelligence artificielle. La compétence de nos chercheurs est reconnue.

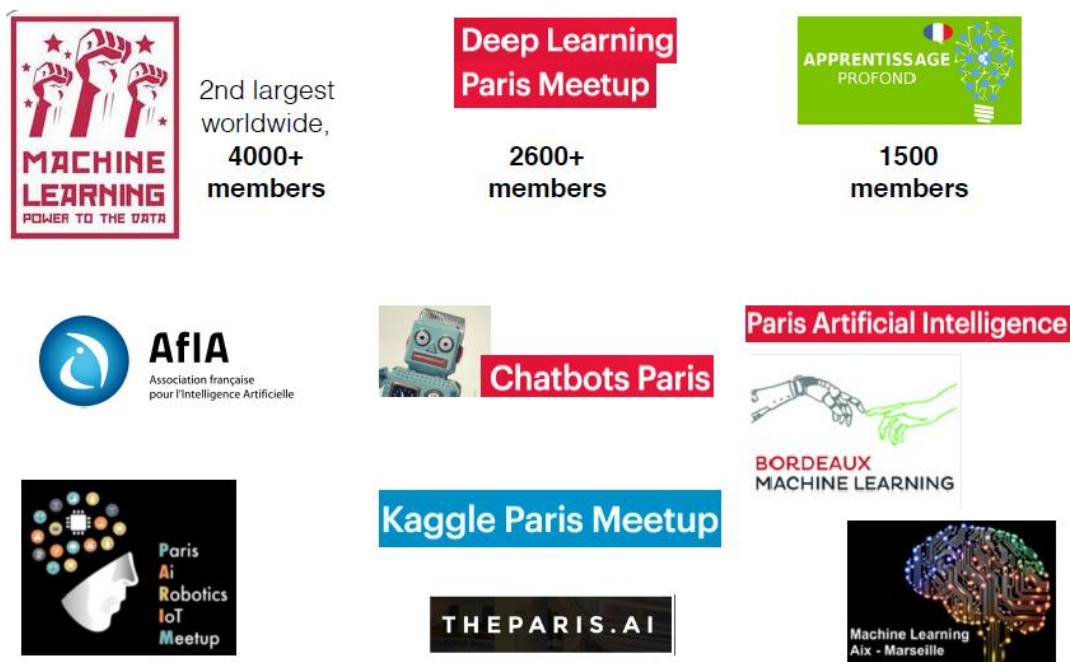
Vos rapporteurs ont relevé, par ailleurs, l'existence d'une **association française pour l'intelligence artificielle (AFIA)**. Société savante composée d'environ **300 membres**, l'AFIA semble souffrir d'une certaine fermeture sur elle-même. Elle a été créée en 1993 en vue de l'organisation de la conférence internationale en intelligence artificielle, c'était une nécessité

---

<sup>1</sup> Voir le site <https://www.meetup.com/fr-FR/Paris-Machine-learning-applications-group/> et le site d'Igor Carron <http://nuit-blanche.blogspot.fr/>

pour gérer l'organisation, mais c'est donc sous l'effet de la structuration internationale de ce champ de recherche plus que de la mobilisation des chercheurs français que l'AFIA a été créée. L'association s'assume comme société savante d'un sous-secteur de l'informatique et aurait tout intérêt à **transcender ses propres limites pour relever le défi d'une intelligence artificielle française ouverte, visible et conquérante.**

**Tableau de la communauté française de l'intelligence artificielle à la fin de l'année 2016**



Source : ISAI/Paul Strachman

La communauté française de l'intelligence artificielle reste encore insuffisamment organisée et visible. Et la question du lien avec les institutions publiques doit être posée.

Votre rapporteure Dominique Gillot se réjouit de **l'expérience des groupements de recherche interdisciplinaires** dans lesquels il y a aussi des entrepreneurs. Les groupements de recherche (GdR) ont pour missions **l'animation de la diffusion des connaissances dans une communauté thématique, l'effort de rapprochement** entre plusieurs types de partenaires (institutionnels, industriels ou prestataires), le développement d'échanges de chercheurs, de doctorants ainsi que de faciliter la mise en place de bourses doctorales ou post-doctorales. Les GdR poursuivent l'objectif **d'accompagner la recherche** dans un domaine spécifique et ses applications, de **fédérer une communauté pluridisciplinaire** et de **diffuser les fruits des réflexions menées**, des avancées tant théoriques que technologiques et des résultats opérationnels obtenus. Un GdR est une **structure originale mise en place par**

---

le CNRS afin d'encourager la coopération des chercheurs des laboratoires qu'il rassemble avec des industriels et des organismes de R&D sur des objectifs fixés à l'avance. Son rôle consiste à coordonner, rapprocher et évaluer les travaux des équipes concernées.

### 5. La sous-estimation des atouts considérables de la France et le risque de « décrochage » par rapport à la recherche internationale en intelligence artificielle

La France se situe à un **stade intermédiaire en matière de publications** dans le domaine de l'intelligence artificielle, mais dispose d'un **réseau de chercheurs très compétent et d'un tissu de *start-ups* très dynamique**. Les atouts considérables du système d'enseignement et de recherche sont connus : excellence de l'école mathématique française, qualité de la recherche et des diplômés...

On dénombre un total de **240 *start-ups* spécialisées en intelligence artificielle** (comme, en France, Bayes Impact, Yseop, Prestashop, Adomik, Gorgias, Owkin, Search'XPR, Placemeter, Otosense, Wit.ai, Wca Robotics, Rhythm, Intuitive surgical, Lore.ai, ou au Royaume-Uni BigRobots et en Russie Brainify).

Le tissu de *start-ups*, que l'initiative FrenchTech vise à renforcer, est très riche et, selon l'investisseur en intelligence artificielle, Paul Strachman, que vos rapporteurs ont auditionné : « *La France est l'un des écosystèmes les plus vibrants en ce qui concerne l'intelligence artificielle. Malheureusement, cela n'est pas très su en-dehors de la France. Et parfois même en-dedans* ». Il convient de noter qu'il a tenu ces propos lors d'une conférence « *France is AI* » qu'il a organisée à Paris du 16 au 18 septembre 2016.



## Les start-ups françaises en intelligence artificielle et robotique

### Agents



### Robotics (Hardware + Software)



### Enterprise



### Platforms



### Industries



Source : ISAI/Paul Strachman

Plusieurs signes positifs peuvent être relevés, dont le succès et l'ambition internationale d'entreprises françaises, comme BlaBlaCar, Criteo ou Drivy, et l'ouverture de formations « alternatives » dans le secteur du numérique (l'école 42 de Xavier Niel par exemple).

Selon les données de Stack Overflow, l'Europe compte désormais **4,7 millions de développeurs** (contre 4,1 millions aux États-Unis). Et Paris est l'une des capitales européennes qui comptent le plus grand nombre de ces profils très recherchés sur le marché.

Pour Mark Zuckerberg, le Président de Facebook, « la France dispose de l'une des communautés de chercheurs en intelligence artificielle la plus forte du monde ». De même Mike Schroepfer, le directeur technique de Facebook, estimait en 2015 que Paris avait « la plus grande concentration de toute l'Europe en matière d'intelligence artificielle ».

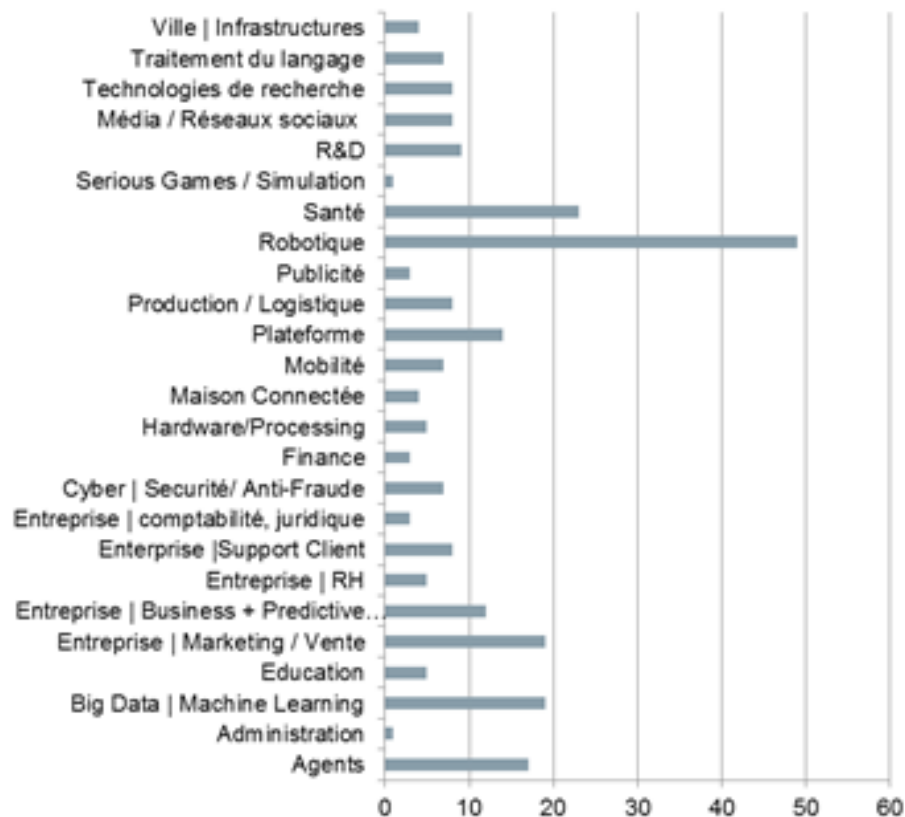
Une déclaration qui s'est accompagnée de l'ouverture en juin 2015 d'un laboratoire de recherche dédié à l'intelligence artificielle à Paris, justement préférée à Londres, sachant que les travaux de Facebook dans ce domaine sont, au niveau mondial, pilotés par le chercheur Yann LeCun, rencontré à plusieurs reprises par vos rapporteurs.

Niklas Zennström, P-DG d'Atomico et cofondateur de Skype, explique dans une tribune au journal *Les Échos* que la capitale française pourrait devenir « *le creuset des grands leaders qui révolutionneront le secteur high-tech dans les dix ans qui viennent* ».

Mais, comme le précise un article de *Venture Beat*<sup>1</sup>, utilisant des données collectées par Paul Strachman, entre janvier 2014 et mi-octobre 2016, une trentaine de *start-ups* françaises spécialisées en intelligence artificielle soulevaient 108 millions de dollars (98 millions d'euros), quand dans le même temps, outre-Manche, huit *start-ups* amassaient à elles seules 900 millions de dollars (814 millions d'euros). Londres possède donc une **avance indéniable sur Paris** en la matière.

La France, qui est à ce stade encore distancée par le Royaume-Uni, serait par ailleurs très en avance par rapport aux autres États européens. Le niveau de financement reste encore en-deçà des besoins.

### Les *start-ups* françaises en intelligence artificielle par domaine d'application



Source : Gouvernement

**Snips.ai** est une *start-up* connue du secteur de l'intelligence artificielle, créée en 2013, par Rand Hindi (prix du MIT 30 en 2015), Mael Primet et Michael Fester. Leur dernière levée de fonds de 5,7 millions

<sup>1</sup> <http://venturebeat.com/2016/11/06/france-makes-its-bid-to-be-recognized-as-a-global-ai-hub/>



---

d'euros en juin 2015 présente la particularité d'associer Bpifrance avec des investisseurs américains, en plus de *business angels* tels que Brent Hoberman et Xavier Niel. L'équipe comprend 35 personnes : des data-scientists, des développeurs, designers et quelques marketeurs. Leur positionnement est large et un peu vague : rendre la technologie invisible et les usages intuitifs *via* de l'intelligence artificielle. À ce titre, la *start-up* a développé des applications expérimentales telles que : snips (un ensemble d'applications de recherche pour iOS dont un clavier virtuel intelligent pour la recherche d'adresses), Tranquilien (qui prédit les places disponibles dans les trains de banlieue), Parkr (la même chose pour prédire les places de parking), Flux (qui identifie le trafic mobile en s'appuyant sur les données des smartphones), RiskContext et SafeSignal (identification de risques d'accidents sur la route). La *start-up* planche aussi sur des applications verticales : pour les véhicules connectés, dans l'hôtellerie, la maison connectée et les loisirs numériques. Le tout s'appuie sur force machine et *deep learning*, modèles probabilistiques, traitement du langage, gestion de graphes et aussi encryption des données pour garantir la vie privée. Derrière la vision, l'implémentation et l'expérimentation, on leur souhaite de réussir la construction de leur « *business model* ».

**CardioLogs Technologies** a été créée en 2014 une solution d'interprétation automatique des électrocardiogrammes (ECG) en temps réel s'appuyant sur du *machine learning*. Il ne s'agit pas à ce stade d'une ubérisation des cardiologues mais de permettre un suivi plus régulier des patients à risques ou atteints de maladies chroniques.

L'attractivité française peut également être relevée avec les cas des entreprises étrangères ayant fait le choix de conduire une part de leur recherche en intelligence artificielle en France.

---

### Exemples d'entreprises étrangères ayant fait le choix de conduire une part de leur recherche en intelligence artificielle en France



Source : ISAI/Paul Strachman

Il convient en outre de relever l'existence d'une **importante diaspora des chercheurs français en intelligence artificielle**. Elle pourrait sans doute être davantage mise à contribution.

Au total, vos rapporteurs déplorent la **sous-estimation des atouts considérables de la France**, mais ils appellent l'attention sur le maintien d'une vigilance forte sur le **risque de « décrochage »** par rapport à la recherche internationale en intelligence artificielle

---

## DEUXIÈME PARTIE : LES ENJEUX DE L'INTELLIGENCE ARTIFICIELLE

### I. LES CONSÉQUENCES ÉCONOMIQUES ET SOCIALES DE L'INTELLIGENCE ARTIFICIELLE

#### A. D'IMPORTANTES TRANSFORMATIONS ÉCONOMIQUES EN COURS OU À VENIR

##### 1. L'évolution vers une économie globalisée dominée par des « plateformes »

Ces technologies, leurs usages et leurs artisans ayant été décrits, il importe d'identifier leurs **impacts sociaux et économiques potentiels**, pour en relever ensuite les enjeux éthiques et juridiques.

Alors qu'Isaac Asimov affirmait qu'« *il est une chose dont nous avons maintenant la certitude : les robots changent la face du monde et nous mènent vers un avenir que nous ne pouvons encore clairement définir* » (*Le Cycle des robots*), les transformations que nous connaissons d'un point de vue économique et technologique semblent être les **signes avant-coureurs de l'évolution vers une économie globalisée de « plateformes »**. Vos rapporteurs ont eu reçu confirmation de cette intuition lors de leur déplacement aux États-Unis. D'après Brian Krzanich, P-DG d'Intel, premier fabricant mondial de microprocesseurs, « *l'intelligence artificielle n'est pas seulement le prochain raz de marée de l'informatique, c'est aussi le prochain tournant majeur dans l'histoire de l'humanité* ». Il est vrai qu'elle peut devenir l'un des fondements de la société et de l'économie de demain.

On parle des « **GAF**A », parfois des « **GAF**AMI », mais il serait plus juste de parler des « **GAF**AMITIS »<sup>1</sup>, « **NATU** »<sup>2</sup> et « **BATX** »<sup>3</sup>. Ces exemples emblématiques des bouleversements en cours sont les prémises de la **place dominante et monopolistique occupée par quelques entreprises** dans un contexte d'économie globalisée de « plateformes ». Un accroissement significatif des investissements dans la recherche en intelligence artificielle est constaté. Chacune de ces entreprises est entrée dans une course pour **acquérir une position de pointe dans les technologies d'intelligence artificielle afin de tirer profit de la situation dominante qui en résultera**. Comme l'affirme Guillaume Devauchelle, directeur de la R&D de Valeo « *dans la ruée vers l'or en Amérique, ceux qui ont fait fortune, ce sont surtout les vendeurs de pioches et de pelles. C'est exactement ce que nous sommes : nous*

---

<sup>1</sup> Google, Apple, Facebook, Amazon, Microsoft, IBM, Twitter, Intel et Salesforce. Ces entreprises américaines représentent la pointe de la recherche et des applications de l'IA.

<sup>2</sup> Netflix, Airbnb, Tesla et Uber.




<sup>3</sup> L'expression désigne les géants chinois du numérique : Baidu, Alibaba, Tencent et Xiaomi.

fournissons des technologies ». Grâce à leurs efforts en R&D et aux rachats de *start-ups*, les grands équipementiers sont bien positionnés dans la chaîne de valeur automobile. Valeo fait la course en tête et, avec la voiture autonome, se positionne d’ores et déjà comme la **future principale plateforme de fourniture de technologies innovantes aux constructeurs automobiles**.


En effet, les technologies d’intelligence artificielle renforcent le modèle « *the winner takes it all* », ou au moins « *the winner takes it most* », ce qui pourrait bien conduire à une **concentration horizontale** progressive des grandes entreprises, conduisant au monopole de « plateformes » dominant une économie globalisée.

On assiste à une **montée en puissance significative du nombre des acquisitions**, après 140 *start-ups* en intelligence artificielle achetées de 2011 à 2016, on a vu, en 2016, 40 *start-ups* en intelligence artificielle être achetées par des grandes entreprises, pour des valeurs allant de 30 millions à 400 millions de dollars.

### Le mouvement de concentration autour de l’intelligence artificielle visible dans les stratégies d’acquisitions

<b>Google</b> Cleversense (2011) DNNresearch(2013) DeepMind (2014) \$660m Emu (2014) Jetpac (2014) Granata Decision Systems Timeful (2015) Api (2016) Moodstocks (2016)  Kifi (2016) Dark Blue Labs (2014) Vision Factory (2014) Hark (2016)  DeepMind	<b>Amazon</b> Orbeus (2015) Angel (2016) Harvest.ai (2017) DeepL, reconnaissance images, assistants	<b>Facebook</b> Face.com (2012) \$55m Wit.ai (2015) Agents, recherche d'images	<b>IBM</b> Cognea (2014) AlchemyAPI (2015) Explorys (2015) DeepL, santé, agents conversationnels (Watson)	<b>Microsoft</b> Netbreeze (2013) <i>Entreprise</i> Equivio (2015) <i>Droit et IA</i> Genee(2016) <i>Assistants</i> SwiftKey (2016) \$250m Maluuba (2017)
<b>Apple</b> Vocal IQ (2015) Perceptio (2015) Turi (2016) \$200m Tuplejump (2016) Emotient (2016) <i>Traitement langage</i> <i>Reconnaissance faciale</i>	<b>Twitter</b> Madbits (2014)  TellApart (2015) Whetlab (2015) Magic Pony (2016) \$150m ML, images	<b>Oracle</b> Crosswise (2016) \$50m Palerra (2016) Big data Client, Cyber-sécurité	<b>Intel</b> IQ Engines (2013) Saffron (2015) Itseez (2016) Nervana (2016) \$408m Movidius (2016) \$400m GPU	
<b>Uber</b> Geometric Intelligence (2016) <i>Nouveau AI Lab: conduite autonome</i>	<b>Yahoo</b> Indisys (2013) LookFlow (2013) SkyPhrase (2013)	<b>Samsung</b> Viv Labs (2016) Assistants virtuels (concurrer Apple)	<b>Aol</b> Gravity (2014) <i>Recomm.</i> Convertro (2014) Sociocast, Velos (2015)	<b>NICE</b> Causata (2013) Nexidia (2016) \$135m Fouilles de données client
<b>Ford</b> SAIPS (2016) Israël Argo.Ai (2017) \$1Md Apprentissage non-supervisé (video) pour voiture autonome	<b>General Electric</b> Wise.io (2016) Bit Stew (2016) \$135m Apprentissage (Plateforme Predict), IoT.	<b>eBay</b> Hunch (2011) Expertmaker (2016) SalesPredict (2016) Prediction achat client	<b>SalesForce</b> Tempo AI (2015) MetaMind (2016) PredictionIO (2016) Calendrier, Deep learning.	<b>Nokia</b> Desti (2014) Medio Systems (2014) Navigation (HERE maps)

**Légende :**

-  Assistants
-  Santé  Vidéo
-  Prédicatif client
-  Cyber-sécurité
-  Voiture autonome

Source : Gouvernement

Légende :

En mars 2017, un **partenariat entre IBM et Salesforce**, premier employeur de San Francisco (dont vos rapporteurs ont rencontré des responsables en son siège), a été conclu sous l’appellation : « *Watson meets Einstein* », du nom de chacun des systèmes d’intelligence artificielle mis en place par ces deux entreprises. Il s’agit d’une exploitation des technologies

---

de manière associée, à travers une interface de programmation applicative (souvent désignée par le terme API, soit *Application Programming Interface*).

Vos rapporteurs observent qu'Intel développe sa plateforme de solutions en intelligence artificielle Nervana. L'entreprise Intel représente, à elle seule, 97 % des serveurs de *data centers* opérant dans le domaine de l'intelligence artificielle. Ils relèvent qu'une **alliance stratégique a été nouée entre Google et Intel** afin de fournir une infrastructure multiple sur le *cloud*, ouverte, flexible et sécurisée.

Dans le secteur médical, des « **Uber** » de la santé pourraient gagner une place de premier plan, le site « *mondocteur.fr* » constituant un exemple de plateforme à fort potentiel marquée par la convergence NBIC.

L'analyse de données et les outils de prédiction sont devenus des incontournables des offres informatiques pour les entreprises, avec l'hébergement dans le *cloud*. Amazon, Baidu, Microsoft et Google y travaillent depuis plusieurs années. Des sociétés spécialisées dans les services professionnels, comme Oracle, Salesforce ou SAP, s'y déploient. Mais, là où ses concurrents se concentrent sur un secteur ou un produit en particulier, IBM imagine Watson comme une plateforme capable de s'adapter à tous les besoins d'une entreprise.

Au début de l'année 2014, IBM avait annoncé consacrer un milliard de dollars pour transformer Watson en division indépendante. Deux ans plus tard, il ajoutait 3 milliards de dollars dans la balance, cette fois-ci consacrés à l'Internet des objets, puis 200 millions de dollars pour achever le centre de recherche de Munich. IBM a aussi investi des milliards de dollars dans des entreprises et *start-ups* spécialisées pour nourrir Watson de données diverses : dans la météorologie, la reconnaissance d'image, le domaine médical, etc. Le nombre d'employés affectés à **Watson frôle désormais les 10 000 personnes**. Selon l'entreprise, aujourd'hui, « à l'ère des systèmes cognitifs, les systèmes peuvent apprendre suite à des expériences, détecter des corrélations, créer des hypothèses, et aussi mémoriser les résultats et en tirer des enseignements ».

Uber a créé un laboratoire de recherche sur l'intelligence artificielle à San Francisco. L'entreprise a en effet racheté la *start-up* « Geometrics Intelligence » pour créer ce laboratoire nommé « Uber AI Labs ». Les applications potentielles sont innombrables. Mais à ce stade, le laboratoire aurait pour mission de se concentrer sur deux innovations en particulier : améliorer l'algorithme d'Uber pour faciliter la rencontre entre un chauffeur et un passager et trouver de nouvelles techniques pour construire un véhicule autonome performant. En août 2016, Uber ajoutait Otto à son actif, une jeune pousse qui développe des kits pour rendre les camions autonomes, un contrat de 680 millions de dollars.

Il convient d'observer que Google a également racheté, en 2013, Waze, la *start-up* israélienne à l'origine du fameux service d'information en matière de trafic routier en temps réel, pour 1 milliard de dollars.

---

**Les entreprises du secteur automobile achètent des entreprises du numérique ou nouent des partenariats** (Google-Fiat-Chrysler, Google-Honda, Amazon-Ford, BMW-IBM, BMW-Mobileye, Volvo-Uber, Tesla-Uber, Renault-Nissan-Microsoft). BMW se rapproche d'IBM pour travailler sur le système cognitif de Watson, les deux entreprises ont ainsi annoncé un partenariat au travers duquel elles travailleront sur le système cognitif de Watson pour améliorer la personnalisation de l'expérience de conduite. Et après le rapprochement entre Google et le groupe Fiat-Chrysler depuis mai 2016, c'est le constructeur japonais Honda qui s'est rapproché de Google-Waymo. Honda est déjà partenaire du service asiatique Grab qui teste des solutions similaires. Grab travaille sur des voitures autonomes capables de conduire seules sur les routes dès 2020 mais requérant la présence d'un conducteur, alors que Waymo veut se dispenser totalement du chauffeur.

Amazon a réussi à intégrer son système de domotique Alexa dans les voitures de Ford, ce qui permettra au propriétaire de contrôler les objets connectés de sa maison depuis sa voiture et vice-versa. Quant à Microsoft, il apporte les conférences audio de Skype dans les voitures Volvo, son assistant vocal Cortana chez BMW et Nissan. Les algorithmes et caméras intelligentes de Mobileye sont utilisés par BMW pour déployer une flotte de 40 véhicules autonomes cette année.

Pour conclure en ce qui concerne le secteur automobile, Rémi Cornubert, d'AT Kearney, explique qu'avant le monde automobile était « simple, très pyramidal : le constructeur était le donneur d'ordres qui travaillait avec des fournisseurs. Aujourd'hui, avec la révolution de la connectivité, de la voiture autonome et des services à la mobilité, les cloisons sautent, il est impossible pour un constructeur de tout maîtriser ». Carlos Ghosn, président de Renault-Nissan reconnaît qu'« on ne peut pas tout faire ».

Vos rapporteurs relèvent que **le patronat français s'inscrit dans la démarche de transformation en cours**. Le MEDEF a ainsi préparé en 2017 le projet « Métamorphose », qui vise à définir une pyramide de changement sur quatre niveaux : un **socle de mobilisation sur la révolution numérique**, des *big data* et des NBIC, un **socle de formation**, un **socle d'accompagnement**, avec les incubateurs et les écosystèmes du numérique et, enfin, **la pointe du financement**, permettant la transformation des PME.

### Le projet « Métamorphose » du MEDEF

L'enjeu des objets connectés permet de capter de nombreuses données, qui deviennent les ressources fondamentales de nombreux domaines en devenir, à l'instar de l'intelligence artificielle ou de la chaîne de blocs (*Blockchain*). A ce titre, l'intelligence artificielle était très présente au *Consumer Electronics Show* de Las Vegas en 2016. **L'intelligence artificielle se niche dans des applications utilisées au quotidien**, telles que les correcteurs orthographiques. L'enjeu essentiel de l'intelligence artificielle est d'alimenter l'apprentissage profond (*deep learning*) et l'apprentissage automatique (*machine learning*), auxquels les « GAFAMI » s'attèlent.

Le MEDEF est actuellement en pleine préparation du projet « Métamorphose ». Ce projet vise à définir une pyramide de la métamorphose bâtie sur quatre niveaux. Tout d'abord, cette pyramide repose sur un **socle de mobilisation sur la révolution numérique**, des *Big data* et des NBIC. Ensuite s'ajoute un **socle de formation**, notamment avec le travail mené sur les campus numérique qui fournissent des outils d'évaluation de numérisation des entreprises. Au-dessus, se trouve le **socle d'accompagnement** : incubateurs, écosystèmes de *start-ups* du numérique, pour accompagner les petites et moyennes entreprises (PME) qui ont besoin de se transformer. Enfin, au sommet se trouve **la pointe du financement**, qui a pour objectif la création de fonds de financement permettant la transformation des PME.

**Ce projet vise deux cibles : les cent mille PME qui doivent se transformer, et les *start-up* françaises.** Il est impératif, selon le MEDEF, que les jeunes pousses se transforment et mutent en PME, voire en grands groupes internationaux. Ces entreprises émergentes regorgent de talents et de chercheurs formidables ; il est dommage que la France se fasse piller de ses talents. Il est donc indispensable d'aider les *start-ups* à se transformer grâce à un financement français. Cela implique de garder les expertises et centres de décision en France, malgré la mondialisation de ce domaine, et que certaines réformes se fassent.

Un second niveau, au niveau européen, est également nécessaire, au travers de la coopération de partenaires tels que l'Allemagne.

Source : OPECST suite à l'intervention de Pierre Gattaz, président du MEDEF, lors de la journée « Entreprises françaises et intelligence artificielle » organisée par le MEDEF et l'AFIA le 23 janvier 2017

## 2. Un risque de redéfinition, sous l'effet de ce nouveau contexte économique, des rapports de force politiques à l'échelle mondiale

Dans ce nouveau contexte économique, **les rapports de force politiques pourraient être progressivement bouleversés** à une échelle mondiale. Le poids pris par les grandes entreprises privées plateformes, GAFAMITIS<sup>1</sup>, NATU<sup>2</sup> et BATX<sup>3</sup> fait courir d'importants risques aux systèmes démocratiques que nous connaissons.

**La question dépasse le sujet de la colonisation numérique et celui de la domination américaine, voire chinoise, ou de pays émergents comme l'Inde.** Ces questions devront faire l'objet d'une grande attention des

<sup>1</sup> Google, Apple, Facebook, Amazon, Microsoft, IBM, Twitter, Intel et Salesforce.

<sup>2</sup> Netflix, Airbnb, Tesla et Uber.

<sup>3</sup> L'expression désigne les géants chinois du numérique : Baidu, Alibaba, Tencent et Xiaomi.



---

pouvoirs publics, car les mouvements de concentration capitalistique qui sont en cours vont dans le sens décrit précédemment, à savoir celui d'une économie globalisée dominée par des « plateformes ». Les monopoles résultant d'une concentration horizontale puis verticale pourraient disposer d'une **puissance à l'échelle mondiale sans équivalent historique**.

### 3. Des bouleversements annoncés dans le marché du travail : perspectives de créations, d'évolutions et de disparitions d'emplois

La **résolution relative aux règles de droit civil sur la robotique** adoptée le 16 février 2017 par le Parlement européen, qui fait suite au rapport de la députée européenne Mady Delvaux, estime, dans l'un de ses considérants, « *que l'utilisation généralisée de robots pourrait ne pas entraîner automatiquement une destruction d'emplois, mais que des emplois moins qualifiés dans les secteurs à forte intensité de main-d'œuvre risquent d'être plus vulnérables à l'automatisation* ». Animé par la même préoccupation, le réseau de l'*European Parliamentary Technology Assessment* (EPTA), auquel appartient l'OPECST, a consacré en 2016 sa conférence annuelle au thème de l'impact des nouvelles technologies numériques et robotiques sur le marché du travail. L'ensemble des contributions présentées par les dix-sept organes d'évaluation scientifique membres de l'EPTA ont été regroupées sous le titre « *L'avenir du travail à l'ère numérique* »<sup>1</sup>.

**Sur cette question hautement sensible, vos rapporteurs ont entendu des pronostics très contrastés lors de leurs auditions.** Selon Cédric Sauviat, président de l'association française contre l'intelligence artificielle (AFCIA), « *ce ne sont pas seulement les emplois de la classe moyenne et des ouvriers qui sont menacés, mais également des emplois d'expertise, comme les emplois de médecins ou d'avocats. L'intelligence artificielle touche des domaines de plus en plus vastes ; de fait, toutes les couches du marché du travail seront concernées par l'intelligence artificielle* ». En sens inverse, Marie-Claire Carrère-Gée, présidente du conseil d'orientation pour l'emploi (COE) estime que le rôle de cette instance est « *d'anticiper, autant que faire se peut, les conséquences du progrès technologique en cours sur l'emploi, dans un contexte où le débat public est marqué par des études donnant au pourcentage près le nombre d'emplois détruits. La situation est très anxiogène. Certes, la crainte du chômage est un grand classique à chaque vague d'innovation technologique. Keynes lui-même avait prédit un chômage massif. Pourtant, l'histoire montre que depuis toujours le progrès technologique a créé des emplois* ».

Ces divergences d'analyses s'expliquent dans une large mesure par des **différences d'approches méthodologiques** et par la difficulté inhérente à toute démarche prospective dans ce domaine éminemment mouvant : finalement, ce sont les convictions personnelles qui l'emportent lorsqu'il

---

<sup>1</sup> Pour consulter les actes de la conférence annuelle 2016 de l'EPTA : <http://epub.oecw.ac.at/ita/ita-projektberichte/EPTA-2016-Digital-Labour.pdf>



---

s'agit d'apprécier globalement l'incidence dans un futur probable de l'intelligence artificielle sur le marché du travail.

Pourtant, l'impact des innovations technologiques sur le volume et la nature des emplois est loin d'être un sujet d'étude nouveau pour la science économique. Depuis les débuts de la révolution industrielle, le débat académique oppose les « **techno-optimistes** » et les « **techno-pessimistes** ». La littérature économique n'aboutit pas à des conclusions univoques quant aux effets des évolutions technologiques sur l'emploi. **Une première distinction est à faire entre les approches micro-économique, sectorielle et macro-économique.** Les préoccupations en termes de gains de productivité et de réduction de la masse salariale sont prédominantes au niveau de chaque entreprise prise isolément, mais les suppressions d'emplois qui en résultent peuvent être, au moins, compensées au niveau global de la branche ou de l'économie nationale, grâce à des effets de baisse des prix et de hausse de la demande. **Une seconde distinction est à faire selon la nature des innovations technologiques.** Une **innovation de production** va permettre de produire davantage avec moins de travail. Mais une **innovation de produit ou de service** peut créer des emplois, en étant à l'origine d'un nouveau marché.

La théorie économique ne tranchant pas ce débat vieux de plus de deux siècles, il est instructif d'examiner les principales conclusions empiriques des **études rétrospectives** qui ont tenté d'évaluer les effets des vagues d'innovation successives au cours des trente dernières années. Le récent rapport du Conseil d'orientation pour l'emploi (COE) intitulé « *Automatisation, numérisation et emploi* »<sup>1</sup> recense et analyse ainsi seize études rétrospectives parues sur ce sujet entre 2009 et 2016. Celles-ci font apparaître très nettement un **double processus de destruction et de créations d'emploi**. Elles soulignent également l'importance des effets indirects positifs sur l'emploi des innovations de produits, qui viennent plus que contre-balancer les effets directs négatifs des innovations de procédés.

Ces études rétrospectives se heurtent à une **question de méthodologie délicate** : comment apprécier les effets distincts des technologies numériques et robotiques sur l'emploi, par rapport à d'autres progrès technologiques ou à d'autres facteurs d'évolution du marché du travail ? Les données statistiques de base qui le permettraient ne sont pas toujours disponibles, et ce genre d'appréciation comporte inévitablement une part de convention. L'OCDE a publié en 2016 une étude de M. Arntz, T. Gregory et U. Zierahn sur « *Les risques de l'automatisation pour l'emploi* »<sup>2</sup>, analysant comparativement 18 États membres, dont la France, sur la période 1990-2012. Elle parvient à la conclusion que les investissements en technologies de l'information et de la communication n'ont pas d'effets

---

<sup>1</sup> Pour consulter le rapport du Conseil d'orientation pour l'emploi : [http://www.coe.gouv.fr/IMG/pdf/COE\\_170110\\_Rapport\\_Automatisation\\_numerisation\\_et\\_emploi\\_Tome\\_1.pdf](http://www.coe.gouv.fr/IMG/pdf/COE_170110_Rapport_Automatisation_numerisation_et_emploi_Tome_1.pdf)

<sup>2</sup> Pour consulter le rapport de l'OCDE : [http://www.oecd-ilibrary.org/social-issues-migration-health/the-risk-of-automation-for-jobs-in-oecd-countries\\_5jlz9h56dvq7-en](http://www.oecd-ilibrary.org/social-issues-migration-health/the-risk-of-automation-for-jobs-in-oecd-countries_5jlz9h56dvq7-en)

---

négatifs sur l'emploi dans ces pays, au niveau agrégé, compte tenu des phénomènes de compensations. **Pour l'avenir, cette étude de l'OCDE estime qu'en moyenne 9 % des emplois au sein de ces pays membres sont menacés par l'automatisation.**

D'autres études prospectives parues ces dernières années sur le même sujet se sont focalisées sur le risque de destruction d'emplois lié aux avancées de la numérisation et de l'automatisation. Une étude originale et fondatrice de la *Oxford Martin School*, conduite par Carl Benedikt Frey et Michael A. Osborne<sup>1</sup>, a suscité en 2013 une forte inquiétude en avançant que l'automatisation représentait un **risque pour 47 % des emplois aux États-Unis**. En 2014, une étude du cabinet de conseil McKinsey estime que **60 % de tous les métiers pourraient voir automatiser 30 % de leurs activités**.

En 2014 également, une étude du cabinet Roland Berger<sup>2</sup> estime, pour la France, que **42 % des métiers présentent une probabilité d'automatisation forte** du fait de la numérisation de l'économie, et que 3 millions d'emplois pourraient être détruits par la numérisation à l'horizon 2025. Cette étude relève que, désormais, les métiers manuels ne sont pas les seuls à être automatisables, mais que des tâches intellectuelles de plus en plus nombreuses pourront faire l'objet d'une prise en charge par les outils numérique.

Lors de sa réunion à Davos en janvier 2017, le Forum économique mondial a également rendu public un pronostic inquiétant, en considérant que la « *quatrième révolution industrielle* », celle de l'intelligence artificielle, des objets connectés et de l'impression 3D, suscite de nouveaux risques mondiaux et tend à détruire plus d'emplois qu'elle n'en crée. Ce qui se traduirait par une **perte nette de 5,1 millions d'emplois dans 15 économies nationales**, dont la France, d'ici à 2020.

**Vos rapporteurs invitent à ne pas céder au « techno-pessimisme » face à cette succession de rapports alarmistes. Ces prévisions leur paraissent trop axées sur le seul aspect des destructions d'emplois, qui est le plus facile à évaluer, et insuffisamment sur celui des créations d'emplois, que les études rétrospectives invitent à ne pas négliger, même si leur caractère indirect les rend plus difficile à quantifier.** A cet égard, une variable essentielle pour l'ampleur et le sens positif ou négatif des effets sur l'emploi de cette « *quatrième révolution industrielle* » est le décalage toujours possible entre le rythme de diffusion des innovations liées à l'automatisation, la numérisation et l'intelligence artificielle au sein de l'économie et de la société, et la temporalité des réactions d'adaptation sur le marché de l'emploi. **Même si les phases de transition peuvent être**

---

<sup>1</sup> Pour consulter l'étude de la Oxford Martin School :

[http://www.oxfordmartin.ox.ac.uk/downloads/academic/The\\_Future\\_of\\_Employment.pdf](http://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf)

<sup>2</sup> Pour consulter l'étude du cabinet de conseil Roland Berger :

[https://www.rolandberger.com/publications/publication\\_pdf/les\\_classes\\_moyennes\\_face\\_la\\_transformation\\_digitale\\_roland\\_berger.pdf](https://www.rolandberger.com/publications/publication_pdf/les_classes_moyennes_face_la_transformation_digitale_roland_berger.pdf)

---

**déliçates, vos rapporteurs sont convaincus qu'à terme ces innovations multiples, qui se renforcent mutuellement, auront globalement un effet positif sur le volume total des emplois, comme sur l'intérêt des tâches professionnelles et la qualification des métiers.**

Dans le cas de la France, l'étude prospective que le Conseil d'orientation pour l'emploi a présentée en janvier 2017, précédemment évoquée, a vocation à devenir une référence par son ampleur et sa précision méthodologique. **Cette étude conclut que l'automatisation et la numérisation devraient avoir un impact relativement limité en termes de créations ou suppressions d'emplois, mais probablement important sur la structure des emplois et le contenu des métiers.**

Depuis les années 1980, l'évolution de la structure des emplois en France a profité surtout aux plus qualifiés. Les nouvelles technologies sont plus facilement substituables aux emplois auxquels sont associés des **tâches manuelles et cognitives routinières**, correspondant à des niveaux de qualification intermédiaires. Le **mouvement de complexification des métiers existants** sous l'effet des technologies nouvelles, est marqué par un essor des compétences analytiques et relationnelles. L'évolution des compétences demandées sur le marché du travail est aussi induite par l'émergence de nouveaux métiers dans le domaine du numérique, qui recouvrent des tâches nouvelles et plus complexes.

La situation actuelle du marché de l'emploi au regard des nouvelles technologies place la France dans une position intermédiaire en Europe. Selon les données d'Eurostat<sup>1</sup>, environ 8 millions de personnes étaient employées comme spécialistes des technologies de l'information et de la communication dans l'Union européenne, soit 3,5 % de l'emploi total. Toutefois, ce taux moyen recouvre des disparités importantes entre les pays, le secteur des TIC représentant plus de 6 % de l'emploi en Finlande et en Suède, mais seulement 1,2 % en Grèce. **La France se situe légèrement au-dessus de la moyenne européenne, avec un taux de 3,6 %.** Le secteur de la robotique est encore très marginal en termes d'emplois. Selon les chiffres de la Fédération internationale de robotique, il représenterait environ 300 000 emplois directs dans le monde. En France, la filière robotique regroupait en 2012 un peu plus de 7 000 emplois, selon les chiffres du Syntec numérique.

La méthodologie retenue par le COE dans son étude repose sur une approche qualitative et sectorielle, fondée sur les données issues de l'enquête « Conditions de travail » de la DARES, et est décomposée par tâche et par métier. Cette étude vise à affiner les approches purement macroéconomiques, et montre que :

**- moins de 10 % des emplois existants présentent un cumul de vulnérabilités susceptibles de menacer leur existence dans un contexte de numérisation et d'automatisation ;**

---

<sup>1</sup> « Les compétences numériques », *Communiqué de presse d'Eurostat n° 207/2016, octobre 2016.*

---

- la moitié des emplois existants est susceptible d'évoluer dans leur contenu de manière significative ou très importante ;

- le progrès technologique devrait continuer à favoriser plutôt l'emploi qualifié et très qualifié.

Vos rapporteurs relèvent que des emplois très qualifiés, du type statisticiens, analystes et conseillers financiers ou bancaires, médecins, notamment radiologues, etc., sont d'ores et déjà impactés par le recours à des applications intelligentes, agiles, interconnectées et ultra rapides. Il en va de même pour les métiers recourant à des techniques de simulation ou de projection dans les domaines de l'environnement, de l'urbanisme, de la gestion des flux, des transports et, d'une manière générale, de tous les systèmes complexes.

L'étude prospective du COE tente également d'anticiper les effets possibles de l'automatisation et de la numérisation sur la **localisation des emplois**. Ces effets apparaissent eux aussi diversifiés. En abaissant le coût de la distance géographique, les technologies de l'information et de la communication ont jusqu'à présent favorisé la délocalisation de certaines activités routinières industrielles et de services vers des pays à faibles coûts de main-d'œuvre. En sens inverse, l'automatisation pourrait atténuer cette tendance, voire l'inverser en incitant à des **relocalisations d'activités** précédemment déplacées vers les pays émergents. Toutefois, même si certaines prémisses de ce phénomène de relocalisation sont perceptibles, celui-ci est encore loin de constituer un mouvement de grande ampleur.

Enfin, toujours dans une approche en termes de localisation des emplois, l'étude du COE estime que ces nouvelles technologies devraient avoir un **impact différencié sur la répartition géographique des emplois sur le territoire national**. Les régions du territoire national dans lesquelles les secteurs industriels traditionnels faiblement intensifs en technologie représentent une grande part de l'emploi, et caractérisées par une forte densité en travailleurs peu qualifiés, sont les plus exposées au risque des destructions d'emplois sous l'effet de l'automatisation. Inversement, les aires urbaines, et notamment les métropoles, pourraient bénéficier de certains « effets d'économies d'agglomération », ainsi que de leurs réserves en emploi pour les compétences complémentaires des nouvelles technologies.

Vos rapporteurs soulignent que le débat sur l'impact du numérique en matière d'emploi n'est pas propre à l'intelligence artificielle : la question se pose beaucoup plus largement, puisque les conséquences du progrès technologique sur le marché du travail sont particulièrement visibles depuis la révolution industrielle, et n'ont d'ailleurs pas débuté avec celle-ci... Le développement de l'informatique et de la robotique implique la disparition de certains emplois, souvent peu qualifiés, mais en crée aussi de nouveau, plus qualifiés.

Les études disponibles ne s'intéressent souvent qu'aux destructions brutes d'emplois, alors qu'il y aura très probablement d'importantes créations d'emplois, non identifiées à ce jour. Dès 1964, un manifeste

---

alarmiste adressé au Président des États-Unis, Lyndon B. Johnson, et signé de plusieurs prix Nobel et chercheurs avait dénoncé une « *large vague de chômage technologique* » causée par « *la combinaison d'ordinateurs et de machines automatiques et autonomes* ».

En 1981, Alfred Sauvy, dans son livre « *Informatisation et emploi* » affirmait : « *Ne vous plaignez pas que le progrès technique détruisse des emplois, il est fait pour ça !* ». Vos rapporteurs jugent la formule un peu provocatrice dans un contexte de chômage élevé, mais l'estiment pertinente en période de plein emploi. La robotisation pousse les travailleurs vers des métiers de plus en plus intéressants, dès lors que des emplois nouveaux apparaissent et que les premiers ont les compétences nécessaires pour les occuper, ou sont aptes à les acquérir. **Vos rapporteurs estiment donc indispensable et urgent d'adapter le système éducatif à ces nouveaux métiers et de développer une offre de formation professionnelle adéquate, afin de garantir aux travailleurs la souplesse de reconversion dont ils ont besoin.**

Il y aura aussi d'importantes **évolutions en contenu des emplois et des tâches** et, au-delà de la stricte robotisation, il existera de plus en plus de **coopération hommes-machines**, à travers des interfaces hommes-machines. Cette coopération devrait être heureuse.

Votre rapporteure Dominique Gillot propose de **réfléchir, en particulier, à un nouveau mode de financement de notre système de protection sociale, qui serait un complément des cotisations existantes et qui pourrait consister en un prélèvement additionnel aux cotisations sociales, soit sur les robots, soit sur les activités automatisées ou les agents autonomes, dans la mesure où ces dispositifs remplacent des emplois précédemment occupés par des êtres humains.** Elle juge une telle proposition plus précise que celle de taxe sur les robots qui circule à la suite de la première version du rapport de Mady Delvaux<sup>1</sup>. Selon votre rapporteure, il faut que le produit de cette imposition n'aille pas au budget de l'État et finance plutôt les régimes sociaux, en particulier les assurances chômage, maladie et vieillesse.

Votre rapporteur Claude de Ganay se déclare, quant à lui, **opposé à toute taxe spécifique sur les activités automatisées**, intelligence artificielle comme robots, remplaçant des emplois occupés par des êtres humains, y compris un prélèvement assis sur les cotisations sociales. Un mécanisme de ce type constituerait, selon lui, un **mauvais signal, pouvant avoir comme effet pervers de décourager la recherche, l'innovation et l'activité économique.** La TVA et l'IS s'appliquent déjà à ces activités. Quand ils ne le font pas parfaitement, il est toujours possible d'y veiller en étendant l'assiette de ces impôts. **Il n'y a donc pas lieu de les taxer davantage en instaurant un nouveau prélèvement qui leur serait spécifique.**

---

<sup>1</sup> Bill Gates a fait cette proposition plus récemment <https://www.weforum.org/agenda/2017/02/bill-gates-this-is-why-we-should-tax-robots>

---

## B. LA SOCIÉTÉ EN MUTATION SOUS L'EFFET DE L'INTELLIGENCE ARTIFICIELLE

### 1. Les défis lancés par l'intelligence artificielle aux politiques d'éducation et de formation continue

Paul Dumouchel et Luisa Damiano, dans *Vivre avec les robots, Essai sur l'empathie artificielle*, estimaient en 2016 que « *vivre avec les robots peut être l'occasion d'un avenir meilleur, non pas économiquement mais moralement et humainement* ». **L'éducation peut être un facteur à la fois levier et bénéficiaire des avancées en intelligence artificielle.**

Intel a, par exemple, noué un partenariat avec Coursera, une vaste plateforme d'enseignement en ligne, pour justement y proposer des cours en intelligence artificielle dès 2017. L'approche de l'entreprise consiste à offrir une offre allant du logiciel au matériel (processeur), en passant par la formation.

La **relation émetteur-récepteur est transformée et modifie tant la pédagogie que les principes d'évaluation**. Les moyens de **prédire la réussite des élèves et d'optimiser les enseignements** seront précisés par les systèmes d'intelligence artificielle. Ces derniers permettront la différenciation des méthodes d'apprentissages, voire des contenus enseignés, la **personnalisation** devant être adaptée à la diversité des élèves. Jean-Marc Merriault, directeur général de Canopé, a donné l'exemple du projet *E-fran*, qui liera aussi bien des pédagogues que des chercheurs et des *start-uppers*. *Matador*, qui était à la base un jeu de plateau transformé dans un environnement numérique, a ainsi été développé. Le projet repose sur un monitoring individuel d'apprentissage du calcul mental par chaque élève. Plus l'élève jouera en classe et à la maison et mieux l'enseignant connaîtra ses compétences acquises et non acquises. Il s'agit de travailler sur le parcours d'un élève qui sera mis en rapport avec tous les autres élèves de son niveau scolaire. L'interaction et l'horizontalité constituent par conséquent des éléments importants de ce type de projet. *E-fran* reposera sur mille cinq cents élèves pendant une année scolaire, avec l'objectif d'analyser plus de sept cent mille opérations chaque année. Selon les profils, des parcours de jeu spécifiques seront proposés à chaque élève. L'ensemble s'appuie sur des chercheurs, aussi bien statisticiens que didacticiens et cognitivistes.

Jean-Marc Merriault a également indiqué à vos rapporteurs qu'un **continuum pédagogique entre le temps scolaire et le hors temps scolaire** devrait émerger, à travers la présence future de systèmes d'intelligence artificielle, de robots ou de *bots* aussi bien à l'école qu'au sein de la maison, il s'agira d'« *accompagner d'outils et d'interfaces pour assurer les usages au sein et en-dehors de la classe. Sur ce point, l'intelligence artificielle peut sûrement apporter un certain nombre de réponses* ».



---

Selon vos rapporteurs **les nouvelles technologies ne seront pas en compétition avec les enseignants**, mais elles leur seront **complémentaires**. Jean-Marc Merriaux lors de son audition a également insisté sur la mutation en cours de la place et du rôle de l'enseignant sous l'effet de l'ensemble des évolutions évoquées, ainsi que sur la nécessité de l'accompagner dès lors que les nouvelles technologies seront parties intégrantes de son enseignement. L'intelligence artificielle intervient pour **compléter le savoir-faire de l'enseignant, en le rendant plus accessible et mieux informé**.

Les **cours en ligne ouverts et massifs**, ou MOOC (*massive open online courses* en anglais) seront, de ce point de vue, des ressources utilisables pour appliquer ces nouvelles méthodes pédagogiques innovantes et **permettre aux jeunes générations d'accéder dans des conditions optimales à la connaissance**, ce dont se félicitent vos rapporteurs.

## **2. Une révolution potentielle de notre cadre de vie et de l'aide aux personnes**

Vos rapporteurs sont convaincus de la **possibilité imminente d'une révolution de notre cadre de vie et de l'aide aux personnes**. Des changements profonds sont à venir dans la connaissance et dans le contrôle de notre environnement et de la santé des populations. Les *smart grids* (réseaux d'énergie permettant une consommation optimisée) et les *smart cities* (villes intelligentes) seront les expressions des bénéfices que nous pouvons tirer de l'intelligence artificielle. Et cela se traduira évidemment en matière de transports, de sécurité, de santé, de dépendance et de handicap. Notre cadre de vie, la qualité de nos vies seront améliorés par l'usage massif de technologies d'intelligence artificielle.

Il sera également possible de **demander de plus en plus de choses à nos assistants personnels**, ils pourront répondre à beaucoup des questions que nous nous posons dans notre vie quotidienne. Ils pourront de mieux en mieux comprendre, organiser et hiérarchiser l'information et la connaissance. Le cabinet d'études Gartner prévoit que 50 % des applications analytiques embarqueront des fonctions d'intelligence artificielle d'ici trois à cinq ans et qu'une large part des analyses sortant de ces applications sera glanée *via* des interactions vocales (chiffre annoncé lors du symposium annuel du cabinet en octobre 2016).

S'agissant des **économies d'énergie**, Google a ainsi l'objectif d'optimiser la consommation de ses serveurs et de la réduire drastiquement grâce à l'intelligence artificielle.

Le calendrier de **déploiement des voitures autonomes** reste incertain (5, 10, 15, 20, 30 ans ?) mais il est plus que probable que la conduite automobile sera dans le futur une activité réservée à des passionnés nostalgiques. La recherche avance en la matière, mais moins vite que la confiance dans les technologies. Ce n'est pas tant pour des raisons techniques que pour des raisons d'oppositions sociales, notamment en

---

France, que le passage aux voitures autonomes risque de tarder. Il peut être constaté que les véhicules collectifs autonomes, tels que les bus, sont mieux acceptés et sont d'ailleurs déjà en service. Pourtant, les tests réalisés par Google-Waymo (3,5 millions de kilomètres parcourus, principalement en zone urbaine) illustrent la fiabilité croissante de ces véhicules. Les constructeurs automobiles déploient de plus en plus de projets en la matière (en particulier Renault, PSA, Volkswagen, Audi, BMW, Daimler-Benz, Honda, Toyota, GM, Ford, Chrysler, Fiat, Volvo...)

Votre rapporteur Claude de Ganay a été surpris par des propos du grand-maître de la confrérie « bérouttes et traditions » de Cernoy-en-Berry qui a expliqué, lors d'une réunion, **tous les bienfaits que les robots et les systèmes d'intelligence artificielle pourraient avoir pour la ruralité, en particulier pour les personnes âgées, isolées ou dépendantes**. Le cas des voitures autonomes a été d'abord évoqué, mais d'autres applications utiles vont émerger.

En matière de **handicap**, la *start-up* américaine BrainRobotics a présenté au salon d'électronique CES de Las Vegas une prothèse de main capable d'interpréter les signaux envoyés par les muscles résiduels de l'utilisateur : le système d'intelligence artificielle repère certaines caractéristiques de ces signaux musculaires, comme leur ampleur par exemple, et les transmet à travers un algorithme de classification qui sépare les différents types de geste (poing fermé, index levé, etc.). Il transmet ensuite celui qu'il a identifié et son intensité au moteur de l'appareil.

L'intelligence artificielle peut aussi venir en aide aux **malvoyants**. Facebook et Microsoft, notamment, ont dévoilé l'an passé des systèmes capables de voir des images et d'en décrire le contenu pour les aveugles.

Hyundai ambitionne de son côté de s'attaquer à la **paralysie** avec des exosquelettes robotisés. Un prototype montré au Consumer Electronics Show (CES) à Las Vegas s'adresse plus particulièrement aux paraplégiques, auxquels il rend la capacité de se lever, marcher ou monter des escaliers. L'appareil, baptisé H-MEX, longe le bas de la colonne vertébrale et tout l'arrière des jambes, avec des attaches au niveau de la taille, des cuisses, des genoux et des pieds. Un système de motorisation permet de déclencher des mouvements de rotation au niveau des articulations, depuis des boutons de commande placés sur les béquilles et par l'intermédiaire d'une connexion sans fil. Hyundai n'a pas encore de plan pour produire des appareils grand public, mais conduit des études cliniques dans des hôpitaux.

La *start-up* Japet de Lille a mis au point un **exosquelette lombaire Atlas**, destiné au milieu médical, et plus particulièrement aux centres de rééducation. Le dispositif repose sur quatre colonnes motorisées qui s'étirent pour décompresser la colonne verticale et soulager les douleurs lombaires. La commercialisation est visée pour la fin de l'année 2017 ou début 2018, et l'entreprise n'exclut pas de le décliner par la suite pour les problèmes au niveau des cervicales, ou pour les myopathies.



### 3. Le défi de la cohabitation progressive avec des systèmes d'intelligence artificielle dans la vie quotidienne

Les agents conversationnels ou *bots*, les robots de service, les agents d'assistance, d'accompagnement, d'aide à la mobilité vont progressivement cohabiter avec nous. Cela nécessitera une grande **vigilance**.

Au quotidien, nos *smartphones* avec des logiciels tels que « Siri », « Cortana » ou « Google Now », nous font d'ores et déjà cohabiter avec des algorithmes puissants, qui connaissent beaucoup d'aspects de chacun de nous.

À ce stade, les *bots* restent encore davantage des **systèmes de questions-réponses**, ils n'ont pas encore la mémoire des échanges et ne savent pas prendre en compte les aspects émotionnels, comme le souligne Laurence Devillers. Les *bots* devront savoir prendre en compte ces deux aspects et, demain, fonctionner sans connexion à Internet, ainsi que l'a expliqué à vos rapporteurs Alex Acero, directeur du projet Siri chez Apple. Aujourd'hui, les *bots* dépendent de l'existence d'une connexion au réseau, cette dépendance les rendant assez peu souples dans leur fonctionnement

Alors que la robotique industrielle connaît des applications selon la règle des 4 D : « *dangerous, dull, dirty, dumb* » (dangereux, ennuyeux, sale, idiot), alors que la robotique suit, comme le rappelle Laurence Devillers, la règle des 4 E : « *everyday, e-health, education, entertainment* » (accompagnement au quotidien, santé, éducation, loisirs).

**L'éducation et la prévention** sont indispensables dans ce contexte de cohabitation croissante entre les *bots* et les humains. Il convient d'apporter une grande attention aux logiques d'empathie et aux aspects émotionnels.

Yann LeCun se déclare convaincu que « *les machines intelligentes du futur auront des sentiments, des plaisirs, des peurs, et des valeurs morales (et que) ces valeurs seront une combinaison de comportements, d'instincts et de pulsions programmées avec des comportements appris* ». Cette perspective mérite que vos rapporteurs s'y attardent car, comme l'indiquent Rodolphe Gélén et Olivier Guilhem, respectivement directeur scientifique et directeur juridique d'Aldebaran puis de SoftBank Robotics, « *la modélisation des émotions est une tâche presque plus facile que l'ensemble des problèmes que les roboticiens ont eu à régler (...), le robot qui utilisera les techniques de perception des émotions pourra quasiment lire à livre ouvert les émotions de son interlocuteur. Elles s'ajouteront aux programmes qui, dès aujourd'hui, sont capables de détecter la joie, la tristesse, la colère, voire le sarcasme, dans la voix du locuteur. La perception d'émotions n'est donc pas ce qui différenciera longtemps l'homme du robot. Quant à l'expression d'émotions, les règles de la politesse élémentaires suffiront largement à en faire un compagnon suffisamment empathique et à le rendre aussi sympathique qu'un bon commerçant sachant jouer au moment opportun la joie, la tristesse, l'excitation ou l'abattement pour s'accorder à son client même si son émotion personnelle est à l'opposé de ce qu'il doit exprimer pour respecter la bienséance. Le robot n'ayant pas*

---

*d'émotion personnelle, il lui est encore plus facile d'exprimer celle que son interlocuteur attend (...) cette façon de jouer sur les émotions pourrait se rapprocher d'une forme de manipulation. Un programmeur doté de quelques connaissances en psychologie, et ils seront de plus en plus nombreux à y être formés, pourra jouer sur l'état émotionnel d'une personne pour la convaincre de prendre ses médicaments, de ne pas boire un verre de plus ou d'aller se coucher quand il est très tard ».*

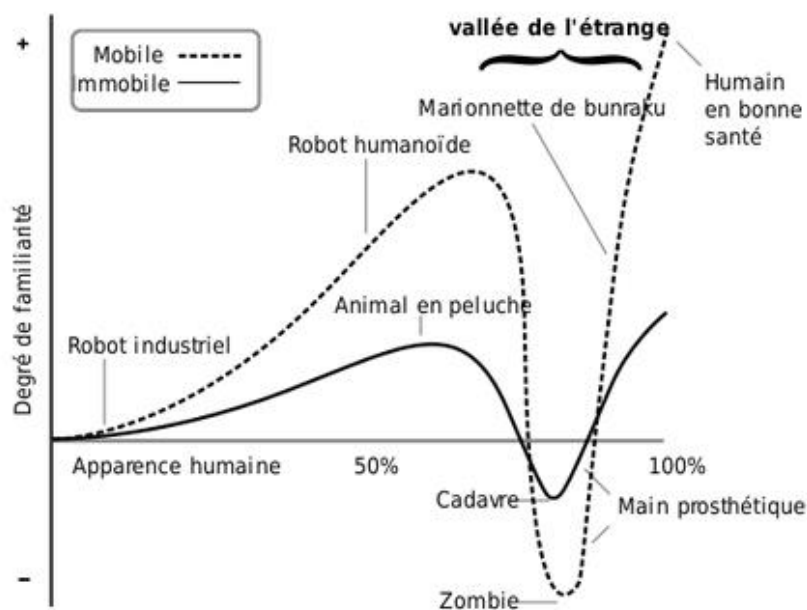
Pour le psychanalyste Serge Tisseron, spécialisé en intelligence artificielle et en robotique<sup>1</sup>, **utiliser ces techniques pour faire croire que les robots seront capables de sentiments serait malhonnête** car la reconnaissance des émotions présente avant tout pour eux l'intérêt de permettre de s'adapter à l'état d'esprit de l'utilisateur et, partant, à le tromper. En donnant l'impression que le robot a des émotions (puisqu'il en exprime), son programmeur trompe l'interlocuteur et peut encourager la création d'un lien affectif disproportionné entre l'homme et la machine.

En outre, la cohabitation avec des machines pose la question de la « **vallée de l'étrange** » ou *Uncanny Valley* (littéralement : vallée dérangement). Ce champ de recherche scientifique, initié par le roboticien japonais Masahiro Mori en 1970, plaide pour les reproductions de certaines caractéristiques du vivant sans chercher une trop grande ressemblance, source de trouble et de gêne, encore difficiles à expliquer. Cette expression imagée peut être représentée graphiquement, comme l'illustre le document suivant, cette vallée étant symbolisée par la zone de perception négative ressentie par un observateur humain face à un robot humanoïde ou un zombie. Cette représentation graphique de la théorie de la vallée de l'étrange pose en abscisse le degré d'apparence humaine (de zéro à 100 %) et en ordonnée le degré de familiarité et/ou d'acceptation.

---

<sup>1</sup> « *Le jour où mon robot m'aimera* », 2015.

### Représentation graphique de la théorie de la vallée de l'étrange



Source : [creativecommons.org/licenses/by-sa/3.0/](https://creativecommons.org/licenses/by-sa/3.0/)

Dans sa nouvelle de science-fiction « *L'histoire du robot et du bébé* », John McCarthy, l'un des pères fondateurs de l'intelligence artificielle, co-organisateur en 1956 avec Marvin Minsky de l'école d'été de Dartmouth, imagine qu'après l'émergence de l'intelligence artificielle forte et avec la vaste pénétration des robots dans la vie quotidienne, les robots d'assistance maternelle, pourtant conçus en conformité avec le principe éthique de « non-figuration humaine », devenu un des articles du code des robots, prennent une place de plus en plus grande : « *de nombreux enfants devinrent plus attachés à leur robot nounou qu'à leurs vrais parents. On y remédia en rendant les robots nounous plus durs et en aidant les parents à rivaliser avec eux pour capter l'amour de leurs enfants. Quelquefois, ça marchait. Dans un deuxième temps, les robots furent programmés de telle sorte que plus les parents étaient sympas, plus le robot l'était, tout en laissant les parents gagner la compétition pour obtenir l'affection de leurs enfants. Le plus souvent, ça marchait* ».

**Cet enjeu de la cohabitation quotidienne avec des intelligences artificielles, de leur acceptation et de leur régulation doit être appréhendé.**

---

## II. LES QUESTIONS ÉTHIQUES ET JURIDIQUES POSÉES PAR LES PROGRÈS EN INTELLIGENCE ARTIFICIELLE

### A. LES ANALYSES PRÉSENTÉES PAR D'AUTRES INSTANCES POLITIQUES

#### 1. Les deux rapports issus des institutions de l'Union européenne : Parlement européen et Comité économique et social européen (CESE)

Du côté des institutions de l'Union européenne, il convient tout d'abord de relever l'initiative conduite par Mady Delvaux, présidente du groupe de travail sur la robotique et l'intelligence artificielle au sein de la Commission des affaires juridiques du Parlement européen. Elle a ainsi rendu public, le 31 mai 2016, un **projet de rapport contenant des recommandations à la Commission européenne concernant des règles de droit civil sur la robotique et une motion portant résolution du Parlement européen**. Ce document, sous la forme d'une résolution du Parlement européen a été adopté le 16 février 2017<sup>1</sup>, dans une version « allégée » à l'issue de sa discussion (plus de 500 amendements ont été déposés). La Commission n'est pas contrainte de suivre les recommandations du Parlement mais elle doit exposer ses raisons en cas de refus.

**L'idée de taxer les robots ou celle de mettre en place un revenu universel font partie des propositions qui ont été écartées lors du vote des amendements.** Par ce texte, les députés européens demandent à la Commission européenne de proposer des règles sur la robotique et l'intelligence artificielle, en vue d'exploiter pleinement leur potentiel économique et de garantir un niveau standard de sûreté, de sécurité et de transparence. Ils y soulignent que des normes communautaires pour les robots devraient être envisagées afin que l'Union européenne prenne l'initiative pour fixer ces normes sans être contrainte de suivre celles édictées par des États tiers.

Il est proposé de clarifier les questions de responsabilité, en particulier pour les voitures autonomes, et de mettre en place un système d'assurance obligatoire et un fonds de garantie permettant le dédommagement des victimes en cas d'accidents causés par ce type de voitures. Il est également proposé un code de conduite éthique pour la robotique, à destination notamment des chercheurs et des concepteurs.

Par ailleurs, les parlementaires européens demandent à la Commission d'envisager, à long terme, la possibilité de créer un statut juridique spécial pour les robots, sous la forme de l'octroi d'une personnalité juridique afin de clarifier la responsabilité en cas de dommages.

---

<sup>1</sup> La résolution a été adoptée par 451 voix pour, 138 voix contre et 20 abstentions.

---

En outre, la création d'une agence européenne pour la robotique et l'intelligence artificielle, afin de fournir aux autorités publiques une expertise technique, éthique et réglementaire, est préconisée.

Une consultation publique sur cette résolution du Parlement européen a été ouverte le 16 février 2017

Vos rapporteurs jugent nécessaires de distinguer clairement les robots « physiques » des robots « virtuels », ce que ne fait pas la résolution du Parlement européen. La confusion doit être levée et le terme de robot réservé à des objets matériels autonomes.

Vos rapporteurs désapprouvent également l'octroi d'une personnalité juridique pour les robots, qui leur paraît soit dépourvue de fondement, soit totalement prématurée.

Le Comité économique et social européen (CESE), l'assemblée consultative des partenaires économiques et sociaux européens, rend des avis qui sont publiés au Journal officiel de l'Union européenne, à raison d'environ 170 par an en moyenne. Catelijne Muller, que vos rapporteurs ont rencontrée, est en cours de finalisation d'un rapport sur l'intelligence artificielle en tant que rapporteure pour le CESE sur l'intelligence artificielle. L'avis sera rendu public à l'issue de la réunion de section prévue le 4 mai 2017. Vos rapporteurs ont pu constater que Mme Muller partageait largement leurs points de vue, analyses et propositions.

## 2. Les trois rapports de la Maison Blanche

La Présidence des États-Unis s'est emparée du sujet de l'intelligence artificielle et a rendu différents rapports dans la période récente, dont trois au cours des six derniers mois. Elle avait auparavant rendu public un **premier rapport en mai 2014 sur le big data**<sup>1</sup> sous-titré « *seizing opportunities, preserving values* », piloté par John Podesta, au sein du bureau exécutif du président des États-Unis<sup>2</sup> (en anglais *Executive Office of the President*, appelé également « *Brain Trust* »). Plus récemment, après avoir mobilisé des experts au sujet de l'intelligence artificielle, en mai 2016<sup>3</sup>, la Présidence des États-Unis a rendu public, en **octobre 2016**, un rapport de l'Office de la politique scientifique et technologique de la Maison Blanche intitulé « *Se préparer à*

---

<sup>1</sup> [https://obamawhitehouse.archives.gov/sites/default/files/docs/20150204\\_Big\\_Data\\_Seizing\\_Opportunities\\_Preserving\\_Values\\_Memo.pdf](https://obamawhitehouse.archives.gov/sites/default/files/docs/20150204_Big_Data_Seizing_Opportunities_Preserving_Values_Memo.pdf) et <http://www.cfr.org/technology-and-science/white-house-big-data---seizing-opportunities-preserving-values/p32916>

<sup>2</sup> Le « cabinet » est, aux États-Unis, la réunion des membres les plus importants de l'exécutif du gouvernement fédéral américain, ce qui est totalement différent d'un cabinet ministériel au sens où nous l'utilisons en français.

<sup>3</sup> Le Président des États-Unis a ainsi lancé un groupe de travail et un sous-comité spécifique au sein du National Science and Technology Council (NSTC), chargés de suivre les évolutions du secteur et de coordonner les activités fédérales sur le sujet. En parallèle ont été tenues quatre sessions de travail publiques entre les mois de mai et juillet 2016, visant à engager la discussion avec le grand public et surtout à produire une large évaluation des opportunités, risques, et implications réglementaires et sociales de l'intelligence artificielle et de ses nouveaux développements, cf. : <https://www.whitehouse.gov/blog/2016/05/03/preparing-future-artificial-intelligence>

---

*l'avenir de l'intelligence artificielle* »<sup>1</sup> (en anglais, *Preparing the Future of Artificial Intelligence*), accompagné d'un autre document intitulé « **Plan stratégique national pour la recherche et le développement de l'intelligence artificielle** »<sup>2</sup> (*National Artificial Intelligence Research & Development Strategic Plan*), ces deux rapports sont résumés ci-après. Enfin, en décembre 2016, le bureau exécutif du président des États-Unis a rendu un nouveau rapport sur **l'impact économique de l'intelligence artificielle et de la robotisation**<sup>3</sup> (il est intitulé « *Artificial Intelligence, Automation, and the Economy* »). L'administration Obama a donc produit pas moins de trois rapports sur le sujet entre octobre et décembre 2016.

Vos rapporteurs ont été marqués par le fait que le Président Barack Obama ait choisi de consacrer **l'une des dernières de ses grandes apparitions publiques** aux défis technologiques, sociaux, économiques et éthiques de l'intelligence artificielle, lors d'un entretien avec Joi Ito, directeur du MIT Media Lab, publié dans le magazine technologique le plus lu au monde<sup>4</sup>. Comme leur ont indiqué les fonctionnaires du service scientifique du consulat de France à San Francisco, il s'agit d'un bon indicateur du fait que « *cette thématique est le sujet brûlant du moment aux États-Unis, que ce soit dans l'expression de politiques publiques, la consolidation d'une puissance de recherche inégalée ou le développement rapide et sans précédent d'activités économiques liées à ces technologies* ».

Le **premier rapport** intitulé « *Se préparer à l'avenir de l'intelligence artificielle* », assez peu détaillé, a pour objectif de produire un **cadre général de réflexion**, en amorçant un état des lieux de la recherche et des applications actuelles tout en posant des jalons prudents quant à la possibilité et la nature d'une régulation. Il dresse moins un bilan de l'état actuel de l'intelligence artificielle, de ses applications effectives et potentielles, points rapidement évoqués, qu'il ne s'intéresse aux questions posées par les progrès en intelligence artificielle pour la société et les politiques publiques, tout en mettant en avant un discours mobilisateur autour du potentiel des avancées en intelligence artificielle pour permettre aux États-Unis de rester à la pointe de l'innovation mondiale. La publication de ce rapport suit une série d'activités publiques menées par l'Office de la politique scientifique et technologique de la Maison Blanche en 2016, qui comprenait cinq ateliers public co-animés tenus à travers le pays, ainsi qu'une demande d'information<sup>5</sup>, en juin 2016, à laquelle cent soixante-et-une réponses ont été apportées. Ces activités ont permis de faire connaître les sujets étudiés et les recommandations incluses dans le rapport.

---

<sup>1</sup> [https://obamawhitehouse.archives.gov/sites/default/files/whitehouse\\_files/microsites/ostp/NSTC/preparing\\_for\\_the\\_future\\_of\\_ai.pdf](https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf)

<sup>2</sup> [https://obamawhitehouse.archives.gov/sites/default/files/whitehouse\\_files/microsites/ostp/NSTC/national\\_ai\\_rd\\_strategic\\_plan.pdf](https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/national_ai_rd_strategic_plan.pdf)

<sup>3</sup> <https://www.whitehouse.gov/sites/whitehouse.gov/files/images/EMBARGOED%20AI%20Economy%20Report.pdf>

<sup>4</sup> 'Barack Obama, neural nets, self-driving cars and the future of the world', entretien du 24 août 2016 dans WIRED. <https://www.wired.com/2016/10/president-obama-mit-joi-ito-interview/>

<sup>5</sup> Cette procédure « Request for information » correspond à une consultation publique formalisée.

---

Ce dernier constate que depuis quelques années, **les machines ont surpassé les humains dans l'accomplissement de certaines tâches spécifiques**, par exemple dans la reconnaissance d'images. Bien qu'il soit peu probable que les machines présentent une intelligence d'application générale comparable ou excédant celle des humains dans les vingt prochaines années, les experts prévoient que les progrès rapides dans le champ de l'intelligence artificielle spécialisée se poursuivront, avec des machines égalant et dépassant les performances humaines sur de nombreuses tâches spécialisées.

L'un des plus importants enjeux de l'intelligence artificielle est son **impact sur l'emploi et l'économie**. Le rapport recommande que la Maison Blanche organise une étude sur l'impact de l'automatisation sur l'économie, ce qui a été fait avec la publication d'un rapport spécifique sur ce sujet en décembre 2016.

En particulier, ce premier rapport se conclut sur l'idée que les technologies d'intelligence artificielle font émerger un **potentiel d'amélioration de la vie des citoyens en ouvrant de nouveaux marchés et de nouvelles opportunités permettant de résoudre certains des grands enjeux sociétaux** : la santé, les transports, l'éducation, l'énergie, l'environnement, la justice, la sécurité ou encore l'efficacité du gouvernement. Il estime crucial que l'industrie, la société civile, et le Gouvernement travaillent ensemble et se mobilisent pour saisir pleinement ces opportunités.

Au-delà des domaines d'application mis en avant, il affirme qu'une **réglementation générale de la recherche en intelligence artificielle semble inapplicable à l'heure actuelle et que la réglementation actuelle est pour l'heure suffisante**, dans l'attente d'une expertise plus fouillée<sup>1</sup>. Ainsi, les éventuelles futures réglementations devront réduire les coûts et les barrières à l'innovation sans mettre en danger la sécurité du public ou la concurrence équitable sur le marché. Le Gouvernement fédéral peut cependant jouer un rôle de pivot, même sans impulser de nouvelles lois<sup>2</sup>.

Un second rapport, intitulé **plan national pour la recherche sur l'intelligence artificielle et le développement stratégique**, a également été

---

<sup>1</sup> L'analyse comparée des risques et des bénéfices devra permettre de justifier les futures évolutions législatives et réglementaires. En effet, le rapport insiste sur l'importance d'ajuster les prochains arbitrages selon le principe suivant : évaluer les risques que l'implémentation de l'IA pourrait réduire, de même que ceux qu'elle pourrait augmenter.

<sup>2</sup> Il a selon le rapport plusieurs rôles à jouer : un rôle d'organisateur du débat public et d'arbitre des mesures à mettre en place à l'échelle du pays ; un rôle de suivi attentif de la sécurité et de la neutralité des applications développées ; un rôle d'accompagnateur de la diffusion de ces technologies tout en protégeant certains secteurs au besoin afin d'éviter des contrecoups économiques dévastateurs ; un rôle de soutien et de financeur de projets de recherche faisant avancer le domaine ; et enfin un rôle d'adoption en son sein même de ces avancées afin d'assurer un service public de meilleure qualité. Il peut, en outre, veiller à la production d'une main d'œuvre en nombre suffisant, ainsi que d'un haut niveau de qualification et de diversité technique, non seulement d'un point de vue professionnel, mais également, et de manière plus large, du point de vue de la formation générale de la population.

---

publié en octobre 2016. Allant de pair avec le premier, il pose les **lignes directrices d'une stratégie nationale pour la recherche et le développement** de l'intelligence artificielle. Il apporte ainsi des recommandations pour des actions spécifiques en R&D financées au niveau fédéral. Il rappelle que le gouvernement fédéral des États-Unis a investi dans la recherche sur l'intelligence artificielle durant de nombreuses années, par exemple à travers la DARPA. Au vu des immenses opportunités qui se présentent, de nombreux facteurs sont à prendre en considération dans la définition de la recherche et du développement (R&D) de l'intelligence artificielle financée au niveau fédéral. Sept priorités ont donc été définies :

1. **Soutenir, par l'investissement fédéral, la recherche à long terme** afin de produire des percées scientifiques et technologiques dans les dix prochaines années et **demeurer le leader mondial de l'intelligence artificielle**. Il s'agit en particulier des méthodes nécessaires à la découverte de savoirs dans de grands volumes de données, l'amélioration des capacités de perception des systèmes d'intelligence artificielle, la compréhension profonde des capacités théoriques et des limites propres au développement de l'intelligence artificielle, la poursuite des efforts visant au développement d'une intelligence artificielle générale par opposition aux intelligences artificielles spécifiques.

2. Développer des **méthodes de collaboration entre hommes et intelligence artificielle**. Plutôt que de remplacer les humains, la plupart des systèmes d'intelligence artificielle vont collaborer avec eux : la recherche est nécessaire afin de créer des interactions effectives entre les humains et les systèmes d'intelligence artificielle.

3. Comprendre et se pencher sur les **implications éthiques, légales et sociétales**, dans le but de concevoir des systèmes d'intelligence artificielle conformes aux principes éthiques, légaux et sociétaux des États-Unis. Le rapport insiste sur l'importance d'assurer la justice, la transparence et la responsabilité des systèmes, dès la phase de conception.

4. Assurer la **sécurité et sûreté des systèmes d'intelligence artificielle**, en particulier dans l'adaptation à des environnements complexes et/ou incertains. Avant que les systèmes d'intelligence artificielle ne soient utilisés à grande échelle, il faut s'assurer que ces systèmes vont opérer de manière sécurisée et fiable. Des progrès plus poussés dans la recherche sont nécessaires pour répondre à ce défi.

5. Développer des **bases de données publiques partagées** pour l'analyse, l'apprentissage, l'entraînement et le test de systèmes d'intelligence artificielle. La profondeur, la qualité et la précision des données d'apprentissage conditionnent leurs performances. La mise à disposition de bases de données fédérales existantes est proposée.

6. Développer des **standards** visant à s'assurer que les technologies émergentes répondent à des objectifs précis en termes de fonctionnalité et d'interopérabilité, ainsi qu'en termes de sécurité et de performance. Des



---

recherches additionnelles sont nécessaires pour développer un large spectre de techniques évaluatives et de plateformes d'essai.

7. Évaluer les **besoins en termes de main-d'œuvre**. Les avancées en intelligence artificielle demandent une forte communauté de chercheurs et d'experts en intelligence artificielle, ce qui plaide pour une meilleure compréhension des besoins actuels et futurs de main-d'œuvre.

Le Plan Stratégique de R&D pour l'intelligence artificielle se conclut avec deux recommandations :

- Développer un cadre de mise en œuvre de la R&D de l'intelligence artificielle afin d'identifier les opportunités scientifiques et technologiques et soutenir une coordination effective des investissements en R&D de l'intelligence artificielle, en accord avec les stratégies 1 à 6 de ce plan.
- Étudier le paysage professionnel national afin de créer et maintenir une main-d'œuvre de bon niveau pour la R&D en intelligence artificielle, en accord avec la stratégie 7 de ce plan.

### 3. Le rapport de la Chambre des Communes du Royaume-Uni

La commission science et technologie de la Chambre des Communes du Royaume-Uni a rendu public, en octobre 2016, un **rapport sur la robotique et l'intelligence artificielle**, intitulé « *Robotics and artificial intelligence* »<sup>1</sup>. Les rapporteurs de ce texte affirment le potentiel disruptif du développement de l'intelligence artificielle sur les manières de vivre et de travailler.

Selon Tania Mathias, présidente par intérim de la commission, « *l'intelligence artificielle a encore du chemin à parcourir avant que les systèmes artificiels et robots ne deviennent comme ceux imaginés dans des œuvres telles que Star Wars. Pour le moment, les "machines à intelligence artificielle" se voient assigner des rôles restreints et spécifiques, comme la reconnaissance vocale ou être un adversaire au jeu de Go. Cependant, la science-fiction est doucement en train de devenir une science factuelle, la robotique et les intelligences artificielles semblent être destinées à jouer un rôle de plus en plus important dans nos vies au cours des prochaines décennies. Il est trop tôt pour établir des réglementations sectorielles pour ce domaine naissant mais il est vital qu'un examen attentif des implications éthiques, légales et sociétales des systèmes d'intelligence artificielle débute dès maintenant* ».

Ce rapport estime que le **rôle d'impulsion joué par le Gouvernement** dans le domaine de la robotique et de l'intelligence artificielle a été déficient. Le fait que les structures privées s'interrogent face aux risques et bénéfices de l'intelligence artificielle ne décharge pas le Gouvernement britannique de ses responsabilités. Ce rapport incite ainsi le Gouvernement à établir une « **Commission nationale sur l'intelligence artificielle** » au sein de l'Institut

---

<sup>1</sup> Pour consulter le rapport :

<https://www.publications.parliament.uk/pa/cm201617/cmselect/cmsctech/145/145.pdf>

---

Alan Turing, afin d'identifier les principes de gouvernance du développement et de l'utilisation de l'intelligence artificielle et d'encourager le débat public.

Ce rapport aborde également la **question de l'emploi**. Selon les rapporteurs de ce texte, les progrès en productivité, permis par la robotique et l'intelligence artificielle, sont largement prévisibles. Bien qu'une incertitude plane sur l'ampleur des destructions d'emploi et la capacité du marché du travail à résorber un chômage potentiellement massif, il demeure que, selon les recommandations de ce rapport, l'accent doit être mis sur **l'ajustement des systèmes d'éducation et de formation** britanniques afin de délivrer les compétences qui permettront à la population de s'adapter et de profiter des opportunités ouvertes au moment où les nouvelles technologies sont et seront mises en oeuvre.

Selon Tania Mathias, « *il est concevable que nous verrons des machines à intelligence artificielle créer de nouveaux emplois dans les décennies à venir tout en remplaçant d'autres. Comme nous ne pouvons pas encore prévoir avec exactitude comment ces changements vont se concrétiser, nous devons répondre avec préparation et réactivité aux besoins de requalification et d'amélioration des compétences* ». Cela implique un **engagement de la part du Gouvernement** qui garantisse la flexibilité des systèmes d'éducation et de formation leur permettant de s'adapter à l'évolution des opportunités et des exigences imposées aux travailleurs. Le rapport déplore, en outre, que le Gouvernement n'ait pas encore publié sa stratégie numérique, ni déterminé ses plans pour doter les futurs travailleurs des compétences numériques essentielles à leur développement professionnel.

Le Gouvernement britannique a apporté en janvier 2017 une réponse<sup>1</sup> au rapport présenté par la Chambre des Communes en reprenant explicitement **trois recommandations** avancées dans celui-ci sans toutefois annoncer de décisions nouvelles. Tout d'abord, le Gouvernement semble bien conscient du bien fondé de développer une stratégie pour soutenir les systèmes autonomes et robotiques (*Robotics and Autonomous Systems*, RAS en anglais) et indique sa volonté d'améliorer **la coordination stratégique entre le Gouvernement, les industries et le monde académique** afin de maximiser les retombées économiques et sociales permises par ces technologies.

Ensuite, le Gouvernement reconnaît et promeut le rôle de l'Institut Alan Turing dans la stratégie à développer, en annonçant la mise en place en son sein d'une « Commission nationale sur l'intelligence artificielle ». Les missions primordiales de l'Institut relatives au développement de l'intelligence artificielle seront de répondre aux enjeux éthiques, sociaux et juridiques, d'assurer que le développement de ces technologies soit

---

<sup>1</sup> La réponse du Gouvernement britannique est disponible ici : <https://www.publications.parliament.uk/pa/cm201617/cmselect/cmsctech/896/89602.htm>

---

pleinement bénéfique à la société, et de bâtir la confiance dans les capacités de développement du Royaume-Uni dans ce secteur.

Enfin, le Gouvernement affirme avoir bien identifié la nécessité de la **formation aux technologies numériques**, cette dimension étant appelée à être incluse dans l'ensemble des filières de formation.

Un groupe de travail de la *Royal Society*, rencontré par vos rapporteurs, conduit également un projet de recherche<sup>1</sup> visant à analyser les opportunités et les défis juridiques, sociaux et éthiques liés au *machine learning* et à ses applications sur les cinq à dix prochaines années. Ce groupe de travail publiera des **recommandations à l'attention du Gouvernement britannique au cours de l'année 2017**.

#### **4. Les initiatives chinoises et japonaises en intelligence artificielle accordent une place contrastée aux questions éthiques**

La Chine et le Japon ont déployé des stratégies nationales pour l'IA qui **accordent une place contrastée aux questions éthiques** : alors que la première les délaisse, le second a mis en place une réflexion officielle de haut niveau en la matière.

La Chine se trouve **au cœur de l'évolution récente des techniques** et déploie d'**importants moyens** avec pour objectif de devenir leader mondial dans le domaine<sup>2</sup>. Elle a l'ambition d'être la **première à disposer, vers 2025-2030, d'une intelligence artificielle générale**, comparable à celle du cerveau humain (« *Artificial General Intelligence* » ou *AGI*), première étape dans le but de parvenir ensuite au développement d'intelligences suprahumaines, ce qui ne soulève pas de questions éthiques majeures dans ce pays.

Même si les progrès visibles de la Chine reposent encore sur des architectures conçues par des scientifiques occidentaux, **ses atouts propres sont réels**, comme l'a indiqué à vos rapporteurs le service scientifique de l'Ambassade de France à Pékin : elle dispose « *des deux supercalculateurs les plus puissants du monde, d'un marché intérieur très important et friand des avancées potentielles du secteur, d'une collusion féconde entre l'État, les instituts de recherche, les universités, les géants de l'internet et de l'informatique, les start-ups* ».

Le 13<sup>ème</sup> plan quinquennal comprend une liste de 15 « nouveaux grands projets – innovation 2030 » qui structurent les priorités scientifiques du pays et correspondent chacun à des investissements de plusieurs milliards d'euros. Parmi ces 15 projets, on en trouve quatre qui sont dédiés indirectement à l'IA, pour un montant de 100 milliards de yuans en trois ans : un projet de « Recherche sur le cerveau » et des projets d'ingénierie

---

<sup>1</sup> Pour consulter le projet de la Royal Society : <https://royalsociety.org/topics-policy/projects/machine-learning/>

<sup>2</sup> Une note complète sur le sujet de la place de l'IA en Chine est annexée au présent rapport.

---

intitulés « Mega données », « Réseaux intelligents » et « Fabrication intelligente et robotique ».

Les recherches envisagées ne sont pas seulement théoriques, les applications multiples sont un moteur important : contrôle de drones, de robots ou d'avatars, interfaces en langage naturel, interfaces homme-machine, détection des émotions, analyse d'images, automobiles autonomes, etc. Les systèmes permettront aussi de faire l'analyse des *big data*.

L'exemple suivant ne semble pas gêner les chercheurs chinois, comme l'a précisé à vos rapporteurs le service scientifique de l'Ambassade de France : « *Piloté par le Gouvernement et l'organisme central de planification, le dispositif de notation de la population devrait récupérer automatiquement les informations sur les citoyens d'ici 2020. Il scrutera les activités en ligne etc, pour générer un score individuel. Il semble que si un seuil est dépassé, l'individu concerné se verra privé d'un certain nombre de droits et de services* ». Une expérimentation a commencé dès 2017 à Suining<sup>1</sup>. D'autres dispositifs d'évaluation individuelle des citoyens ou des consommateurs sont étudiés, comme le *Sesame Credit* du distributeur en ligne Alibaba<sup>2</sup>.

**L'opinion publique chinoise semble donc, à ce stade, peu préoccupée par les questions éthiques et philosophiques soulevées par ces applications**, ni même par les questions concernant la protection des données et de la vie privée qui sont des enjeux importants pour l'évolution de la recherche et pour les applications de l'IA. En effet, les jeux de données sont l'une des dimensions majeures des progrès des algorithmes de *deep learning*.

L'intelligence artificielle est considérée, par ailleurs, comme **l'élément clé de la révolution numérique au Japon** selon les informations transmises à vos rapporteurs par le service scientifique de l'Ambassade de France au Japon<sup>3</sup>. Le gouvernement japonais a fait le choix d'annoncer dès mai 2015 un **grand plan quinquennal pour la robotique et l'intelligence artificielle**<sup>4</sup>, entré en vigueur le 1<sup>er</sup> avril 2016.

Il a fait de ces deux dernières technologies le socle de sa nouvelle stratégie en science, technologie et innovation, qui vise à mettre en place, au terme de la « 4<sup>ème</sup> révolution industrielle », une « société 5.0 », société super-intelligente et fer-de-lance à l'échelle mondiale du dynamisme japonais. On estime que le marché de l'intelligence artificielle au Japon devrait passer **de 3,7 milliards de yens en 2015 à 87 milliards en 2030**, dont

---

<sup>1</sup> Cf. [https://www.washingtonpost.com/world/asia\\_pacific/chinas-plan-to-organize-its-whole-society-around-big-data-a-rating-for-everyone/2016/10/20/1cd0dd9c-9516-11e6-ae9d-0030ac1899cd\\_story.html?utm\\_term=.71ca38d649e1](https://www.washingtonpost.com/world/asia_pacific/chinas-plan-to-organize-its-whole-society-around-big-data-a-rating-for-everyone/2016/10/20/1cd0dd9c-9516-11e6-ae9d-0030ac1899cd_story.html?utm_term=.71ca38d649e1)

<sup>2</sup> Cf. <http://www.bbc.com/news/world-asia-china-34592186>

<sup>3</sup> Une note complète sur le sujet de la place de l'IA au Japon est annexée au présent rapport.

<sup>4</sup> Soutenu par 200 entreprises et universités, le 5<sup>ème</sup> Plan-cadre pour la Science et la technologie prévoit le triplement des investissements dans ce domaine d'ici 2020 en vue d'inciter les entreprises à répandre l'utilisation de la robotique dans tous les secteurs de l'économie et de la société afin de surmonter le vieillissement et la baisse du nombre d'actifs.

---

30,5 milliards de yens dans le domaine du transport et 42 milliards en incluant la production industrielle pour le transport.

Le gouvernement japonais a annoncé une **vague d'investissements massifs** dans le domaine, avec **27 milliards de yens pour la seule année 2016**, à travers la création de centres de recherche et technologies dédiés. Ce montant est à rapprocher des plus de 300 milliards de yens d'investissement que les grands groupes japonais ont lancé sur les trois ans à venir pour financer des programmes ou des laboratoires dédiés à l'intelligence artificielle<sup>1</sup>. Les partenariats publics-privés entre les centres publics et les grands groupes japonais se multiplient, afin d'exploiter le potentiel de création de valeur que constitue l'intelligence artificielle sur des applications ciblées. L'intelligence artificielle ouvre la voie à des développements très attendus au Japon, selon le service scientifique de l'Ambassade, dans le domaine des **transports, de la communication, de la traduction automatique ou de la robotique**, notamment à l'horizon des Jeux Olympiques et paralympiques de 2020, que le Japon envisage comme une vitrine technologique pour se présenter comme « *le pays leader mondial* » en termes d'innovation.

Il peut être relevé qu'à l'occasion de l'organisation du sommet 2016 du G7 au Japon, qui s'est tenu à Shima les 26 et 27 mai 2016, le Japon a pris l'initiative d'organiser une **réunion ministérielle du G7 dédiée aux Sciences et technologies de l'information et de la communication**, format qui n'avait pas été mis en œuvre depuis 20 ans.

Le ministère japonais de l'Éducation, de la culture, des sports, des sciences et de la technologie a rédigé en 2016 un **Livre blanc pour la science et la technologie**, intitulé « *Vers la société ultra-intelligente mise en œuvre par l'IoT, le big data et l'IA- pour que le Japon soit un précurseur mondial* », adopté en Conseil des ministres le 20 mai 2016. Ce rapport signale l'insuffisance au Japon de la formation et de la recherche fondamentale en informatique, qui devraient venir soutenir le développement des technologies de l'information et de la communication.

Trois ministères, respectivement en charge de la recherche, de l'industrie et des communications se sont mobilisés avec une vitesse étonnante pour développer des initiatives pour la robotique et l'intelligence artificielle et ont, chacun, annoncé en 2016 la création d'un centre de recherche spécifique sur l'intelligence artificielle : un premier centre auprès du RIKEN (principal centre de recherche japonais, du type CNRS), portant le nom de « *AIP Center* »<sup>2</sup> (*Advanced Integrated Intelligence Platform Project*

---

<sup>1</sup> Plusieurs exemples de projets sont détaillés dans la note sur le sujet de la place de l'IA au Japon annexée au présent rapport.

<sup>2</sup> Les activités s'articulent autour de 5 objectifs : développer des technologies pour l'intelligence artificielle « fondamentale » (basé essentiellement sur le machine learning) ; contribuer à l'accélération de la recherche scientifique (analyse et synthèse d'articles scientifiques, d'expériences de brevets...) ; contribuer à des applications concrètes à fort impact sociétal (problématique des soins dans le contexte du vieillissement de la population, gestion des infrastructures et des ouvrages de

---

Center), l'AIRC (*Artificial Intelligence Research Center*), hébergé par l'AIST (*Advanced Institute for Science and Technologies*), et, enfin, le NICT (*National Institute for Information and Communication Technologies*).

Le gouvernement japonais a mis en place **deux structures pour le suivi des stratégies du grand plan quinquennal pour la robotique et l'intelligence artificielle** : un Conseil dédié à la stratégie des technologies liées à l'IA (qui a notamment pour but de coordonner les actions des trois centres de recherche précités, disposant d'un site web commun et mutualisant leurs ressources informatiques) et un Comité de délibération sur l'IA et la société humaine. Ces deux structures relèvent directement du gouvernement japonais.

**Le Conseil de la stratégie des technologies liées à l'IA** a tenu sa réunion inaugurale le 18 avril 2016 et sert de « quartier général de la recherche et du développement (R&D) des technologies de l'IA et de leurs applications industrielles » en regroupant les trois ministères impliqués dans l'IA. Il est présidé par Yuichiro Anzai, Président de la JSPS (agence de financement de la recherche du MEXT dédiée à la recherche fondamentale), et composé de deux représentants du Keidanren (syndicat patronal des entreprises japonaises), des présidents de deux universités (Université de Tokyo et Université d'Osaka) et des présidents de cinq grands instituts de recherche et agences de financement : le NICT (*National Institute for Information and Communication Technologies*, dépendant du Ministère des affaires intérieures et communication, MIC), le RIKEN (principal centre de recherche japonais), l'AIST (*Advanced Institute for Science and Technologies*, centre de recherche du Ministère de l'Économie), la JST (agence de financement orientée vers les projets de recherche appliquée) et la NEDO (agence de financement dépendant du Ministère de l'Économie). Deux sous-comités ont été placés sous l'autorité de ce Conseil : le Comité de collaboration de recherche (conseil des présidents des instituts de recherche et des agences de financement) et le Comité de collaboration industrielle qui se réunissent chacun mensuellement.

**Le Comité de délibération sur l'IA et la société humaine** a tenu sa première réunion le 30 mai 2016 en présence de Aiko Shimajiri, ministre chargée de la politique de la science et de la technologie, conformément à une annonce en Conseil des ministres, le 12 avril 2016. Il s'agit de la **première structure gouvernementale ayant pour mission d'étudier les enjeux liés à l'IA**, selon cinq points de vue, à savoir : l'aspect éthique, l'aspect légal, l'aspect économique, l'aspect social et l'aspect R&D. Ce Comité, présidé par Mme Yuko Harayama et composé de onze experts, se réunit à une fréquence mensuelle pour analyser les activités nationales et internationales et a choisi de se baser sur des cas d'application précis, mettant en œuvre des technologies qui devraient voir le jour à court terme : le véhicule autonome, l'automatisation de l'appareil de production et la

---

*génie civil, résilience aux catastrophes naturelles...)* ; prise en compte des aspects éthiques, légaux et sociaux ; développement des ressources humaines.



---

communication homme/machine. Il souhaite également engager un débat avec le grand public, par le biais essentiellement de séminaires ouverts et de questionnaires en ligne.

Le Comité a remis ses conclusions, qui seront prises en compte dans la nouvelle Stratégie japonaise globale pour la Science, la technologie et l'innovation qui sera publiée en juin 2017. Des discussions au niveau international en la matière sont prévues à partir de 2017. Mme Harayama a toutefois présenté à Paris le 17 novembre 2016, dans le cadre du *Technology Foresight Forum 2016* de l'OCDE<sup>1</sup> dédié à l'intelligence artificielle, les premières pistes de réflexion du Comité :

- Éthique : Le citoyen peut-il accepter d'être manipulé pour modifier ses sentiments, convictions ou comportements, et d'être catégorisé ou évalué, sans en être informé ? Quel impact aura le développement de l'IA sur notre sens de l'éthique et les relations entre les hommes et les machines ? Dans la mesure où elle étend notre temps, notre espace et nos sens, est-ce que l'IA viendra affecter notre conception de l'humanité, notamment notre conception des facultés et des émotions humaines ? Comment évaluer les actions et la création à partir de l'IA ?

- Légal : comment trouver le juste équilibre entre les bénéfices du traitement des *big data* par l'IA, et la protection des informations personnelles ? Est-ce que les cadres légaux existants pourront s'appliquer aux nouvelles problématiques soulevées par l'IA ? Comment clarifier la responsabilité dans le cas d'incidents impliquant de l'intelligence artificielle (par exemple dans le cas du véhicule autonome) ? Quels sont les risques encourus en utilisant l'IA ou en ne l'utilisant pas ?

- Économique : comment l'IA va-t-elle changer notre manière de travailler ? Quelle politique nationale mettre en place pour favoriser l'utilisation de l'IA ? Comment l'IA va-t-elle modifier le monde de l'emploi ?

- Sociétal : Comment réduire les divisions liées à l'IA et répartir de manière équitable le coût social de l'IA ? Y-a-t-il des pathologies ou des conflits de société potentiellement engendrés par l'IA ? Peut-on garantir la liberté d'utiliser ou non l'IA, et assurer le droit à l'oubli ?

- Éducation : Quelle politique nationale mettre en place pour faire face aux inégalités que pourrait provoquer l'utilisation de l'IA dans le domaine de l'éducation ? Comment développer notre capacité à exploiter l'IA ?

- R&D : Quelle R&D développer pour l'IA en respect de l'éthique, la sécurité, la protection de la vie privée, la transparence, la contrôlabilité, la visibilité, la responsabilité ? Comment rendre disponible l'information liée à l'IA de manière à ce qu'un utilisateur puisse prendre la décision d'utiliser ou non l'IA ?

---

<sup>1</sup> <http://www.oecd.org/sti/ieconomy/Yuko%20Harayama%20-%20AI%20Foresight%20Forum%20Nov%202016.pdf>

---

Le Comité cherche notamment à approfondir trois voies pour définir des politiques adaptées : la coévolution de la société et de la technologie ; la recherche d'un équilibre entre les bénéfices (services personnalisés à coût abordable) et les risques liés à l'IA (discrimination, perte de protection des données à caractère privé, perte d'anonymat) ; la définition des limites de la prise de décision automatisée.

Les questions éthiques occupent donc, au total, une **place contrastée au sein des initiatives chinoises et japonaises en IA**. Autant elles **ne semblent pas être, pour le moins et à ce stade, au cœur des démarches impulsées par les autorités chinoises**, autant **le gouvernement japonais les évoque très largement à travers la mise en place de son Comité national de délibération sur l'IA et la société humaine**. Le Japon a fait le choix **d'accompagner la transition produite par l'intelligence artificielle de manière très étroite**, avec des investissements publics massifs, mais également par une **réflexion publique institutionnalisée sur l'impact de l'utilisation de l'intelligence artificielle pour la société**.

### 5. La stratégie du Gouvernement français pour l'intelligence artificielle : un plan qui arrive trop tard pour être intégré dans les stratégies nationales destinées au monde de la recherche

Thierry Mandon, secrétaire d'État chargé de l'Enseignement supérieur et de la Recherche et Axelle Lemaire, secrétaire d'État au Numérique et à l'Innovation, ont lancé une **stratégie nationale en intelligence artificielle**, appelée « France IA », le 20 janvier 2017 dans les locaux de l'incubateur Agoranov à Paris. Cette stratégie, qui consistait à mobiliser **sept groupes de travail** rassemblant des chercheurs et des entreprises afin de définir les **orientations stratégiques** de la France dans ce domaine, a été suivie le 21 mars 2017 de l'annonce du **plan du Gouvernement** à la Cité des sciences et de l'industrie et de la remise d'un rapport au Président de la République<sup>1</sup>. Ce plan prévoit :

- la mise en place d'un comité stratégique « FranceIA » rassemblant les sphères académique, scientifique, économique ainsi que des représentants de la société civile, chargé de mettre en œuvre les recommandations des groupes de travail ;

- la coordination par la France d'une candidature à un « projet phare de technologie émergente » (« *FET flagship* ») sur l'IA, co-financé par l'Union européenne (montant estimé d'un milliard d'euros) ;

- le lancement d'un nouveau programme mobilisant les institutions de recherche pour identifier, attirer et retenir les meilleurs talents en IA,

---

<sup>1</sup> Cf. [http://www.economie.gouv.fr/files/files/PDF/2017/Rapport\\_synthese\\_France\\_IA\\_.pdf](http://www.economie.gouv.fr/files/files/PDF/2017/Rapport_synthese_France_IA_.pdf)  
[http://www.economie.gouv.fr/files/files/PDF/2017/Conclusions\\_Groupes\\_Travail\\_France\\_IA.pdf](http://www.economie.gouv.fr/files/files/PDF/2017/Conclusions_Groupes_Travail_France_IA.pdf) et  
[http://www.economie.gouv.fr/files/files/PDF/2017/Dossier\\_presse\\_France\\_IA.pdf](http://www.economie.gouv.fr/files/files/PDF/2017/Dossier_presse_France_IA.pdf)



---

dans le cadre de l'action Programmes prioritaires de recherche du troisième volet du plan pour les investissements d'avenir (« PIA 3 ») ;

- le financement d'une infrastructure mutualisée pour la recherche ;
- la constitution d'un consortium public-privé en vue de l'identification ou de la création d'un centre interdisciplinaire pour l'intelligence artificielle ;
- l'inclusion systématique, d'ici fin 2017, de l'IA dans les priorités de l'ensemble des dispositifs publics de soutien à l'innovation ;
- la mobilisation des ressources publiques (Bpifrance, PIA) et privées pour atteindre l'objectif d'ici cinq ans d'investir dans dix *start-ups* françaises pour plus de 25 millions d'euros chacune ;
- la mobilisation des filières automobile, relation client, finances, santé et transport ferroviaire pour que chaque filière définisse une stratégie sectorielle en IA d'ici fin 2017 ;
- le lancement d'un appel à projets pour des plateformes sectorielles de partage de données pour trois à six secteurs, d'ici la fin 2017 ;
- la conclusion du débat sur l'éthique des « algorithmes » animé par la CNIL en octobre 2017 et la remise de recommandations au Gouvernement ;
- le lancement d'ici l'été 2017 d'une concertation de France Stratégie sur la question des effets de l'intelligence artificielle sur l'emploi.

Vos rapporteurs jugent **regrettable** que l'évolution récente des connaissances et des technologies ainsi que la pertinence de l'actualité n'ait pas permis que cette démarche du Gouvernement soit inscrite dans les **différentes stratégies nationales de recherche à 10 ans** et le **livre blanc de l'enseignement supérieur et de la recherche**<sup>1</sup>.

## **B. DES « LOIS D'ASIMOV » À LA QUESTION CONTEMPORAINE DE LA RÉGULATION DES SYSTÈMES D'INTELLIGENCE ARTIFICIELLE**

### **1. Dépasser les « lois d'Asimov » pour envisager un droit de la robotique**

Dès ses premiers romans, l'écrivain **Isaac Asimov** a formalisé ses « **trois lois** » applicables aux robots. Ces « trois lois » sont les suivantes :

- première loi, « *un robot ne peut porter atteinte à un être humain ni, restant passif, laisser cet être humain exposé au danger* » ;

---

<sup>1</sup> La loi pour l'enseignement supérieur et la recherche du 22 juillet 2013 a conduit à la préparation de deux grandes stratégies à 10 ans : une Stratégie nationale de l'enseignement supérieur (StraNES) et une Stratégie nationale de la recherche (S.N.R.). Cette loi prévoit également, dans son article 17, la réalisation d'un livre blanc de l'enseignement supérieur et de la recherche par le Gouvernement au Parlement tous les cinq ans.

---

- deuxième loi, « *un robot doit obéir aux ordres donnés par les êtres humains, sauf si de tels ordres sont en contradiction avec la première loi* » ;

- troisième loi, « *un robot doit protéger son existence dans la mesure où cette protection n'entre pas en contradiction avec la première ou la deuxième loi* ».

Au-delà de l'articulation des « trois lois » de la robotique entre elles, Isaac Asimov a imaginé une **quatrième loi, dite « loi zéro », élaborée par les robots eux-mêmes**. Cette invention suit le changement d'échelle de la sphère d'influence des robots. Elle consiste en une généralisation de la première loi, par le passage d'un individu à l'humanité toute entière : « *nulle machine ne peut porter atteinte à l'humanité ni, restant passive, laisser l'humanité exposée au danger* ».

**Bien qu'elles puissent avoir l'air infaillibles, ces règles peuvent être prises en défaut et atteindre leurs limites.** L'œuvre d'Isaac Asimov montre que l'application et l'articulation entre ces trois lois ne vont pas de soi. Ces règles, **interprétées par les robots**, peuvent même **finir par nuire aux êtres humains**.

**La Corée du Sud s'est tout de même inspirée de ces « lois » pour rédiger un projet de charte sur l'éthique des robots**, dans le but « *d'éviter les problèmes de société qui pourraient découler de mesures sociales et juridiques inadéquates prises pour encadrer l'existence de robots dans la société* ».

Vos rapporteurs rappellent que **les « lois d'Asimov » sont des règles issues de la fiction et qu'elles posent de profonds problèmes de mise en œuvre**. Les experts en robotique rencontrés par vos rapporteurs ont tous souligné la difficulté à traduire ces « lois » dans des systèmes matériels. Au total **il s'agit davantage de principes éthiques généraux que de prémisses à un droit de la robotique**.

Les travaux d'**Andrea Keay** visant à compléter les « lois d'Asimov » participent aussi de cette logique éthique. Pour mémoire, elle propose que les robots ne soient pas utilisés comme des armes, qu'ils doivent se conformer aux lois, notamment celles sur la protection de la vie privée, être sûrs, fiables et donner une image exacte de leurs capacités, qu'ils ne doivent pas être utilisés pour tromper les utilisateurs les plus vulnérables (ce qui plaide pour éviter les robots humanoïdes trop ressemblant) et qu'il doit être possible de connaître le responsable de chaque robot.

Reconnaître une **personnalité juridique des robots** est une des pistes innovantes qui parcourent le débat public sur la robotique.

En France, **l'avocat Alain Bensoussan**, rencontré à plusieurs reprises par vos rapporteurs, **milite en faveur de l'adoption d'un droit des robots** au sein de l'association pour le droit des robots qu'il préside. Il a ainsi rédigé **un projet de charte des droits des robots**, qui fait de ces derniers des êtres artificiels dotés d'une personnalité juridique particulière et d'un droit à la

dignité<sup>1</sup>. Il réfléchit également aux implications en matière de responsabilité et d'assurance. En outre, il demande à ce que tout robot dispose de systèmes de sécurité permettant un arrêt d'urgence.

Selon Alain Bensoussan, les textes actuellement en vigueur, à l'instar de la loi du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés<sup>2</sup>, **n'offrent pas un cadre juridique suffisant** face aux évolutions en cours de la robotique, à l'amélioration des capacités d'apprentissage et de la liberté décisionnelle du robot et à la question de la confidentialité des enregistrements et des données que celui-ci peut recueillir. En particulier, pour Alain Bensoussan, il serait essentiel d'intégrer aux corpus normatifs existants un « droit des robots », qui se déclinerait sur trois axes : les règles générales applicables à tous les types de robots ; les règles spécifiques par type de robot (véhicule autonome, robot humanoïde...) ; et les référentiels robotiques sur les plans éthiques, culturels et normatifs.

Vos rapporteurs ont sur le sujet de la reconnaissance de la personnalité juridique des robots un avis très réservé. Ils ne sont pas convaincus de l'intérêt de reconnaître une **personnalité juridique aux robots** et se demandent **à qui il conviendrait d'accorder la personnalité juridique**, au robot dans son ensemble ou à son intelligence artificielle ?

Dans la mesure où **le système d'intelligence artificielle pourrait migrer d'un corps robotique à un autre**, la partie physique du robot ne serait qu'un contenant, destiné à recevoir pour un temps donné un système. Il serait alors **opportun d'opérer une discrimination entre la partie physique et la partie informatique du robot** en vue de les soumettre à des régimes juridiques différents, notamment en matière de responsabilité. **Il faudrait alors pouvoir, tel l'historien Ernst Kantorowicz distinguant les deux corps du roi, discerner les deux corps du robot**<sup>3</sup>.

Vos rapporteurs estiment qu'il est **urgent d'attendre en la matière** et que le sujet de la personnalité juridique des robots n'est pas à une question qui mérite d'être posée à ce stade.

## 2. Les questions juridiques en matière de conception (design), de propriété intellectuelle et de protection des données personnelles et de la vie privée

S'agissant des autres aspects juridiques de l'intelligence artificielle et de la robotique, il sera loisible de conduire une réflexion et de faire de la **prospective concernant la conception (design) et l'autorisation de**

---

<sup>1</sup> Cette charte précise que « les données à caractère personnel conservées par un robot sont soumises à la réglementation Informatique et libertés. Un robot a le droit au respect de sa dignité limitée aux données à caractère personnel conservées ».

<sup>2</sup> L'intégralité de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés est disponible ici : <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000886460>

<sup>3</sup> Ernst Kantorowicz, « Les Deux Corps du roi. Essai sur la théologie politique au Moyen Âge », Gallimard, 1957.

---

**commercialisation**<sup>1</sup>. Pour Rodolphe Gélén et Olivier Guilhem, respectivement directeur scientifique et directeur juridique d'Aldebaran puis de Softbank robotics, il n'existe **pas de vide juridique béant**. Les rapports parus sur le sujet, notamment dans le monde anglo-saxon, semblent aller dans le même sens.

Dans l'état actuel du droit, en cas de **commercialisation de robots entre professionnels**, ces derniers disposent d'une certaine **liberté contractuelle qui leur permet de combler les incertitudes législatives et jurisprudentielles**. Ainsi, leur appréciation totale des possibilités et leur maîtrise des contraintes et limites technologiques leur offrent une approche technique permettant la distribution de la responsabilité finale de chaque partie prenante de cet échange commercial (fabricant, développeur, propriétaire et utilisateur).

Concernant la commercialisation de robots à destination des consommateurs, le **droit de la consommation** a vocation à s'appliquer.

La propriété intellectuelle pose des questions auxquelles vos rapporteurs n'ont pas de réponses définitives. Quel est le statut de ce qui est créé par des technologies d'intelligence artificielle ? **Ces œuvres sont-elles la propriété de son acheteur, de son fabricant, de l'éditeur du logiciel ?** En tout cas **elles n'appartiennent pas à la machine** d'après vos rapporteurs. Plusieurs juristes, tel l'avocat Alain Bensoussan imagine à l'inverse une telle solution, qui implique de doter les intelligences artificielles et les robots d'une personnalité propre, comme il a été vu cette idée n'emporte pas leur adhésion.

Dans l'état actuel du droit français, la reconnaissance d'une création pleinement robotique ou issue de technologies d'intelligence artificielle n'existe pas et **seul un être humain peut bénéficier d'un régime de propriété intellectuelle pour ses créations**. Il existe néanmoins certains mécanismes juridiques permettant à une œuvre ou à un ouvrage dont le processus de création fut partiellement assuré par une machine ou un système d'intelligence artificielle de bénéficier d'un régime de protection au titre de la propriété intellectuelle. Ainsi, **il est possible d'accorder un brevet à une création résultant d'un processus industriel impliquant un ordinateur ou un robot**. De même, si les créations produites par des composants robotiques ne sont pas éligibles à la protection assurée par le code de la propriété intellectuelle, le savoir-faire, qui représente un ensemble d'informations non brevetées résultant de l'expérience et de l'expérimentation, peut être utilisé comme un outil pour protéger la création robotique, suite à la construction jurisprudentielle et doctrinale reconnaissant la notion de savoir-faire. Enfin, conformément à la

---

<sup>1</sup> Un régime du type de celui qui est applicable aux médicaments avant autorisation de mise sur le marché, avec une période de tests et d'observations, pourrait devenir obligatoire pour les systèmes autonomes d'intelligences artificielles, au stade où leur commercialisation massive sera envisagée.

---

« Classification de Nice »<sup>1</sup>, qui est le système de classification des produits et des services établi dans le cadre de l'Organisation mondiale de la propriété intellectuelle et ratifié par la France, les robots et technologies d'intelligence artificielle sont considérés comme des biens et leurs actions en tant que fournisseur de services ne sont, de fait, pas prises en compte.

À titre de comparaison, les robots et les technologies d'intelligence artificielle n'étant pas dotés de personnalité juridique propre au regard du droit international, ils demeurent considérés aux yeux de nombreux systèmes juridiques nationaux comme des objets, et ne sont donc pas porteurs de droits. Le droit de l'Union européenne ne prévoit pas la création par un robot ou un ordinateur ; de fait, **leurs créations sont exclues du champ de la protection de la propriété intellectuelle, du brevet, du dépôt de marque et du droit d'auteur** tel que défini par le droit communautaire. Ainsi, au-delà de la France et plus généralement des États membres de l'Union européenne, de nombreux pays tels que l'Afrique du Sud, le Brésil, la Chine, les États-Unis ou le Japon, considèrent que **la création ne peut être qu'humaine, et non issue d'une machine ou d'une technologie d'intelligence artificielle**. Dans ce cas, les créations effectuées par le recours à un robot sont éligibles à la protection de la propriété intellectuelle, la propriété de la création étant attribuée au propriétaire ou à l'utilisateur de la machine ou du système.

L'avocat Laurent Szuskin propose, quant à lui, une **application distributive de la propriété intellectuelle**. Par exemple, une invention dans un système d'intelligence artificielle pourrait être protégeable par le brevet, les logiciels sous-jacents par le droit d'auteur, les bases de données par le droit spécifique à celles-ci, etc. Sur la question de savoir si le résultat produit par l'intelligence artificielle appartient au développeur de la solution ou au fournisseur ou encore à l'entreprise ayant intégré la solution à ses systèmes de production ou encore à une autre personne telle que celle ayant fourni les données Laurent Szuskin plaide, en l'absence de régime légal ou de jurisprudence à ce jour, pour une **solution contractuelle**. Les clauses « propriété intellectuelle et savoir-faire » encadrant le développement d'intelligence artificielle ou la fourniture de services d'intelligence artificielle doivent attribuer la propriété ou du moins l'affectation contractuelle des résultats qui découlent de l'usage de la solution.

**Le développement de la robotique « intelligente » et des technologies d'intelligence artificielle soulève également des questions en matière de protection des données personnelles.** Au quotidien, nos ordinateurs connectés à Internet et nos *smartphones* avec des logiciels tels que « Siri », « Cortana » ou « Google Now », nous font d'ores et déjà cohabiter avec des algorithmes puissants, qui connaissent beaucoup de chacun de nous, **sans que nous ne connaissions l'usage qui peut être fait de ces millions d'informations à caractère personnel**. Les agents conversationnels

---

<sup>1</sup> Les 45 classes établies par la « Classification de Nice » sont disponibles ici : <http://web2.wipo.int/classifications/nice/nicepub/en/fr/edition-20170101/taxonomy/>

---

étant appelé à jouer un rôle croissant dans nos sociétés, ce sujet doit faire l'objet d'une prise en charge satisfaisante.

**Les activités robotiques sont soumises au respect de la loi n°78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés.** Ainsi, les propriétaires de robots gérant le système de traitement des données doivent respecter les obligations posées par la Commission nationale de l'informatique et des libertés (CNIL), c'est-à-dire la notification standard sur les utilisations du robot, sur le type de logiciel utilisé, sur les systèmes de sécurité installés pour protéger les données d'intrusions tierces non-autorisées, sur les données personnelles stockées dans le robot, et sur les informations fournies aux utilisateurs concernant le traitement de leurs données personnelles. Tout traitement de données personnelles doit être signalé en amont à la CNIL, et l'utilisation de données « sensibles », comme les données médicales, doit être autorisée par la CNIL. En outre, les propriétaires de systèmes de traitement de données personnelles sont soumis à une obligation de sécurité et de confidentialité des données.

Un effort d'adaptation du cadre juridique de la protection des données à caractère personnel aux nouvelles réalités du monde numérique a été mené récemment à la fois au niveau communautaire et au niveau national.

**Le règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données<sup>1</sup>**, également appelé « règlement général européen sur la protection des données » et abrogeant la directive 95/46/CE, vise à doter les États membres d'une législation uniforme et actualisée en matière de protection des données à caractère personnel. Ce « règlement général sur la protection des données », entré en vigueur le 24 mai 2016 et qui sera applicable à partir du 25 mai 2018, est destiné à remplacer l'actuelle loi du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés.

L'application de ce règlement général européen sur la protection des données poursuit **trois objectifs**. Tout d'abord, l'application de ce règlement vise à « *renforcer les droits des personnes, notamment par la création d'un droit à la portabilité des données et de dispositions propres aux personnes mineures* »<sup>2</sup>. Ensuite, ce règlement impose un accroissement de la transparence et la responsabilisation des acteurs traitant des données, selon une logique de conformité dont les acteurs sont responsables, sous le contrôle et avec l'accompagnement du régulateur. Enfin, l'application de ce règlement a pour objectif de « *crédibiliser la régulation grâce à une coopération renforcée entre les autorités de protection de données, qui pourront notamment adopter des décisions*

---

<sup>1</sup> Le texte intégral du règlement est disponible ici : <http://eur-lex.europa.eu/legal-content/FR/ALL/?uri=CELEX%3A32016R0679>

<sup>2</sup> La CNIL propose une analyse détaillée et synthétique des dispositions contenues dans le règlement général européen sur la protection des données, disponible ici : <https://www.cnil.fr/fr/reglement-europeen-sur-la-protection-des-donnees-ce-qui-change-pour-les-professionnels>

---

*communes et des sanctions renforcées dans les cas de traitements de données transnationaux* »<sup>1</sup>.

Les **dispositions contenues** dans le règlement général européen sur la protection des données **encadrent la collecte et le traitement de données à caractère personnel par de nombreux droits et responsabilités**. Ainsi, ce règlement introduit la définition de « *l'expression du consentement renforcé* », indiquant que les utilisateurs doivent être informés de l'usage de leurs données et doivent donner leur accord, ou s'opposer, au traitement de leurs données personnelles. De même, le droit à la portabilité des données est affirmé, et les responsables de traitements des données à caractère personnel devront assurer des opérations respectant la protection des données personnelles, à la fois dès la conception du produit ou du service et par défaut.

Le règlement général européen sur la protection des données s'accompagne de l'adoption de la **directive (UE) 2016/680** du Parlement européen et du Conseil du 27 avril 2016 **relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel** par les autorités compétentes à des fins de prévention et de détection des infractions pénales, d'enquêtes et de poursuites en la matière ou d'exécution de sanctions pénales, et à la libre circulation de ces données. Cette directive, entrée en vigueur le 5 mai 2016 et que les États membres sont tenus de transposer dans leur ordre juridique interne au plus tard le 6 mai 2018, s'applique aux opérations de données effectuées à la fois au niveau transfrontalier et au niveau national par les autorités compétentes des États membres à des fins répressives, comprenant notamment la prévention et la détection des infractions pénales et la protection contre les menaces pour la sécurité publique.

Vos rapporteurs notent que certaines dispositions de la **loi n° 2016-1321 du 7 octobre 2016 pour une République numérique anticipent ce règlement européen sur la protection des données personnelles**.

La loi crée, en effet de nouveaux droits informatique et libertés et permet ainsi aux personnes de **mieux maîtriser leurs données personnelles, par l'affirmation du droit à l'autodétermination informationnelle**, qui constitue un renforcement des principes énoncés à l'article 1<sup>er</sup> de la loi du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés. De même, elle introduit le **droit à l'oubli** par les mineurs *via* une procédure spécifique accélérée permettant un effacement des données « problématiques » sur les plateformes, prévu désormais par l'article 40 de la loi du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés. L'article 40-1 de cette même loi permet désormais aux personnes d'organiser **la conservation, l'effacement et la communication de leurs données personnelles après leur décès**, notamment en désignant une personne pour exécuter des directives générales ou particulières souhaitées par le défunt. En outre, le nouvel article 43 bis de la loi du 6 janvier 1978 relative à l'informatique, aux fichiers et aux

---

<sup>1</sup> *Ibid.*



---

libertés offre la possibilité aux citoyens **d'exercer leurs droits par voie électronique.**

**La loi du 7 octobre 2016 pour une République numérique élargit également les pouvoirs de sanctions de la CNIL et lui confie de nouvelles missions.** Ainsi, le **plafond maximal des sanctions** de la CNIL est désormais de trois millions d'euros, cette augmentation anticipant celle prévue par le règlement général européen sur la protection des données. La loi introduit la **consultation systématique** de la CNIL afin que celle-ci soit saisie pour avis dès lors qu'un projet de loi ou une disposition d'un projet de loi ou de décret est relatif à la protection et au traitement des données à caractère personnel. Tous les **avis** de la CNIL seront par ailleurs **automatiquement rendus publics**. De plus, la loi renforce la CNIL de nouvelles missions en matière de protection des données personnelles. Ainsi, la CNIL doit assurer la **promotion des technologies de protection de la vie privée**, certifier la **conformité des processus d'anonymisation des données à caractère personnel** lors de leur mise en ligne et de leur utilisation, et **conduire une réflexion sur les problèmes éthiques et les questions de société face à l'évolution des technologies numériques**. Cette dernière mission confiée à la CNIL l'a ainsi menée à initier un cycle de débats publics, ateliers ou rencontres, intitulé « *Éthique et numérique : les algorithmes en débat* »<sup>1</sup>. Ce point sera abordé plus loin dans la partie consacrée au cadre national de la réflexion sur les enjeux éthiques de l'intelligence artificielle

Enfin, **la loi du 7 octobre 2016 pour une République numérique contribue également à une meilleure ouverture des données publiques.** L'article 4 de la loi modifie, en effet, le droit à la communication des documents administratifs, et crée l'article L311-3-1 du Code des relations des relations entre le public et l'administration, qui dispose que « *sous réserve de l'application du 2° de l'article L. 311-5, une décision individuelle prise sur le fondement d'un traitement algorithmique comporte une mention explicite en informant l'intéressé. Les règles définissant ce traitement ainsi que les principales caractéristiques de sa mise en œuvre sont communiquées par l'administration à l'intéressé s'il en fait la demande. Les conditions d'application du présent article sont fixées par décret en Conseil d'Etat* »<sup>2</sup>.

À ce titre, **la loi signe le passage d'une logique de la demande d'un accès à la logique de l'offre de données publiques**, bien que les critères de communicabilité de ces données demeurent inchangés.

---

<sup>1</sup>La CNIL rendra publique dès l'automne 2017 la synthèse des échanges et des contributions. Comme elle l'affirme, « il s'agira d'établir une cartographie de l'état du débat public et un panorama des défis et enjeux. Des pistes ou propositions pour accompagner le développement des algorithmes dans un cadre éthique pourraient faire par la suite l'objet d'arbitrages par les pouvoirs publics ». La présentation du projet est disponible ici : <https://www.cnil.fr/fr/ethique-et-numerique-les-algorithmes-en-debat-0>

<sup>2</sup> Le droit applicable à la communication des documents administratifs est disponible ici : [https://www.legifrance.gouv.fr/affichCode.do;jsessionid=9D32BEC5092D1A4CC3D228A1256F8ABA.tpdila23v\\_2?idSectionTA=LEGISCTA000031367696&cidTexte=LEGITEXT000031366350&dateTexte=20170312](https://www.legifrance.gouv.fr/affichCode.do;jsessionid=9D32BEC5092D1A4CC3D228A1256F8ABA.tpdila23v_2?idSectionTA=LEGISCTA000031367696&cidTexte=LEGITEXT000031366350&dateTexte=20170312)

---

Néanmoins, **vos rapporteurs soulignent le fait que l'introduction de ce nouveau droit réinterroge les questions d'explicabilité et de responsabilité des algorithmes.** En effet, le principe d'explicabilité d'un algorithme implique que toute décision prise par celui-ci doit être accessible et compréhensible par les personnes concernées par cette décision, afin de permettre à ces derniers de fournir une meilleure contestation des erreurs constatées ou des données erronées. De même, le principe de responsabilité d'un algorithme implique que l'algorithme ou son utilisateur rende compte des effets de leurs procédés et de leurs actions. Les algorithmes étant souvent caractérisés par leur opacité et qualifiés de boîtes noires insondables, **la collecte et le traitement de données publiques désormais librement accessibles par des algorithmes soulève de nombreuses préoccupations sur la transparence des algorithmes.**

Vos rapporteurs notent que **la loi** ne fournit toujours **pas de régime juridique spécifique** de protection des données personnelles dans les cas de collecte et de traitement de ces données par des robots intelligents ou des technologies d'intelligence artificielle, c'est le **droit commun de la protection des données à caractère personnel dans les traitements informatiques** qui continue de s'appliquer.

Afin de répondre plus spécifiquement aux préoccupations d'explicabilité et de responsabilité des algorithmes, l'Institut national de recherche en informatique et en automatique (INRIA), suite à un rapport remis au Gouvernement, a créé une **plateforme scientifique d'évaluation de la responsabilité et de la transparence des algorithmes**, ce projet ayant été placé sous la direction de Nozha Boujemaa, directrice de recherche à l'INRIA. La plateforme « Transalgo » a ainsi été mise en place en janvier 2017.

La disposition sur l'ouverture des données publiques introduite par la loi du 7 octobre 2016 pour une République numérique et la question de la transparence des algorithmes ont été abordées, le 15 novembre 2016, dans le cadre de **débats au Sénat** sur les « inégalités devant l'orientation après le bac ». Au cours de ce débat, notre collègue sénatrice Sylvie Robert a abordé la question de l'algorithme de répartition utilisé par la plateforme « Admission post-bac » (APB), dont les résultats semblaient refléter des inégalités subies par les candidats, notamment en fonction de leur origine sociale, dans l'orientation dans des filières d'enseignement supérieur.

### **Inégalités devant l'orientation après le bac (extrait des débats<sup>1</sup>)**

**Mme Sylvie Robert.** Madame la secrétaire d'État, une récente étude menée par l'INSEE dans l'académie de Toulouse souligne que les résultats d'admission post-bac reposent, dans une large mesure, sur un déterminisme social évident. À dossiers équivalents, les élèves issus de milieux favorisés s'orientent beaucoup plus vers les filières d'excellence ou les grandes écoles.

Plusieurs facteurs peuvent expliquer ce constat : les différences de ressources financières, la position sociale des parents, qui influe souvent sur le choix des enfants, l'asymétrie d'information concernant les établissements d'enseignement supérieur ou encore les disparités en matière d'orientation dans les lycées. Cette configuration tend à conférer un poids déterminant au capital social et culturel détenu par l'élève et sa famille.

Or le niveau de diplôme demeure un facteur prépondérant en matière d'insertion sur le marché du travail. À preuve, quatre ans après la sortie de la formation initiale, le taux de chômage des peu ou non diplômés, qui s'élève à 45 %, est quatre fois plus important que celui des diplômés du supérieur.

Pour remédier à cette situation, il se révèle donc essentiel d'agir en amont, en garantissant une égalité réelle devant l'orientation, laquelle n'est pas seulement un « processus de répartition des élèves dans différentes voies de formation », mais aussi « une aide dans le choix de leur avenir scolaire et professionnel », comme le rappelle le Haut Conseil de l'éducation.

À ce titre, il est reconnu que le système APB, admission post-bac, requiert un accompagnement et un suivi personnalisés de chaque élève.

Néanmoins, l'impossibilité parfois, pour l'élève, d'obtenir dans le cadre scolaire des informations pertinentes sur les filières et établissements envisagés, ainsi que des conseils quant aux stratégies à mettre en œuvre pour formuler ses vœux, constitue l'une des causes principales d'erreur, voire d'échec, d'orientation.

D'ailleurs, dans le rapport d'information sénatorial intitulé « Une orientation réussie pour tous les élèves », il est préconisé d'intégrer le conseil en orientation dans la formation initiale et continue des enseignants. Dans cette même perspective, les rectorats ont proposé des améliorations du système APB : ouvrir le dispositif à l'ensemble des filières sélectives ; abandonner le tirage au sort utilisé pour certaines formations, qui est source de frustration, d'injustice et parfois de contentieux ; associer au processus, dès la classe de première, l'élève et sa famille, afin de les familiariser à l'outil APB et de leur permettre d'anticiper et de réfléchir posément à l'orientation ; renforcer la transparence du système APB par la publication de son code source, conformément aux dispositions de l'article 2 du projet de loi pour une République numérique, qui crée un droit d'accès aux règles définissant le traitement algorithmique.

---

<sup>1</sup> L'intégralité des débats conduits lors de la séance du 15 novembre 2016 suite à la question n° 1489 adressée à Mme la ministre de l'éducation nationale, de l'enseignement supérieur et de la recherche au Sénat est disponible ici :

[https://www.senat.fr/seances/s201611/s20161115/s20161115\\_mono.html#cribkmk\\_questionnaire\\_1489\\_109137](https://www.senat.fr/seances/s201611/s20161115/s20161115_mono.html#cribkmk_questionnaire_1489_109137)

Je souhaiterais connaître la position du Gouvernement sur ces quelques pistes de réflexion. Par ailleurs, madame la secrétaire d'État, envisagez-vous de prendre d'autres mesures afin que tous les élèves puissent faire un choix éclairé et aient des chances égales, avec le système APB, de poursuivre ses études dans la filière et l'établissement supérieur de ses vœux ?

**Mme Clotilde Valter, secrétaire d'État auprès de la ministre du travail, de l'emploi, de la formation professionnelle et du dialogue social, chargée de la formation professionnelle et de l'apprentissage.** Madame la sénatrice Sylvie Robert, je suis mandatée par Mme la ministre de l'éducation nationale et par mon collègue Thierry Mandon pour répondre à votre question.

L'orientation des élèves est un des champs de réflexion et de travail du Gouvernement depuis 2012, dans la perspective de la lutte que nous menons contre le décrochage scolaire.

C'est dans ce cadre que le parcours Avenir a été mis en place, à la rentrée 2015, pour délivrer une information personnalisée à chaque élève, et ainsi favoriser l'élaboration d'une orientation cohérente. Cet accompagnement personnalisé en lycée, dispensé dès la classe de seconde, représente d'ores et déjà deux heures par semaine en moyenne.

Des actions ont également été engagées pour améliorer le continuum de formation bac-3/bac+3, telles que la généralisation du conseil d'orientation anticipé en classe de première, le renforcement du rôle de la commission académique des formations post-baccalauréat, l'amélioration de l'articulation des programmes du second degré et du supérieur par la rénovation en profondeur des programmes, le renforcement des passerelles et l'évolution de l'offre pédagogique.

Je tiens également à rappeler que le dispositif admission post-bac n'est, pour les élèves, qu'un outil d'expression des vœux. Le choix de l'orientation se fait bien évidemment en amont de la formulation de ces derniers sur le portail ; c'est le fruit d'une réflexion que l'élève mène avec l'aide de l'équipe pédagogique et grâce aux ressources de l'ONISEP, l'Office national d'information sur les enseignements et les professions.

Ce portail a fait l'objet d'évolutions importantes, qui visent à améliorer l'information et à permettre à chaque élève de formaliser un choix réfléchi, que ce soit en le confortant dans son choix ou en lui conseillant une autre orientation. De plus en plus, ce portail permet en effet aux élèves de recevoir un conseil. La très grande majorité des universités l'utilisent désormais pour formuler des avis : on recense plus de 500 000 avis ainsi délivrés par les universités au cours de la dernière année. Les équipes éducatives ont été formées à cet effet dans chaque académie, au niveau des bassins de formation des établissements.

Les actions mises en œuvre sur le terrain, à l'instar des Cordées de la réussite et des parcours d'excellence, lancés à la rentrée de 2016, doivent aussi être mentionnées.

Ces politiques commencent à porter leurs fruits : nous enregistrons des résultats extrêmement positifs, avec une baisse du nombre de jeunes sortis sans qualification, inférieur cette année à 100 000, le taux de jeunes de 18 à 24 ans non qualifiés étant désormais, dans notre pays, plus faible qu'en Allemagne ou au Royaume-Uni.

J'ai bien pris note, madame la sénatrice, des questions très précises que vous avez posées sur un certain nombre de points. Je ne suis pas en mesure d'y répondre, mais je les transmettrai à Mme la ministre de l'éducation nationale et à Thierry Mandon.

**Mme Sylvie Robert.** Madame la secrétaire d'État, je vous remercie vivement de cette réponse, et j'ai bien noté que mes questions précises obtiendront des réponses précises. Mon intention n'était vraiment pas de critiquer le système APB, qui s'est en effet beaucoup amélioré. Je souhaitais simplement souligner la difficulté que rencontrent certains élèves, et leurs familles avec eux, pour élaborer de façon libre et éclairée leur parcours professionnel.

### 3. Les divers régimes de responsabilité envisageables et ceux envisagés

Les régimes de responsabilité envisageables ont tendance à **mettre en cause soit le fabricant, soit le propriétaire, soit l'utilisateur**. Les **cas d'accident** risquent en effet de se multiplier à raison de la diffusion de systèmes autonomes, notamment de robots.

Vos rapporteurs jugent donc **nécessaire de se poser d'autres questions que celle d'une reconnaissance de la personnalité juridique des robots**.

L'association **EuRobotics** (« European Robotics Coordination Action »), en charge du programme de recherche de l'Union européenne en robotique avec l'objectif de favoriser le développement de la robotique en Europe, a proposé le 31 décembre 2012 un projet de **livre vert sur les aspects juridiques de la robotique**<sup>1</sup>. Dans son projet de livre vert, deux situations sont distinguées : celles où un robot cause un dommage du fait d'un défaut de fabrication et qui justifient une responsabilité du fait des produits défectueux<sup>2</sup> et celles où un robot cause un dommage dans le cadre de ses interactions avec des humains dans un environnement ouvert. Dans ce dernier cas, avec des robots de nouvelle génération dotés de capacité d'adaptation et d'apprentissage et dont le comportement présente un certain degré d'imprévisibilité, le régime de la responsabilité du fait des produits défectueux n'est pas approprié. Le cadre juridique applicable pourrait donc s'inspirer, selon les auteurs du livre vert, de deux régimes traditionnels (**responsabilité du fait des animaux** ou **responsabilité du fait d'autrui**, comme celle des parents si l'on choisit d'assimiler les robots cognitifs aux

<sup>1</sup> Cf. [http://www.eu-robotics.net/cms/upload/PDF/euRobotics\\_Deliverable\\_D.3.2.1\\_Annex\\_Suggestion\\_GreenPaper\\_ELS\\_IssuesInRobotics.pdf](http://www.eu-robotics.net/cms/upload/PDF/euRobotics_Deliverable_D.3.2.1_Annex_Suggestion_GreenPaper_ELS_IssuesInRobotics.pdf)

<sup>2</sup> La directive n° 85/374/CEE du Conseil du 25 juillet 1985 sur la responsabilité du fait des produits défectueux établit le principe de la responsabilité objective (responsabilité sans faute) du fabricant en cas de dommages provoqués par un produit défectueux.

enfants) ou, encore, d'un **code de conduite éthique applicable aux robots** et qui reste à rédiger.

Vos rapporteurs notent que quatre régimes de responsabilité pourraient en réalité trouver à s'appliquer aux accidents causés par des robots : le régime de **responsabilité du fait des produits défectueux**<sup>1</sup>, celui de la **responsabilité du fait des animaux**<sup>2</sup>, celui de la **responsabilité du fait d'autrui**<sup>3</sup>, ou, encore, celui, traditionnel<sup>4</sup>, de la **responsabilité du fait des choses**, mais qui ne s'applique que de façon résiduelle par rapport au régime de responsabilité du fait des produits défectueux.

Le robot est considéré comme une chose dans le droit civil français. **La responsabilité du fait des choses**, en tant que régime de responsabilité objectif codifié par l'ancien article 1384 du Code civil, **signifie que pour qu'elle soit appliquée, la chose doit être impliquée dans le dommage et qu'elle joue un rôle actif** (comme le fait d'être en mouvement ou de toucher la victime) dans l'occurrence dudit dommage. C'est l'individu considéré comme « gardien » de la chose qui est responsable de la réparation du dommage causé. Cependant, si le dommage est causé par une faille de sécurité du robot, le régime de responsabilité pour le dommage causé par la chose s'applique au fabricant du robot.

Selon Arnaud Touati et Gary Cohen<sup>5</sup>, **le régime de responsabilité du fait des choses laisse planer des incertitudes face à des biens autonomes**. En effet la jurisprudence requiert, pour appliquer ce régime, d'avoir la garde de la chose pour en être tenu responsable. Or cette garde se matérialise par un pouvoir de contrôle, de direction et d'usage. Mais alors que l'on conçoit facilement l'usage d'une intelligence artificielle (utiliser le logiciel, exploiter ses capacités), en avoir la direction et le contrôle semblent deux éléments beaucoup plus difficiles à envisager face à des systèmes d'intelligence artificielle autonomes, de surcroît non matérialisés physiquement<sup>6</sup>.

Concernant ce régime de responsabilité applicable à l'intelligence artificielle et aux robots, Rodolphe Gélén et Olivier Guilhem estiment, quant à eux, intéressant de noter que la responsabilité du fait des choses peut appréhender **certaines caractéristiques propres des robots comme leur polyvalence, leur capacité d'apprentissage et d'interaction**. En revanche, l'autonomie décisionnelle semble davantage poser problème. Si le robot agit

---

<sup>1</sup> Codifié aux articles 1386-1 et suivants du code civil, il vise à engager la responsabilité du producteur de robots dès lors que ces derniers, ayant causé un dommage, n'offrent pas « la sécurité à laquelle on peut légitimement s'attendre ». Ce critère de sécurité légitime présente l'intérêt d'être flexible.

<sup>2</sup> Pour les cas où un individu est déclaré responsable des actes dommageables commis par un animal dont il a la garde ou la propriété.

<sup>3</sup> Pour les cas où un individu est déclaré responsable des actes dommageables commis par un tiers.

<sup>4</sup> Issu du fameux ancien article 1384 du code civil.

<sup>5</sup> Article « Le droit à l'épreuve de l'intelligence artificielle » du 28 novembre 2016 paru dans la revue *Village de la Justice*.

<sup>6</sup> Même en présence d'une application matérielle de l'intelligence, tel qu'un robot, le problème de la garde reste posé, de sorte que l'utilisateur ne contrôle pas effectivement le système, il peut simplement l'allumer ou l'éteindre.

---

de façon autonome, qui est son gardien ? Le concepteur de son intelligence artificielle ou le propriétaire qui a réalisé son apprentissage ?

À ce niveau, le fait de mettre en place une **responsabilité en cascade** pourrait être envisagée. Dans la mesure où trois ou quatre acteurs sont en présence (le producteur de la partie physique du robot, le concepteur de l'intelligence artificielle, l'utilisateur et s'il est distinct de ce dernier, le propriétaire), il est possible d'imaginer que chacun puisse supporter une part de responsabilité selon les circonstances dans lesquelles est survenu le dommage. Arnaud Touati et Gary Cohen plaident de même pour offrir à l'intelligence artificielle un statut particulier, différent de celui réservé à la chose et protecteur en cas d'accident, du type « **chaîne de responsabilité** », allant du concepteur à l'utilisateur, en passant par le fabricant, le fournisseur et le distributeur. À l'heure où d'autres juristes, tel Alain Bensoussan, prônent la création d'une personnalité juridique autonome pour les systèmes d'intelligence artificielle, il est important d'**identifier des pistes qui ne fassent pas courir le risque de déresponsabiliser les acteurs du secteur, à commencer par les industriels de la robotique.**

En outre, il conviendrait de réfléchir à la possibilité d'instituer des **systèmes d'assurance spécifiques, voire des assurances obligatoires**. La Fédération Française de l'Assurance a ainsi mis en place dès la fin 2014 une commission spécialisée dans les questions du numérique qui a pour objectif de structurer un écosystème plus favorable au numérique tout en respectant les enjeux concurrentiels entre les assureurs. Cette Commission, composée de 26 représentants des sociétés d'assurances, est présidée par Virginie Fauvel, en charge du Digital & Market Management d'Allianz France. La commission a notamment pour mission :

- d'analyser les enjeux collectifs attachés à la transformation digitale pour le secteur,
- d'étudier les moyens de consolider la confiance entre les assureurs et les assurés dans cette transformation,
- de promouvoir l'innovation et notamment une réglementation adaptée et graduée (principe du bac à sable réglementaire).

Cette commission a lancé plusieurs actions concrètes en 2016, dont :

- une initiative pour le *Legal Design*, qui vise à lutter contre la surabondance de l'information, grâce à des infographies et vidéos permettant de rendre l'information juridique plus visuelle et plus facilement compréhensible par les assurés ;

- une rencontre avec une trentaine de *start-ups InsurTech* en décembre 2016, afin de renforcer les liens entre les assureurs et les « jeunes pousses » ;

- L'organisation d'une *Learning Expedition* en Silicon Valley et au CES de Las Vegas en janvier 2017.



---

Pour l'année 2017, plusieurs thématiques ont été définies comme prioritaires par la commission :

- *Véhicules connectés/autonomes* : une bonne utilisation des données des véhicules connectés permettra une meilleure prévention des risques d'accidents (ex : localisation de zones accidentogènes). Par ailleurs, il est nécessaire d'anticiper l'arrivée des véhicules autonomes afin de proposer des produits d'assurance adaptés ;

- *Blockchain* : cette technologie pourrait permettre de simplifier l'identification et la preuve d'assurance, ainsi que d'automatiser les procédures d'indemnisation (l'un des exemples étant l'indemnisation automatique des voyageurs en cas de retard d'avion) ;

- *Intelligence artificielle* : la puissance de calcul des ordinateurs et l'augmentation exponentielle du nombre de données vont permettre à l'intelligence artificielle d'offrir de très nombreuses applications nouvelles : reconnaissances vocales, reconnaissances d'images, assistants virtuels, véhicules autonomes...

Selon la Fédération Française de l'Assurance interrogée par vos rapporteurs, l'intelligence artificielle est un sujet naissant sur lequel les impacts ne sont pas encore connus. En termes de réglementation, il faudra par conséquent trouver un équilibre entre un encadrement qui ne bride pas l'innovation mais qui apporte suffisamment de protection aux consommateurs. De nouvelles questions vont émerger avec ces nouvelles technologies, et notamment certaines concernant l'assurance.

Mais la Fédération Française de l'Assurance estime qu'il est **encore trop tôt pour y répondre** car elle n'a pas aujourd'hui de visibilité suffisante sur les applications futures de ces technologies, elle assure qu'en tout état de cause **le droit et l'assurance accompagneront les nouveaux risques**, ce dont se réjouissent vos rapporteurs. Avec l'émergence de nouvelles formes d'intelligence artificielle et de robotique, il pourrait être **envisagé de mettre en place de nouveaux régimes d'assurance**, voire de créer une **assurance obligatoire**.

Enfin, **vos rapporteurs s'interrogent sur la question de la responsabilité juridique des algorithmes**, par exemple le cas d'un moteur de recherche pour les suggestions qu'il peut proposer à ses utilisateurs.

Dans l'arrêt n° 625 du 19 juin 2013 de la Première chambre civile de la Cour de cassation<sup>1</sup>, la plus haute juridiction judiciaire a en effet considéré que **Google ne pouvait pas être tenu pour responsable des mots proposés d'après un algorithme construit par ses soins**. L'explication du raisonnement des juges est le suivant : « *la fonctionnalité aboutissant au rapprochement critiqué est le fruit d'un processus purement automatique dans son fonctionnement et aléatoire dans ses résultats, de sorte que l'affichage des « mots clés » qui en résulte est exclusif de toute volonté de l'exploitant du moteur de*

---

<sup>1</sup> L'intégralité de l'arrêt rendu par la Cour de la Cassation est disponible ici : [https://www.courdecassation.fr/jurisprudence\\_2/premiere\\_chambre\\_civile\\_568/625\\_19\\_26825.html](https://www.courdecassation.fr/jurisprudence_2/premiere_chambre_civile_568/625_19_26825.html)

*recherche d'émettre les propos en cause ou de leur conférer une signification autonome au-delà de leur simple juxtaposition et de leur seule fonction d'aide à la recherche ».*

Dans la mesure où la saisie semi-automatique de Google fonctionne uniquement à partir d'algorithmes, la Cour de cassation estime qu'il n'est pas possible d'en déduire que la responsabilité de l'entreprise puisse être engagée. **Pour la Cour de cassation, les algorithmes ne sont donc pas coupables.**

#### **4. Les différenciations du droit applicable selon le type d'agents autonomes : robots industriels, robots de service, voitures autonomes et dilemmes éthiques afférents**

Le livre blanc « droit de la robotique » que le SYMOP a publié en 2016 contient d'utiles réflexions à ce niveau, notamment sur les robots industriels. La question de la sécurité des robots implique, en amont, l'établissement d'une définition de la collaboration et de l'interaction homme-robot. La **norme ISO 8373:2012** établit la **définition** de certains **termes caractérisant une interaction entre l'homme et le robot**<sup>1</sup>. Ainsi, à l'article 2.29 de la norme, l'interaction homme-robot est définie comme un « *échange d'information et d'actions entre l'homme et le robot pour exécuter une tâche, au moyen d'une interface utilisateur* », notamment au moyen d'échanges vocaux, visuels ou tactiles.

Le Parlement européen et le Conseil de l'Union européenne ont adopté la directive 2006/42/CE<sup>2</sup>, dite directive « Machines », dont l'objectif est d'assurer « *la libre circulation des machines au sein du marché intérieur tout en garantissant un haut niveau de protection de la santé et de la sécurité* », impliquant une **harmonisation des exigences de chaque État membre en termes de santé et de sécurité** concernant la conception et la production de machines. Les parties prenantes concernées par la directive Machines doivent respecter des obligations directement liées à leur statut :

- le fabricant est responsable de la conformité de la machine aux exigences de santé et de sécurité. La conformité du produit est certifiée par un marquage « CE » sur la machine ;
- l'importateur peut, en plus de transmettre à l'autorité de surveillance du marché certaines informations concernant le produit, assumer une responsabilité juridique par rapport au produit importé ;

<sup>1</sup> Cette norme ISO définit ainsi les termes de « *fonctionnement collaboratif* » (article 2.25), de « *robot de collaboration* » (article 2.26), et « *[d'] interaction homme-robot* » (article 2.29). L'intégralité du texte est disponible ici : <https://www.iso.org/obp/ui/#iso:std:iso:8373:ed-2:v1:fr>

<sup>2</sup> Directive 2006/42/EC du Parlement européen et du Conseil du 17 mai 2006 relative aux machines et modifiant la directive 95/16/CE. L'ensemble des dispositions contenues dans cette directive est accessible ici : <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:157:0024:0086:fr:PDF>

- 
- le distributeur doit veiller à la conformité des produits qu'il met sur le marché, et l'assembleur et l'installateur doivent veiller à ce que le produit demeure conforme.

Le **code du travail** contient également de **nombreuses dispositions concernant les exigences de santé et de sécurité**. Une partie de la directive 2006/42/CE a ainsi été transposée dans le droit français aux articles R. 4311-4 et suivants du code du travail. En outre, les différentes dispositions relatives aux exigences de santé et de sécurité présentes dans le code du travail affirment que **la sécurité et la protection de la santé des travailleurs sont assurées par l'employeur et s'appliquent dans le cadre de l'utilisation de robots industriels ou de services**. L'employeur doit respecter les exigences de conformité en vigueur et est tenu de prendre des mesures adaptées à l'utilisation de robots, telles que des actions de prévention des risques professionnels, des actions d'information et de formation, ou encore assurer des conditions d'utilisation sécurisée des robots.

Il est important de noter, comme le relève le livre blanc du Symop « *Droit de la robotique* », que « *la cour de Cassation a déjà retenu la responsabilité d'employeurs en cas d'infractions à la législation relative à la sécurité des travailleurs dans le cadre d'utilisation de robots* », notamment dans le cas du décès d'un travailleur dans une usine d'emballage<sup>1</sup> ou d'un employeur ayant fait travailler un salarié sur une ligne de fabrication robotisée sans prendre les mesures de sécurité nécessaires<sup>2</sup>.

Les robots de service posent différentes questions juridiques, dont les développements précédents ont montré que **les enjeux en termes de responsabilité ou de sécurité n'étaient pas insurmontables**.

En revanche, pour ce qui concerne les **voitures autonomes**, le besoin d'essais à grande échelle et en situation réelle appelle une **clarification du cadre juridique**. Il s'agit à la fois d'**enjeux économiques et de sécurité**.

Le cadre légal auquel sont soumises les expérimentations des voitures autonomes peut être examinée sous différents aspects. **Au regard du droit privé international**, certaines dispositions contenues dans la **Convention de Vienne sur la circulation routière du 8 novembre 1968**, comme **l'obligation de présence d'un conducteur et le contrôle de celui-ci sur le véhicule en mouvement**, peuvent constituer d'éventuels obstacles juridiques à la généralisation de voitures autonomes en France. Seuls les véhicules dotés de systèmes partiellement autonomes (comme les systèmes d'aide à la conduite) sont autorisés à la circulation sur la voie publique, et le chauffeur doit avoir le contrôle du véhicule. Cette convention internationale, bien qu'encore récemment amendée, nécessitera de nouvelles modifications du fait de l'intégration croissante de systèmes d'aide à la conduite, de systèmes autonomes et de systèmes d'intelligence artificielle dans les véhicules automobiles.

---

<sup>1</sup> Arrêt n°02-87666 du 30 septembre 2003 de la chambre criminelle de la Cour de cassation.

<sup>2</sup> Arrêt n°01-21192 du 16 septembre 2003 de la deuxième chambre civile de la Cour de cassation.

---

L'état actuel du droit communautaire peut également constituer un obstacle juridique à la circulation de véhicules intelligents sur la voie publique. L'adoption de la **directive 2010/40/UE du Parlement européen et du Conseil du 7 juillet 2010 concernant le cadre pour le déploiement de systèmes de transport intelligents dans le domaine du transport routier et d'interfaces avec d'autres modes de transport** a permis l'instauration d'un cadre juridique accélérant le déploiement des systèmes de transport intelligents, qui pourrait servir de modèle pour l'adoption d'une directive spécifique aux voitures numériques afin de coordonner les législations des États membres sur le déploiement des voitures intelligentes. Néanmoins, vos rapporteurs constatent que **le droit communautaire ne prévoit pour le moment pas de cadre normatif spécifique permettant l'harmonisation de l'expérimentation, du déploiement et de l'exploitation des véhicules intelligents au sein de l'espace européen.**

Cependant, **l'adoption du règlement (UE) 2015/758** du Parlement européen et du Conseil du 29 avril 2015 concernant les exigences en matière de réception par type pour le déploiement du système eCall embarqué fondé sur le service 112 **rendant obligatoire l'installation de terminaux permettant aux véhicules de communiquer entre eux (V2V) et avec les infrastructures de transport intelligent (V2I)**, permettent de poser les bases du déploiement des véhicules intelligents en Europe.

**Vos rapporteurs notent que les États-Unis disposent de lois autorisant l'expérimentation de voitures autonomes sur la voie publique**, au niveau des États du Nevada depuis juin 2011, de Floride depuis avril 2012, de Californie depuis septembre 2012, du district de Columbia depuis janvier 2013, et du Michigan depuis fin 2013. **D'autres projets de loi concernant les voitures autonomes sont en cours d'adoption dans une dizaine d'États.** Ces lois existantes ou en cours d'adoption fixent les conditions des tests sur la voie publique ainsi que les normes de sécurité applicables. Cependant, aucune harmonisation au niveau fédéral n'est à ce jour à l'étude.

En France, **l'article R. 412-6-I du Code de la route indique que tout véhicule en mouvement doit avoir un conducteur.** Dans l'état actuel de la législation française, le conducteur du véhicule est responsable en cas d'accident de la route. Cependant, cette législation ne peut être appliquée telle quelle aux accidents causés par des véhicules autonomes car le conducteur n'a pas le contrôle direct du véhicule.

La course à la voiture autonome et la perspective de ses débouchés massifs entre en tension avec le risque d'accident et surtout le flou juridique en matière de responsabilité. Le Gouvernement s'est vu confier la tâche de déterminer le régime juridique applicable et il n'y a donc pas lieu pour vos rapporteurs et l'OPECST d'interférer avec cette mission.

La loi n°2015-992 du 17 août 2015 relative à la transition énergétique pour la croissance verte contient des dispositions introduisant **le cadre**

---

**expérimental visant à promouvoir l'expérimentation et le déploiement de véhicules propres, incluant les voitures sans chauffeur.**

Cette loi habilite le Gouvernement à agir par ordonnance concernant l'autorisation d'expérimentation de véhicules à délégation partielle ou totale de conduite sur la voie publique. En ce sens, **l'ordonnance n° 2016-1057 du 3 août 2016** relative à l'expérimentation de véhicules à délégation de conduite sur les voies publiques **autorise l'expérimentation de véhicules intelligents sur la voie publique** sous condition de la délivrance d'une autorisation accordée par le ministre chargé des transports, après avis du ministre de l'intérieur. Pour mémoire, le projet « Nouvelle France industrielle », annoncé à la fin de l'année 2013 avec le but de réindustrialiser les territoires, anticipait l'arrivée des véhicules à pilotage automatique d'ici 2020.

Néanmoins, vos rapporteurs rappellent qu'**une mise en circulation effective de véhicules autonomes sur la voie publique soulève des questionnements éthiques**. Dans l'article « *The social dilemma of autonomous vehicles* »<sup>1</sup> paru le 24 juin 2016 dans le magazine *Science*, Jean-François Bonnefon, Azim Shariff et Iyad Rahwan affirment que le choix opéré par l'algorithme peut représenter de véritables dilemmes. Certains cas peuvent conduire l'algorithme à prendre une décision basée sur un critère moral qui lui aura été programmé à l'avance. Deux conceptions s'affrontent selon les auteurs : **une conception utilitariste**, qui postule qu'il faut **minimiser les pertes humaines**, et une **conception self-protective** qui postule que **les systèmes algorithmiques embarqués doivent avant tout protéger ses passagers à tout prix**.

Au cours des six études qu'ils ont menées, les auteurs ont constaté que les participants ont largement été en accord avec le fait qu'il était plus moral qu'un véhicule autonome sacrifie son passager si cela permettait de sauver un grand nombre de vies. Cependant, **face à des situations concrètes, de nombreux participants avaient la tentation de faire le choix du « passager clandestin »** en privilégiant les choix des véhicules autonomes protégeant ses passagers à tout prix. De fait, si des véhicules autonomes dotés de codes moraux utilitaristes et des véhicules étant programmés pour protéger leurs passagers étaient commercialisés, les participants orienteraient davantage leur choix vers les véhicules les protégeant à tout prix.

Le laboratoire du MIT « *moral machine* », visité par vos rapporteurs, travaille notamment sur les **dilemmes éthiques en voiture**. Le dilemme du tramway inventé par Philippa Foot a été réaffiné et testé. Les résultats provisoires des tests conduisent à identifier différents facteurs de choix : le nombre de tués (on préfère la solution qui réduit le nombre de morts), le fait de sacrifier en priorité des personnes qui transgressent les règles (exemple du voleur), le fait de sacrifier en priorité un animal contre un humain, le fait

---

<sup>1</sup> Les résultats de l'étude menée par les auteurs sont disponibles sur le site Internet du magazine *Science* : <http://science.sciencemag.org/content/352/6293/1573.full>

---

de sacrifier en priorité une personne plus âgée face à une personne plus jeune et *a fortiori* un enfant, le fait de sacrifier en priorité un homme face à une femme...

Le 6 février 2017 a été donné le coup d'envoi d'un **programme européen de trois ans baptisé « Autopilot »**. Versailles fera partie des cinq lieux d'expérimentation en Europe, avec l'objectif d'améliorer l'efficacité des véhicules autonomes grâce à l'exploitation des données externes, produites par l'infrastructure, les objets connectés... et les usagers. On compte 43 acteurs impliqués dans le projet, constructeurs automobiles et sous-traitants, acteurs des télécoms, instituts de recherche... On y trouve notamment PSA, IBM, Valeo, Continental, TomTom, Stmicro ou Thales. Cinq territoires ont été choisis pour tester des concepts de communication entre véhicules autonomes et systèmes d'information externes, en France, Finlande, Espagne, Italie et Pays-Bas. Un projet similaire sera lancé en Corée. Plusieurs types de configuration seront testés : conduite en milieu urbain, sur autoroute, stationnement autonome. En France, c'est Versailles qui accueillera les tests, sous l'impulsion de l'institut Vedecom, qui a mis au point un prototype fonctionnel de véhicule autonome de niveau 4 (100 % autonome dans des zones précises).

Par ailleurs, une autre expérimentation de voitures autonomes est conduite dans un **partenariat franco-allemand**, visant la mise en place à partir de mars 2017 du **premier site expérimental transfrontalier** de tests de voitures autonomes. Il s'agira en effet d'une zone allant de Metz à Merzig dans la Sarre, avec des tronçons d'autoroutes, de routes et de zones urbaines. Il s'agit d'un complexe ouvert à tous les constructeurs, équipementiers ou entreprises du numérique ou des télécommunications. La France et l'Allemagne entendent ainsi rattraper leur retard en matière d'expérimentation de voitures autonomes et se placer ensuite à l'avant-garde de la définition des futures règles applicables (standards ou réglementation).

Concernant l'assurance applicable aux voitures autonomes, il semble nécessaire de se doter à terme, sur un plan global et au moins dans tous les États membres de l'UE, d'un **système d'assurance obligatoire** afin de garantir le dédommagement total des victimes d'accidents causés par ce type de véhicules.

## C. LA PRISE EN COMPTE GRANDISSANTE DES ENJEUX ÉTHIQUES

### 1. Le cadre national de la réflexion sur les enjeux éthiques de l'intelligence artificielle

La place des systèmes d'intelligence artificielle et des machines utilisant ces technologies, notre dépendance à leur égard et la maîtrise que nous conservons de leur évolution sont des questions qui méritent d'être débattues dès aujourd'hui. Il convient d'anticiper les problèmes posés par l'intelligence artificielle. Comme il a été vu, la science-fiction, avec Isaac



Asimov, a réfléchi à la question et proposé des lois de la robotique, mais comment garantir, au-delà du statut juridique du robot et de ces lois, que ces technologies puissent être maîtrisées, utiles et conformes à nos valeurs ? La réflexion sur les enjeux éthiques de l'intelligence artificielle doit clarifier le cadre dans lequel s'inscrit la recherche en intelligence artificielle, ses usages ainsi que les limites éventuelles qu'il faut fixer à l'intelligence artificielle. Pour Gérard Sabah, les aspects pertinents sur lesquels l'éthique de l'intelligence artificielle doit réfléchir et se prononcer sont « *les impacts de telles machines sur la vie privée, sociale, politique et économique, ainsi que les valeurs qui les sous-tendent et celles qu'elles impliquent* ». Il poursuit en affirmant que « *la société doit définir clairement les limites acceptables entre la science et la fiction, le progrès et les risques encourus, afin de préserver notre identité et notre liberté* ».

La CERNA d'Allistene, déjà évoquée plusieurs fois, joue en la matière un rôle majeur, elle a d'ailleurs produit des rapports, dont il sera rendu compte plus loin. Sa création est récente et fait suite aux demandes parallèles en 2009 du Comité d'éthique du CNRS (COMETS) et de l'INRA. Le rapport du COMETS sur l'éthique des Sciences et technologies de l'information et de la communication (STIC) a particulièrement donné naissance à la CERNA<sup>1</sup>. Il convient de relever que le COMETS s'est, à plusieurs reprises, penché sur les problèmes éthiques posés par les STIC<sup>2</sup>.

Le rôle de la CERNA est cependant remis en cause par les **nouvelles missions dévolues à la CNIL**, suite à la loi du 7 octobre 2016 pour une République numérique<sup>3</sup>. Aux termes de cette loi, la CNIL, qui se définit comme l'autorité française de contrôle en matière de protection des données personnelles<sup>4</sup>, a en effet été **chargée de conduire une réflexion sur les questions d'éthique liées au numérique et aux algorithmes**, ce qui la mène à animer le débat public en la matière. Une page Internet dédiée a été créée<sup>5</sup>. Il s'agit aussi de s'intéresser aux questions de sociétés soulevées par l'évolution des technologies numériques. La CNIL a choisi d'y répondre par l'organisation de débats publics, d'ateliers et de rencontres. Selon elle, son rôle consiste à « *initier un processus de discussion collectif que feront vivre tous*

---

<sup>1</sup> Rapport du COMETS sur l'éthique des STIC : <http://www.cnrs.fr/comets/IMG/pdf/08-rapportcomets091112-2.pdf>

<sup>2</sup> Il est loisible de mentionner l'existence de rapports sur le partage des données ([http://www.cnrs.fr/comets/IMG/pdf/2015-05\\_avis-comets-partage-donnees-scientifiques-3.pdf](http://www.cnrs.fr/comets/IMG/pdf/2015-05_avis-comets-partage-donnees-scientifiques-3.pdf)), sur le contrôle des publications scientifiques avec les nouveaux médias (<http://www.cnrs.fr/comets/IMG/pdf/mediaaviscometsavril16-2.pdf>) ou, encore sur les sciences citoyennes ([http://www.cnrs.fr/comets/IMG/pdf/comets-avis-entier-sciences\\_citoyennes-25\\_juin\\_2015.pdf](http://www.cnrs.fr/comets/IMG/pdf/comets-avis-entier-sciences_citoyennes-25_juin_2015.pdf)).

<sup>3</sup> Loi n° 2016-1321 du 7 octobre 2016 pour une République numérique.

<sup>4</sup> Sa mission historique, conformément à la loi n° 78-17 du 6 janvier 1978 modifiée à plusieurs reprises, dont par la loi pour une République numérique précitée, est de veiller à ce que l'informatique soit au service du citoyen et qu'elle ne porte atteinte ni à l'identité humaine, ni aux droits de l'Homme, ni à la vie privée, ni aux libertés individuelles ou publiques. Elle assure ainsi la protection des données à caractère personnel dans les traitements informatiques mis en œuvre sur le territoire français.

<sup>5</sup> Cf. <https://www.cnil.fr/fr/ethique-et-numerique>



*ceux – institutions publiques, société civile, entreprises – qui souhaitent y prendre part en organisant des débats et manifestations multiformes* ». Elle a ainsi mis en place une plateforme<sup>1</sup> pour contacter l'équipe de la mission « éthique et numérique », afin de permettre à toute « institution publique, membre de la société civile ou entreprise » de pouvoir prendre part au débat sur les algorithmes. Elle assurera la coordination et la cohérence des diverses manifestations.

Un cycle de débats publics intitulé « *Les algorithmes en débat* » est ainsi organisé par la CNIL en 2017. Pour la CNIL, la réflexion doit porter cette année sur « *les algorithmes à l'heure de l'intelligence artificielle* ». Elle retient que « *ceux-ci occupent dans nos vies une place importante, bien qu'invisible. Résultats de requêtes sur un moteur de recherche, ordres financiers passés par des robots sur les marchés, diagnostics médicaux automatiques, affectation des étudiants à l'Université : dans tous ces domaines, des algorithmes sont à l'œuvre. Ces derniers mois, le sujet des algorithmes s'est invité dans le débat public et a suscité une forte attention médiatique* ». Différentes questions sont posées<sup>2</sup>.

À l'automne 2017, la CNIL rendra publique la synthèse des échanges et des contributions. Il s'agira d'établir une « *cartographie de l'état du débat public* » et un « *panorama des défis et enjeux* ». Des **pistes ou des propositions pour accompagner le développement des algorithmes dans un cadre éthique** pourraient faire par la suite l'objet d'arbitrages par les pouvoirs publics.

**L'articulation et la complémentarité entre le travail de la CERNA et celui de la CNIL sont à rechercher.** À ce stade, vos rapporteurs relèvent que les manifestations envisagées<sup>3</sup> sont le plus souvent organisées conjointement par la CNIL avec la CERNA, le COMETS, l'AFIA, Universcience ou, encore, le Genotoul de Toulouse (plateforme éthique et bioscience).

Le Conseil général de l'économie (CGE) a rendu le 15/12/2016 un rapport au ministre de l'économie et des finances portant sur les « *Modalités de régulation des algorithmes de traitement des contenus* »<sup>4</sup>. Ses auteurs, Jacques Serris et Ilarion Pavel, montrent que les algorithmes de traitement des contenus sont inséparables des données qu'ils traitent et des plateformes qui les utilisent pour proposer un service. Mais alors qu'il y a de **nombreux travaux sur la protection des données et sur la loyauté des plateformes, il y**

<sup>1</sup> Cf. <https://www.cnil.fr/fr/webform/contacter-lequipe-de-la-mission-ethique-et-numerique>

<sup>2</sup> *S'agit-il d'une nouvelle révolution industrielle, ou d'un simple moyen d'améliorer la productivité ? Les algorithmes sont-ils les nouveaux décideurs ? Ont-ils pour effet de nous enfermer dans une bulle informationnelle, mettant en danger ouverture culturelle et pluralisme démocratique ? Sont-ils au contraire un moyen d'accéder à des idées, contenus, données ou personnes inaccessibles ou invisibles jusqu'alors ? Quelle transparence à l'ère des algorithmes : comment concilier transparence et propriété intellectuelle ? Faut-il repenser, face aux progrès de l'intelligence artificielle, la responsabilité des acteurs publics et privés ? Comment construire le libre-arbitre dans un monde « algorithmé » ?*

<sup>3</sup> Cf. <https://www.cnil.fr/fr/les-partenaires-et-evenements>

<sup>4</sup> <http://www.economie.gouv.fr/cge/modalites-regulation-des-algorithmes-traitement-des-contenus>

---

**en a encore peu sur les algorithmes eux-mêmes.** Ceux-ci sont pourtant des moteurs d'innovations, avec la révolution des réseaux neuronaux et de l'apprentissage profond. Ce rapport ne propose pas une nouvelle régulation sectorielle qui s'appliquerait aux algorithmes. En revanche, il souligne qu'il faut **développer la capacité à tester et contrôler les algorithmes** – tout en préservant l'innovation. Ses auteurs proposent cinq pistes d'action qui ont pour objet **la montée en compétence et le développement de l'expertise des pouvoirs publics**, mais aussi d'appeler au **développement de bonnes pratiques** dans les différents secteurs économiques. Ils soulignent aussi qu'il faut **préserver une image positive des technologies** utilisées pour concevoir ou opérer des algorithmes. C'est essentiel pour continuer à attirer les jeunes générations de françaises et de français dans des filières de formation exigeantes (mathématiques, ingénieurs ou *data scientists*) où la France est aujourd'hui bien placée.

Vos rapporteurs relèvent que l'INRIA avec sa **plateforme « Transalgo »**, placée sous la direction de Nozha Boujemaa, développe en 2017 de manière utile une **plateforme scientifique d'évaluation de la responsabilité et de la transparence des algorithmes**<sup>1</sup>, afin de répondre aux préoccupations exprimées d'**explicitabilité des algorithmes**.

Pour la CERNA, dans son **rapport sur l'éthique de la recherche en robotique**, le **respect de la vie privée** doit être une priorité dans la mesure où les systèmes d'intelligence artificielle et les robots posent de nouvelles difficultés<sup>2</sup>. La conception des robots doit donc intégrer **l'exigence de confidentialité des données personnelles qu'ils traitent**.

Ce rapport remarque également que les capacités d'autonomie des systèmes d'intelligence artificielle et des robots portent surtout actuellement sur l'autonomie opérationnelle, mais **demain leur autonomie sera de plus en plus décisionnelle**, issue de systèmes plus élaborés. Les robots auront, de plus, outre les développements en informatique, la possibilité **d'une plus grande ressemblance avec l'être humain**, comme tente de le montrer d'ores et déjà Hiroshi Ishiguro au Japon. **Les interrogations éthiques sur les finalités d'un tel projet d'intelligence artificielle humanoïde et sur ses effets s'imposent**, surtout que la ressemblance avec l'humain renvoie à l'hypothèse de la « Vallée de l'étrange », introduite par Masahito Mori en 1970, sorte de malaise ressenti par les êtres humains face à des entités presque semblables à eux, mais pas au point de s'y tromper.

La CERNA a formulé **neuf préconisations générales, sept sur l'autonomie, cinq sur l'imitation du vivant et quatre sur l'homme augmenté**. Vos rapporteurs les rappellent ici de manière synthétique :

---

<sup>1</sup> Cf. <https://www.inria.fr/actualite/actualites-inria/transalgo>

<sup>2</sup> Ces systèmes ont la capacité de capter des données personnelles (photos ou vidéos de personnes, voix, paramètres physiologiques, géolocalisation...), leur déploiement soulève donc des questions liées à la protection de la vie privée et des données personnelles.

---

- lorsque le chercheur s'exprime en public sur une question de société relative à son activité professionnelle, il doit **distinguer son intervention experte de l'expression de son opinion personnelle** ;

- les établissements de recherche se dotent de **comités opérationnels d'éthique en sciences et technologies du numérique** ;

- les établissements de recherche et les acteurs concernés mettent en place des groupes de travail et des projets de recherche **interdisciplinaires** ouverts à l'international et incluant **des chercheurs et des juristes** pour traiter des aspects juridiques des usages de la robotique ;

- les établissements de recherche et les acteurs concernés mettent en place des **actions de sensibilisation et de soutien** auprès des chercheurs et des laboratoires de recherche dans le numérique. Lors de l'élaboration et dans la conduite de ses projets le chercheur saisira, si nécessaire, le comité opérationnel d'éthique de son établissement ;

- lors de la conception d'un système numérique ayant la capacité de capter des **données personnelles**, le chercheur se demandera si ce système peut être équipé de dispositifs facilitant le **contrôle de sa conformité à la réglementation** lors de sa mise en usage ;

- le chercheur veillera à prendre en compte l'exposition potentielle de ses recherches à des **attaques numériques** ;

- si le chercheur considère que le projet vise un développement pouvant avoir **un impact important sur la vie des utilisateurs**, il veillera à **en délibérer** avec les acteurs et les utilisateurs potentiels afin d'éclairer au mieux les choix scientifiques et technologiques ;

- le chercheur veillera à **documenter l'objet ou le système conçu et à en exposer les capacités et les limites**. Il sera attentif aux retours d'expérience à tous les niveaux, du développeur à l'utilisateur ;

- le chercheur veillera à faire une **communication mesurée et pédagogique** sachant que les capacités des objets et systèmes qu'il conçoit peuvent susciter des questionnements et des interprétations hâtives dans l'opinion publique ;

- concernant l'autonomie, le chercheur doit se poser la question des **reprises en main que l'opérateur ou l'utilisateur peut effectuer** et étudier la **possibilité ou non laissée à l'humain de « débrayer » les fonctions autonomes du robot**. Il doit faire en sorte que les décisions du robot ne soient **pas prises à l'insu de l'opérateur**, être conscient des phénomènes de **biais de confiance**, et être attentif à **explicitier les limites des programmes de perception, d'interprétation et de prise de décision**, en particulier les programmes qui visent à conférer une conduite morale au robot. Le chercheur doit évaluer **jusqu'à quel point les logiciels d'interprétation du robot peuvent caractériser correctement une situation** et discriminer entre plusieurs situations qui semblent proches, **surtout si la décision d'action est fondée uniquement sur cette caractérisation**. Il faut en particulier évaluer

---

comment les incertitudes sont prises en compte. Le chercheur doit analyser **la prévisibilité du système humain-robot** considéré dans son ensemble, en prenant en compte les incertitudes d'interprétation et d'action, ainsi que les **défaillances possibles du robot et celles de l'opérateur**, et analyser l'ensemble des états atteignables par ce système. Il doit intégrer des **outils de traçage** dès la conception du robot. Ces outils doivent permettre d'élaborer des explications, même limitées, à plusieurs niveaux selon qu'elles s'adressent à des experts de la robotique, à des opérateurs ou à des utilisateurs ;

- en matière d'émotions, le chercheur étudiera, au regard des fonctions utiles du robot, **la pertinence et la nécessité de susciter des émotions et de recourir à des aspects ou des comportements biomimétiques**, notamment dans les cas de forte ressemblance visuelle ou comportementale entre un robot et un être vivant. Dans les cas où l'apparence ou la voix humaines sont imitées, le chercheur s'interrogera sur **les effets que pourrait avoir cette imitation**. Le chercheur doit avoir conscience que la démarche biomimétique peut **brouiller la frontière entre un être vivant et un artefact**. Le chercheur consultera sur ce brouillage le comité opérationnel d'éthique de son établissement ;

- pour les projets de recherche qui ont trait au développement de la robotique affective, le chercheur s'interrogera sur les répercussions éventuelles de son travail sur les **capacités de socialisation de l'utilisateur** ;

- pour les projets qui mettent en présence des enfants et des robots, le chercheur doit se poser la question de **l'impact de l'interaction enfant-robot** sur le développement des capacités émotionnelles de l'enfant, tout particulièrement dans la petite enfance ;

- pour les projets de recherche relatifs à des robots susceptibles d'avoir des effets sur **l'affectivité des utilisateurs et de susciter leur attachement**, le chercheur devra élaborer un **protocole de conception et d'évaluation** en veillant à impliquer les compétences multidisciplinaires nécessaires et des utilisateurs potentiels ;

- le chercheur doit être **prudent dans sa communication sur les capacités émotionnelles des robots et sur l'imitation de la nature et du vivant**, notamment parce que l'expression des émotions, au sens humain, par un robot, est un **leurre**, et parce que l'imitation du vivant peut amener, volontairement ou pas, à **prêter à l'artefact des caractéristiques du vivant** ;

- les chercheurs en robotique réparatrice ou d'assistance doivent appliquer, en coordination avec les professionnels de santé, les aidants et les patients, **les principes d'éthique en usage dans le secteur médical** afin d'arbitrer entre les exigences d'efficacité et de sécurité des soins, celles d'autonomie et d'intégrité de la personne et, enfin, de protection de la vie privée. Ces questions relèvent de l'éthique et non uniquement du droit en cela qu'elles demandent à être arbitrées dans chaque cas particulier et qu'elles ne reçoivent pas de réponse générale. Pour en traiter, il faudra prendre avis auprès des comités opérationnels d'éthique des sciences

médicales et veiller à ce que les compétences technologiques y soient étroitement associées. Dans le cas d'organes robotisés à vocation réparatrice, le chercheur aura le **souci de la préservation de l'autonomie de l'individu équipé**, à savoir de la maîtrise qu'il conservera autant que faire se peut sur ses actions, et de la **conservation de l'intégrité des fonctions autres que celles concernées par la réparation**. Dans le cas des dispositifs robotisés visant l'augmentation, le chercheur veillera à la **réversibilité** de celle-ci : les dispositifs doivent être **amovibles sans dommages** pour la personne, autrement dit, sans que la personne perde l'usage de ses fonctions initiales. En vue de **prévenir les discriminations induites par l'augmentation**, le chercheur se posera la question de l'incidence de l'augmentation des facultés et des capacités humaines induites par les dispositifs qu'il développe sur le comportement social de ceux qui en bénéficient ainsi que, symétriquement, de ceux qui n'en bénéficient pas.

La CERNA a formulé dans son second rapport, intitulé « *Éthique en apprentissage machine* », et rendu public en 15 mars 2017, ses préconisations sur le *machine learning* structurées autour de six thèmes directeurs : les données des systèmes d'apprentissage (1-4), l'autonomie des systèmes apprenants (5-6), l'explicabilité des méthodes d'apprentissage et leur évaluation (7-9), les décisions des systèmes d'apprentissage (10), le consentement dans le domaine du numérique (11), la responsabilité dans les relations homme-machine (12-13), et l'organisation de la recherche française sur l'éthique du numérique (14-17) :

1. Les concepteurs et entraîneurs des systèmes d'apprentissages veillent à la **qualité des données d'apprentissage** et aux conditions de leur captation.
2. Les entraîneurs doivent garantir que **les données représentent un miroir de la diversité culturelle**.
3. **Les variables dont les données sont réglementées**, les entraîneurs doivent veiller à ce qu'elles ne soient pas discriminantes (âge, sexe, race, etc.), tout en respectant le principe de confidentialité des données.
4. Le concepteur d'un système d'apprentissage automatique doit prévoir des dispositifs de **traçabilité du système**.
5. La machine ne doit pas introduire de biais de caractérisation et induire en erreur l'utilisateur sur l'état de son système.
6. Le concepteur doit maintenir un certain niveau de **vigilance dans la communication** sur les capacités d'un système apprenant, afin de ne laisser aucune place à l'interprétation ni à des fantasmes ou craintes irrationnelles.
7. Le concepteur doit veiller à l'**explicabilité**, la transparence des actions de son système apprenant, tout en maintenant un niveau de performance suffisant.

8. Tout en garantissant une meilleure explicabilité du système, le concepteur doit décrire les limitations des **heuristiques d'explication** du système, en évitant notamment la création de biais.
9. Le concepteur d'un système apprenant apporte sa contribution à l'**élaboration des normes** et des protocoles d'évaluation de l'apprentissage machine.
10. Le concepteur doit garantir la **place de l'humain dans les décisions assistées par des machines apprenantes**, afin d'éviter notamment la création de biais ou l'installation de dépendance de l'humain par rapport aux décisions des machines.
11. La **mémorisation des traces** des données personnelles utilisées dans le processus d'apprentissage devra obtenir le consentement de l'utilisateur et en accord avec la législation sur la protection des données personnelles en vigueur.
12. Le concepteur du système doit y inclure des **mécanismes de contrôle** automatiques ou supervisés.
13. Le concepteur doit fournir une **déclaration des intentions d'usage** du système informatique « de manière sincère, loyale et complète » au cours de son apprentissage.
14. La création d'un réseau national de recherche dénommé « *Initiative Fédérative de Recherche Numérique, Éthique et Société* » permettrait de faire émerger un positionnement français sur les questions d'impact sociétal et éthique des sciences et technologies du numérique.
15. La création de **comités d'éthique opérationnels d'établissements** en science et technologie du numérique est conseillée.
16. Les établissements sont également incités à lancer des **initiatives sur les aspects juridiques** des usages des innovations du numérique, au travers de groupes de travail et projets de recherches avec d'autres acteurs concernés.
17. Des actions de **sensibilisation et soutien du chercheur par les établissements** doivent être mises en place.

Le rapport produit par la CERNA présente, au total, des **recommandations plus opérationnelles que celles fournies** par les autres structures ayant rendu public des rapports similaires, telles que l'association mondiale des ingénieurs électriciens et électroniciens (*Institute of Electrical and Electronics Engineers* ou IEEE), qui regroupe plus de 400 000 membres, ou les « 23 points d'Asilomar » issus de la conférence « *Beneficial AI 2017* » (initiatives qui seront évoquées plus loin dans le présent rapport). La singularité du rapport de la CERNA s'observe également par le fait que les **recommandations mises en avant concernent, pour la majeure partie d'entre elles, des aspects techniques du développement de l'apprentissage automatique**, alors que le rapport de l'IEEE et les « 23 points d'Asilomar » abordent des problématiques plus vastes, telles que la question des

---

investissements, des relations entre les scientifiques et les décideurs, la course à l'innovation entre les chercheurs, le bénéfice collectif, la vie privée, la défense ou encore les problèmes économiques et humanitaires.

En outre, les 17 propositions avancées par ce rapport de la CERNA misent davantage sur la **pluridisciplinarité**, qui est, selon les auteurs, **essentielle à la réflexion sur les considérations éthiques, juridiques et scientifiques de l'apprentissage automatique**.

Le Club informatique des grandes entreprises françaises (CIGREF), créé en 1970 à l'initiative de dirigeants de grandes entreprises, a lui aussi réfléchi aux questions éthiques posées par l'intelligence artificielle et la robotique. Deux rapports ont ainsi été rendus publics, dont vos rapporteurs ont rencontré les auteurs. Le premier porte sur le thème de « l'éthique du numérique »<sup>1</sup> et a été rédigé par Flora Fischer, chercheuse en philosophie des technologies à la Sorbonne et chargée de recherche au CIGREF, que vos rapporteurs ont eu le plaisir d'auditionner. Le second, qui se veut livre blanc, porte sur la « Gouvernance de l'intelligence artificielle dans les grandes entreprises », et a été réalisé en partenariat avec le cabinet Alain Bensoussan Avocats<sup>2</sup>.

Le premier rapport du CIGREF sur « l'éthique du numérique » montre que la technologie numérique est à la fois relationnelle, d'usage et fabriquée, comme toute technologie. L'éthique du numérique est donc à la fois une éthique des usages et de la conception. Flora Fischer aime à rappeler que le philosophe Gilbert Simondon disait que « *toutes les technologies sont des médiations* » et que le numérique crée un nouveau rapport au monde, dans lequel il faut simplement être attentif à l'éthique des nouveaux usages et à la démultiplication. Beaucoup d'entreprises qui naissent aujourd'hui avec le numérique, tel que les *start-ups*, mettent déjà en œuvre la *privacy by design*, et l'éthique *by design* ce qui suppose d'anticiper les usages et la façon dont tels ou tels outils vont adapter les pratiques et de voir quelles questions éthiques cela va engendrer. Pour respecter ces objectifs éthiques, il faut anticiper les usages des outils dès la conception et prévoir des architectures souples, suffisamment pour agir rétroactivement sur les usages et sur la conception. Il faut aussi explorer les limites des nouveaux outils et des nouveaux services. Les GAFAs ayant été les pionniers, ils ont imposé leurs propres règles, ce qui pourra justifier une régulation, par exemple en matière de stockage et de traitement des données privées. Les entreprises qui vont mettre en œuvre des services numériques passeront par la *privacy by design*, et l'éthique *by design*, ce qui sera plus facile pour des jeunes et petites entreprises que pour des grandes entreprises plus anciennes. Cela nécessitera une révision de l'architecture des plateformes. Dans ce contexte général, l'État ne peut et ne pourra qu'agir *a posteriori*.

---

<sup>1</sup><http://cigref.fr/Publication/2014-CIGREF-Ethique-et-Numerique-une-ethique-a-reinventer-Rapport-mission-F-FISCHER.pdf>

<sup>2</sup><http://www.cigref.fr/wp/wp-content/uploads/2016/09/Gouvernance-IA-CIGREF-LEXING-2016.pdf>



---

Le second rapport du CIGREF, le livre blanc sur la « *Gouvernance de l'intelligence artificielle dans les grandes entreprises* », apporte une **vision prospective** visant à permettre aux entreprises d'anticiper le passage de la transition digitale que nous vivons déjà à la « transition intelligente ». Il présente l'intérêt de partir du cadre historique de l'intelligence artificielle, en passant par ses définitions (générale et technique) et de ses modes d'expression (robots, avatar, *chatbots*...), afin de montrer que l'intelligence artificielle confronte le management de l'entreprise à des situations émergentes multiples, à la fois culturelles, humaines (impact sur les manières de travailler), éthiques et juridiques... Les entreprises devront être en mesure d'anticiper, par exemple, l'évolution des compétences, l'évolution de la réglementation, que ce soit sur la robotique intelligente ou sur les différentes formes de l'intelligence artificielle. Le Livre Blanc aborde également la question du **droit prospectif** : l'intelligence artificielle, du fait de son autonomie, a un degré d'imprévisibilité dans le cadre de son interaction avec les êtres humains, or, en l'état actuel du droit, aucune règle ne serait directement applicable à la responsabilité délictuelle de l'intelligence artificielle, ce qui peut apparaître discutable. Des craintes liées au développement de l'intelligence artificielle font débat et posent questions. Par exemple, la délégation de tâches à haute responsabilité (décision, recommandation) à des machines interroge sur le libre arbitre et la place laissée à la pertinence de l'interprétation humaine. L'entreprise ne saurait donc contourner des **questions éthiques** de deux ordres : l'éthique des usages et l'éthique de la conception (*by design*)... ». L'enjeu est de « *saisir la complexité du sujet et mieux comprendre les freins et leviers à actionner pour accompagner au mieux les opportunités à venir* ».

Vos rapporteurs ont, en outre, rencontré les animateurs et des chercheurs du **projet ETHICAA** (Éthique et Agents Autonomes), financé par l'Agence nationale de la recherche (ANR) pour la période 2014 - 2017, dont les rapports techniques et les publications sont disponibles<sup>1</sup>. Coordonné par Grégory Bonnet (GREYC - Normandie Université), il associe différents partenaires<sup>2</sup> et chercheurs.

Les **objectifs initiaux du projet peuvent être rappelés**. Les machines et les logiciels (agents) deviennent de plus en plus autonomes et agissent de plus en plus sans être contrôlés par des utilisateurs ou des opérateurs humains. C'est pourquoi, la question de doter ces agents autonomes de comportements éthiques se pose. L'objectif du projet ETHICAA est de définir ce que devrait être un système composé d'un ou plusieurs agents artificiels capables de gérer des conflits éthiques, aussi bien au niveau individuel qu'au niveau collectif. Il y a en effet des verrous scientifiques majeurs. En premier lieu, les théories éthiques sont difficiles à mettre en

---

<sup>1</sup> <https://ethicaa.org>

<sup>2</sup> Les responsables scientifiques du projet pour les partenaires sont : Alain Berger (société Ardans) ; Olivier Boissier (Institut Henry Fayol - ARMINES) ; Pierre-Antoine Chardel (Institut Mines Télécom) ; Jean-Gabriel Ganascia (LIP6 - Université Paris 6) ; Catherine Tessier (ONERA).

---

œuvre sous forme de principes éthiques opérationnels. En second lieu, ces principes éthiques opérationnels sont eux-mêmes difficiles à implanter parce qu'ils sont liés à l'évaluation courante de la situation, dont l'automatisation rencontre de fortes limites. En troisième lieu, d'un point de vue philosophique, il existe de nombreux principes éthiques et aucun d'eux n'est meilleur que les autres, rendant ainsi difficile de choisir celui qui doit être mis en œuvre. Enfin, les systèmes informatisés sont de plus en plus ouverts et décentralisés, c'est-à-dire impliquant des agents artificiels autonomes en interaction avec d'autres agents, des opérateurs ou des utilisateurs humains. Dans ces circonstances, la gestion des conflits éthiques entre différents agents devient une question cruciale et des méthodes originales sont nécessaires pour répondre à cette problématique.

**Un état des lieux des travaux peut être fait en mars 2017 :** les travaux du projet ETHICAA se structurent en trois points : (1) une réflexion autour des concepts éthiques à employer et des domaines d'application sensibles ; (2) une production de modèles de décision et de raisonnement éthiques ; (3) une fédération d'une communauté de recherche autour de ces thématiques.

1. Dans le domaine de la réflexion, deux études ont été produites :

- Un état de l'art dans le domaine de la philosophie morale et de l'intelligence artificielle nous ont permis de formuler des définitions : agent artificiel éthique (cadre idéal non réalisable en pratique) ; agent artificiel éthique compétent (agent capable de justifier ses actes en fonction de critères explicites) ; situation de conflit éthique.
- Une identification de quatre scénarios-clés intéressants à des fins de modélisation et d'expérimentation : un scénario de véhicule autonome, un scénario de véhicule aérien piloté en tandem par un agent artificiel et un agent humain, un scénario d'agent de surveillance médicale, un scénario d'organisation et de coopération d'agents de gestion de portefeuilles.

2. Dans le domaine de la production, deux approches ont été mises en œuvre et expérimentées sur une partie des scénarios mentionnés précédemment :

- Un travail sur des architectures d'agents se fondant sur l'utilisation des logiques d'actions, de logique modale et de logique argumentative pour modéliser et expliquer *a priori* un raisonnement moral.
- Un travail sur la vérification formelle de propriété éthique ayant pour but de vérifier *a posteriori* si la spécification d'un système multi-agent répond à des critères éthiques.

3. Dans le domaine de la fédération d'une communauté de recherche, le projet s'axe sur deux points :

- Une interaction avec la communauté internationale de recherche par l'organisation d'un atelier international ainsi qu'à la participation au dépôt

---

d'une action COST *Responsible Artificial Intelligence* et à la IEEE *Global Initiative on Ethical Considerations in the Design of Autonomous Agents*.

- La diffusion grand public au travers d'articles et conférences de vulgarisation. À cela, s'ajoutent de nombreuses interventions en séminaires, journées d'études, tables rondes et interventions radiophoniques.

**Quant aux perspectives du projet ETHICAA**, il s'agit de rassembler les différents modèles de décision produits au sein d'un même cadre de conception auquel s'ajoutera la dimension multi-agent, offrant ainsi une grille de lecture unifiée pour concevoir des méthodes répondant aux problématiques de régulation éthiques d'agents autonomes. D'un point de vue de développement, le projet ETHICAA envisage de finaliser un **logiciel de démonstration orchestrant une simulation du scénario de gestion éthique d'actifs financiers**, il s'agit en effet de finaliser une preuve de concept à travers l'organisation et la coopération d'agents de gestion de portefeuilles selon des règles éthiques.

## 2. Les nombreuses expériences anglo-saxonnes de réflexion sur les enjeux éthiques de l'intelligence artificielle

Les **expériences de réflexion sur les enjeux éthiques de l'intelligence artificielle**, que ce soit aux États-Unis ou au Royaume-Uni, sont particulièrement **nombreuses** et se sont **multipliées de façon impressionnante** dans la période récente. Ces **expériences de réflexion** sont le plus souvent **non gouvernementales** et sont fréquemment financées par des **donations privées** : on peut relever que cette observation s'inscrit dans le contexte d'une **tradition de mécénat scientifique et technologique** aux États-Unis, alors qu'en France le mécénat reste largement consacré à l'art et à la culture.

Vos rapporteurs ont, en effet, observé une multiplication surprenante d'initiatives anglo-saxonnes, souvent **coordonnées**, visant la prise en compte de principes éthiques dans la recherche et les usages de l'intelligence artificielle. Ils souhaitent les rappeler.

Il existe tout d'abord une structure plus ancienne que les autres, qui reste tout de même jeune puisqu'elle a été créée en 2000, *The International Society for Ethics and Information Technology* (INSEIT), qui édite aussi sa propre revue. Cette structure, qui rassemble beaucoup de chercheurs du monde académique, est de plus en plus éclipsée par la multiplication récente des collectifs ou des instituts se donnant pour rôle d'animer une réflexion sur les enjeux éthiques du numérique et plus spécifiquement de l'intelligence artificielle. Ces initiatives plus récentes sont souvent financées par des **fonds privés issus de grandes entreprises du secteur et ont recours à des plans de communication aboutis**.

L'une des principales initiatives est l'**Institut du futur de la vie** ou « *Future of Life Institute* » (FLI)<sup>1</sup>, fondé en mars 2014, qui est à l'origine en janvier 2015 de la lettre d'avertissement sur les dangers potentiels de l'intelligence artificielle, qui affirmait qu'étant donné le grand potentiel de l'intelligence artificielle, « *il était important d'étudier comment la société peut profiter de ses bienfaits, mais aussi comment éviter ses pièges* ». Le FLI s'interroge ainsi sur les **conséquences économiques, légales et éthiques de l'intelligence artificielle** et de l'automatisation des tâches et promeut le développement d'une intelligence artificielle bénéfique et fiable.

Le FLI, visité par vos rapporteurs en janvier 2017, situé à Cambridge, près de Boston (avec le MIT et Harvard), est une organisation à but non lucratif, dont le financement repose sur d'importantes donations privées. Il se donne pour mission de « *catalyser et soutenir la recherche et les initiatives visant la sauvegarde de la vie et proposant une vision optimiste de l'avenir* ». Il s'agit de « *tirer le meilleur profit des nouvelles technologies et de prévenir les risques potentiels pour l'humanité du développement de l'intelligence artificielle* ». Il soutient ainsi en 2016 et 2017, suite à un appel à projets lancé en 2015, pas moins de 37 projets de recherche destinés à prévenir les risques liés à l'intelligence artificielle<sup>2</sup>. Selon Max Tegmark, président du FLI, il existerait une « *course entre le pouvoir grandissant de la technologie et le bon sens avec lequel on la gère* » : alors que « *jusqu'ici, tous les investissements ont eu pour objectif de rendre les systèmes plus intelligents, c'est la première fois qu'il y a un investissement sur l'autre aspect* ».

Lors d'un colloque à New York sur les défis posés par l'émergence de l'intelligence artificielle, organisé le 14 octobre 2015 par l'Institut de recherche sur la criminalité et la justice des Nations Unies (UNICRI), Max Tegmark était invité avec un autre expert<sup>3</sup> à s'exprimer devant quelques 130 délégués de 65 États. Ils ont clairement souligné les **risques liés à l'intelligence artificielle** et appelé à la mise en place d'une **réflexion solide sur l'éthique de l'intelligence artificielle**.

Le second expert, Nick Bostrom, philosophe, fondateur du **Future of Humanity Institute** (FHI) en 2005 au sein de l'Université d'Oxford (unité de l'*Oxford Martin School*), rencontré par vos rapporteurs, a également fondé, dès 2004, un **Institute for Ethics and Emerging Technologies** (IEET), proche du mouvement transhumaniste.

De manière similaire au Future of Humanity Institute, ont été créées plusieurs structures qui travaillent toutes en réseau les unes avec les autres,

<sup>1</sup> L'Institut a été fondé en mars 2014 par Max Tegmark cosmologiste au MIT, Jaan Tallinn co-fondateur de Skype, Anthony Aguirre physicien à l'UCSC et deux étudiants (Viktoriya Krakovna et Meia Chita-Tegmark), figurent à son conseil consultatif l'informaticien Stuart J. Russell, le biologiste George Church, le physicien Frank Wilczek, les cosmologistes Stephen Hawking et Saul Perlmutter, ou, encore, l'entrepreneur Elon Musk.

<sup>2</sup> Il s'agit, par exemple, de développer une intelligence artificielle capable d'expliquer ses décisions ou, encore, de travailler sur l'alignement de l'intelligence artificielle sur les valeurs humaines.

<sup>3</sup> Outre Max Tegmark, président du FLI, dont il a déjà été question, le second expert était Nick Bostrom, philosophe, fondateur du Future of Humanity Institute (FHI) de l'université d'Oxford.

---

au sein de l'Université de Cambridge, un *Centre for the Study of Existential Risks* (CSER) créé en 2012 et un *Leverhulme Centre for the Future of Intelligence* créé en 2016 (tous les deux visités par vos rapporteurs), au sein de l'Université de Berkeley, un *Machine Intelligence Research Institute* (MIRI), créé lui aussi en 2016 et animé par Stuart Russel<sup>1</sup>, rencontré par vos rapporteurs.

Il peut être relevé que les anciens dirigeants de Paypal, Elon Musk (actuellement patron de Tesla et SpaceX) et Sam Altman se sont fixés pour but de promouvoir et de développer des outils d'intelligence artificielle en *open-source*. Ils ont ainsi fondé, le 11 décembre 2015, la « **fondation OpenAI** », qu'ils président, association à but non lucratif visant à réfléchir aux questions de société que pose l'intelligence artificielle. Vos rapporteurs ont eu la chance de visiter cette association basée dans la Silicon Valley et d'en rencontrer des responsables.

Le dernier exemple, peut être le plus significatif est le « *Partnership on AI* » formé en septembre 2016 par Google, Microsoft, Facebook, IBM et Amazon afin de réfléchir et de faire avancer de manière collective les discussions sur l'intelligence artificielle. Yann LeCun et Demis Hassabis ont joué un rôle essentiel dans ce partenariat. Vos rapporteurs se sont réjouis du fait qu'**Apple a rejoint cette initiative le 26 janvier 2017 le jour de leur visite du siège de l'entreprise**<sup>2</sup>. La responsable des affaires publiques d'Apple a alors expliqué à vos rapporteurs **que les valeurs spécifiques à Apple, en particulier la protection des données personnelles et de la vie privée, pourraient être mieux prises en compte**.

Apple a donc rejoint officiellement, et en tant que membre fondateur, avec Facebook, Google, Microsoft, IBM et Amazon, les activités du collectif voué au développement éthique de l'intelligence artificielle, ce « *Partnership on AI* ». Les six « GAFAMI » se sont donc dotés d'un outil commun, réputé être **ouvert aux entreprises, aux chercheurs et à toute personne morale intéressée par la démarche**. Le « *Partnership on AI* » a aussi intégré dans son conseil d'administration **six nouveaux membres provenant d'ONG et d'universités**. De cette manière, son conseil d'administration se compose désormais de douze membres, six responsables d'entreprises des nouvelles technologies (GAFAMI) et de six autres appartenant aux ONG.

---

<sup>1</sup> Stuart Russell travaille à l'université de Berkeley, où il est professeur au département d'informatique et directeur du centre pour l'étude des systèmes intelligents. Ancien membre du bureau exécutif de l'AAAI (American Association for Artificial Intelligence), il a reçu de nombreux prix scientifiques. Il est l'auteur de plus de cent articles et de plusieurs best-sellers sur l'intelligence artificielle. Il a co-rédigé le principal manuel disponible sur l'IA avec Perer Norvig, ancien professeur à l'université de Californie du Sud, directeur scientifique chez Google, qui auparavant travaillé pour la NASA sur l'intelligence artificielle et la robotique, ainsi que pour Junglee sur l'extraction d'informations par Internet., membre de l'AAAI et de l'ACM (Association for Computing Machinery).

<sup>2</sup> Lors de la visite de vos rapporteurs au siège appelé « campus Apple », situé en plein centre de la Silicon Valley au 1 InfiniteLoop à Cupertino en Californie, un communiqué de presse leur a été communiqué par des responsables de l'entreprise. Il expliquait qu'Apple rejoint ce partenariat en tant que membre fondateur (« founding partner »). Le communiqué de presse se trouve ici : <https://www.partnershiponai.org/2017/01/partnership-ai-update/>

---

Ces six représentants d'ONG et d'universités sont : Dario Amodei d'OpenAI, une organisation fondée par Elon Musk, Deirdre Mulligan de l'Université de Californie Berkeley, Jason Furman du *Peterson Institute of International Economics*, un prestigieux think tank de Washington, Subbarao Kambhampati de l'*Association for the Advancement of Artificial Intelligence*, Carol Rose de l'Union Américaine pour les Libertés Civiles (ACLU) et Eric Sears de la fondation MacArthur. Le communiqué de presse de « Partnership on AI » affirme : « *C'est un moment important pour Partnership on AI, alors que nous établissons un conseil d'administration diversifié et équilibré qui **étendra et élargira notre leadership**. L'inclusion de perspectives différentes et d'une réflexion critique constante a été une mission centrale depuis le début, et nous continuerons à ajouter de nouvelles voix à mesure que nous avançons* ». D'après le collectif, le conseil d'administration contrôlera les activités générales de l'organisation avec un directeur exécutif (encore inconnu), avec un Comité exécutif de direction (inconnu lui aussi) qui jugera et développera des initiatives selon les objectifs de l'organisation. Ces initiatives concerneront notamment la résolution « *des problématiques importantes, [...] sur l'éthique, la sécurité, la transparence, la vie privée, l'influence, et l'équité* » engendrées par les intelligences artificielles du futur. La première réunion du conseil a eu lieu le vendredi 3 février à San Francisco. L'organisation promet de révéler plus de détails sur son programme de recherche et d'activités.

Le rapport « *L'Intelligence artificielle et la vie en 2030* » (*Artificial Intelligence and Life in 2030*<sup>1</sup> en anglais) publié en septembre 2016 par l'Université Stanford, visitée par vos rapporteurs, dévoile les résultats de l'étude « *One Hundred Year Study of Artificial Intelligence* », un projet universitaire débuté en 2014 et initié par Eric Horvitz, chercheur au laboratoire Microsoft Research. Le rapport a été réalisé par les membres du Comité permanent de l'étude, présidé par Barbara J. Grosz et composé de six autres membres, dont Eric Horvitz et Russ Altman, et par un groupe d'étude rassemblant dix-sept chercheurs spécialisés dans le sujet de l'intelligence artificielle.

L'étude et les résultats présentés au sein de ce rapport portent sur **l'analyse, à long-terme, des avancées techniques de l'intelligence artificielle et de ses implications sur la société** au cours des cent dernières années. Ce projet reprend les travaux d'une autre étude menée entre 2008 et 2009, informellement connue sous le titre de « *Étude d'Asilomar de l'Association pour les progrès de l'intelligence artificielle* » (AAAI Asilomar Study, en anglais). Au cours de cette étude, un groupe d'experts de l'intelligence artificielle appartenant à des établissements, institutions et champs disciplinaires différents ainsi que des spécialistes des sciences cognitives, des juristes et des philosophes, avaient été mobilisés par Éric Horvitz, qui était à cette époque le président de l'AAAI.

---

<sup>1</sup> L'intégralité du rapport publié par l'Université Stanford est disponible ici : [https://ai100.stanford.edu/sites/default/files/ai\\_100\\_report\\_0831fnl.pdf](https://ai100.stanford.edu/sites/default/files/ai_100_report_0831fnl.pdf)



---

Selon les auteurs de ce rapport, **les technologies issues de l'intelligence artificielle, à l'instar de la reconnaissance vocale ou de l'utilisation de systèmes de recommandations automatiques, ont déjà et continueront de bousculer des pans entiers de l'économie mondiale et des sociétés.** L'intelligence artificielle représente en cela un enjeu considérable, tant du point de vue économique que politique et sociétal, et un facteur de création de valeur pour nos sociétés. **Il est, de fait, essentiel que les pouvoirs publics et les citoyens en saisissent les enjeux,** dont la complexité interdit tout angélisme ou catastrophisme *a priori*. Les gouvernants sont ainsi encouragés à produire une **législation encourageant l'innovation, la production et le transfert d'expertise,** et promouvant les responsabilités que doivent endosser le monde de l'entreprise et la société civile afin d'affronter les défis apportés par l'utilisation de ces technologies, notamment sur la question de la répartition des fruits de la croissance numérique.

Les auteurs de ce rapport soumettent de nombreuses recommandations à l'attention des décideurs, couvrant un large spectre de domaines. Des propositions de politiques publiques sont ainsi proposées en matière de protection de la vie privée, d'innovation, de responsabilité civile et pénale, ou encore de fiscalité. Trois recommandations majeures peuvent être retenues. Tout d'abord, les auteurs recommandent **d'accroître le niveau d'expertise des gouvernants** en matière d'intelligence artificielle, afin que cela leur permette de mieux apprécier les impacts de ces technologies. Ensuite, ils incitent à **l'élimination des obstacles et des freins à la transparence et aux recherches** sur la sécurité, la protection de la vie privée et les répercussions sociales entraînées par l'utilisation de technologies douées d'intelligence artificielle, afin de prévenir les usages abusifs. Enfin, les auteurs de ce rapport appellent au **financement d'études d'impact pluridisciplinaires.**

**Suite à ce rapport de septembre 2016** et puisqu'il n'existait **aucun guide commun encadrant le domaine de l'intelligence artificielle,** notamment en édictant de bonnes pratiques en la matière, plusieurs spécialistes de **l'intelligence artificielle et de la robotique se sont réunis lors de la conférence dénommée « Beneficial AI 2017 » organisée par le Future of Life Institute** (le FLI, dont il a été question plus haut). La conférence s'est tenue à Asilomar, en Californie du 5 au 8 janvier 2017<sup>1</sup>, avec le soutien de sponsors tels que Alexander Tamas, Elon Musk, Jaan Tallinn et deux associations « *The Center for Brains, Minds, and Machines* » et « *The Open Philanthropy Project* ».

Aux termes de la rencontre, les spécialistes ont procédé à l'adoption de vingt-trois principes baptisés « **Les 23 principes d'Asilomar** » et dont l'objectif est d'encadrer le développement de l'intelligence artificielle. D'après les informations recueillies, les principes ont été signés par 846 chercheurs spécialisés dans l'intelligence artificielle et la robotique et par 1 270 autres spécialistes dans divers domaines. Ces « **23 principes d'Asilomar** »

---

<sup>1</sup> Son programme figure ici : <https://futureoflife.org/bai-2017/>



---

représentent un guide de référence d'encadrement éthique du développement de l'intelligence artificielle, qui explique que « *l'intelligence artificielle a déjà fourni de nombreux outils utiles qui sont utilisés au quotidien à travers le monde. Son développement continu, guidé par les principes suivants, offrira des opportunités extraordinaires pour aider, responsabiliser et rendre plus performants les humains pour les décennies et les siècles à venir* ». Les « 23 principes d'Asilomar »<sup>1</sup> se présentent comme suit :

1) **Objectif des recherches** : Le développement de l'intelligence artificielle ne doit pas servir à créer une intelligence sans contrôle mais une intelligence bénéfique.

2) **Investissements** : Les investissements dans l'intelligence artificielle doivent être orientés vers le financement de recherches visant à s'assurer de son usage bénéfique, qui prend en compte des questions épineuses en matière d'informatique, d'économie, de loi, d'éthique et de sciences sociales. Parmi ces questions :

- « Comment rendre les futures intelligence artificielle suffisamment solides pour qu'elles fassent ce qu'on leur demande sans dysfonctionnement ou risque d'être piratées ? »
- « Comment améliorer notre prospérité grâce à cette automatisation tout en maintenant les effectifs humains ? »
- « Comment adapter le cadre juridique afin d'être plus juste et efficace, de suivre le rythme de l'intelligence artificielle et de gérer les risques qui y sont associés ?
- « Quel ensemble de valeurs l'intelligence artificielle devra respecter, et quel statut éthique devrait-elle revêtir ?

3) **Relations entre les scientifiques et les législateurs** : Un échange constructif et sain entre les développeurs d'intelligence artificielle et les législateurs est souhaitable.

4) **Esprit de la recherche** : Un esprit de coopération, de confiance et de transparence devrait être entretenu entre les chercheurs et les scientifiques en charge de l'intelligence artificielle.

5) **Éviter une course** : Les équipes qui travaillent sur les intelligences artificielles sont encouragées à coopérer pour éviter des raccourcis en matière de standards de sécurité.

6) **Sécurité** : Les intelligences artificielles devraient être sécurisées tout au long de leur existence, une caractéristique vérifiable et applicable.

7) **Transparence en cas de problème** : Dans le cas d'une blessure provoquée par une intelligence artificielle, il est nécessaire d'en trouver la cause.

---

<sup>1</sup> <https://futureoflife.org/ai-principles/>

---

8) **Transparence judiciaire** : Toute implication d'un système autonome dans une décision judiciaire devrait être accompagnée d'une explication satisfaisante contrôlable par un humain.

9) **Responsabilité** : Les concepteurs et les constructeurs d'intelligence artificielle avancée sont les premiers concernés par les conséquences morales de son utilisations et de ses détournements. Il leur incombe donc d'assumer la charge de les anticiper.

10) **Concordance de valeurs** : Les intelligences artificielles autonomes devraient être conçues de façon à ce que leurs objectifs, leur comportement et leurs actions s'avèrent conformes aux valeurs humaines.

11) **Valeurs humaines** : Les intelligences artificielles doivent être conçues et fonctionner en accord avec les idéaux de la dignité, des droits et des libertés de l'homme, ainsi que de la diversité culturelle.

12) **Données personnelles** : Chacun devrait avoir le droit d'accéder et de gérer les données le concernant au vu de la capacité des intelligences artificielles à analyser et utiliser ces données.

13) **Liberté et vie privée** : L'utilisation d'intelligence artificielle en matière de données personnelles ne doit pas rogner sur les libertés réelles ou perçue des citoyens.

14) **Bénéfice collectif** : Les intelligences artificielles devraient bénéficier au plus grand nombre, les valoriser et les rendre plus performants.

15) **Prospérité partagée** : La prospérité économique découlant de l'utilisation de systèmes d'intelligence artificielle devrait être partagée avec le plus grand nombre, pour le bien de l'humanité.

16) **Contrôle humain** : Les humains devraient pouvoir choisir comment et s'ils veulent déléguer des décisions de leur choix aux intelligences artificielles.

17) **Anti-renversement** : Le pouvoir obtenu en contrôlant des intelligences artificielles très avancées devrait être soumis au respect et à l'amélioration des processus civiques dont dépend le bien-être de la société plutôt qu'à leur détournement à d'autres fins.

18) **Course aux IA d'armement** : Une course aux intelligences artificielles d'armement mortelles est à éviter.

19) **Avertissement sur les capacités** : En l'absence de consensus sur le sujet, il est recommandé d'éviter les hypothèses au sujet des capacités maximum des futures intelligences artificielles.

20) **Importance** : Les intelligences artificielles avancées pourraient entraîner un changement drastique dans l'histoire de la vie sur Terre, et devront donc être gérées avec un soin et des moyens considérables.

21) **Risques** : Les risques causés par les IA, particulièrement les risques catastrophiques ou existentiels, sont sujets à des efforts de préparation et d'atténuation adaptés à leur impact supposé.

---

22) **Auto-développement infini** : Les IA conçues pour s'auto-développer à l'infini ou s'auto-reproduire, au risque de devenir très nombreuses ou très avancées rapidement, doivent faire l'objet d'un contrôle de sécurité rigoureux.

23) **Bien commun** : Les intelligences surdéveloppées devraient seulement être développées pour contribuer à des idéaux éthiques partagés par le plus grand nombre et pour le bien de l'humanité plutôt que pour un État ou une entreprise.

Il convient, en outre, d'observer que **Microsoft** a lancé en 2017 un **fonds d'investissement en capital-risque dédié à l'intelligence artificielle avec un objectif de ciblage sur les investissements à impact positif pour la société**. D'après le vice-président de Microsoft Ventures, Nagraj Kashyap, « *l'intelligence artificielle doit être conçue pour assister l'humanité, être transparente, maximiser l'efficacité sans détruire la dignité humaine, protéger intelligemment la vie privée et assurer la responsabilité de l'imprévu, et se garder des préjugés. Ce sont ces principes qui guideront l'évolution de ce fonds* ». Le premier investissement de Microsoft est destiné à la plateforme « *Element AI* » basée à Montréal, cofondée avec Yoshua Bengio, professeur à l'université de Montréal. Cet investissement confirme à vos rapporteurs le fait que la capitale du Québec se place parmi les principaux pôles mondiaux de l'intelligence artificielle, alors que Google y avait également basé une de ses divisions de recherche en intelligence artificielle.

De même que Microsoft, avec Pierre Omidyar et Reid Hoffman, les fondateurs d'eBay et de LinkedIn, ont lancé en 2017, avec le fondateur de Raptor Group James Pallotta et les fondations Knight et William et Flora Hewlett, un fonds d'investissement de 27 millions de dollars, qui porte le nom de *Ethics and Governance of Artificial Intelligence Fund*, qui accompagnera les projets R&D axés sur les problématiques d'éthique dans le domaine de l'intelligence artificielle. Le fonds est piloté par le Media Lab du MIT et le Berkham Klein Center for Internet and Society de Harvard. Cette enveloppe sera exploitée pour soutenir les initiatives repérées dans le monde académique et orientées vers le développement d'une intelligence artificielle éthique, capable de ne pas « reproduire et amplifier les biais humains. Dans cette optique, il s'agira d'impliquer, au-delà des ingénieurs, des sociologues, des philosophes, des juristes, des économistes... et les régulateurs, au croisement des sciences informatiques, humaines et sociales. L'*Ethics and Governance of Artificial Intelligence Fund* se donne aussi pour mission de vulgariser l'intelligence artificielle auprès du grand public.

Vos rapporteurs s'interrogent sur les objectifs précis des GAFAMI et d'Elon Musk à travers ces nombreuses initiatives. La volonté de ces nouveaux géants pourrait-elle être celle de **se dédouaner** ou de **créer un nuage de fumée pour ne pas parler des vrais problèmes éthiques** posés à court terme par les technologies d'intelligence artificielle, telles que **l'usage des données ou le respect de la vie privée** ? Vos rapporteurs n'ont pas tranché et laissent aux auteurs de ces initiatives le bénéfice du doute. Chris Olah, l'un des auteurs de « *Concrete Problems in AI Safety* », travail de

---

recherche publié par Google fait valoir qu' « *alors que les risques potentiels de l'IA ont reçu une large attention de la part du public, les discussions autour de ce sujet sont restées très théoriques et basées sur des spéculations* » et qu'il faut développer « *des approches pratiques d'ingénierie de systèmes d'intelligence artificielle opérant de façon sûre et fiable* ». Dépassant les débats théoriques entre pro et anti-IA, ces initiatives permettent d'aborder avec des experts la réalité concrète de l'intelligence artificielle, même si elles donnent, selon vos rapporteurs, une **place trop grande au risque de l'émergence d'une IA forte qui dominerait et pourrait faire s'éteindre l'espèce humaine.**

### **3. Le travail en cours sur les enjeux éthiques au sein de l'association mondiale des ingénieurs électriciens et électroniciens (Institute of Electrical and Electronics Engineers ou IEEE)**

Vos rapporteurs soulignent l'important travail en cours sur les enjeux éthiques actuellement au sein de l'association mondiale des ingénieurs électriciens et électroniciens (*Institute of Electrical and Electronics Engineers* ou IEEE), qui regroupe plus de 400 000 membres.

L'initiative mondiale de l'IEEE pour « les considérations éthiques dans l'Intelligence Artificielle et les Systèmes Autonomes » a en effet pour principal objectif de **proposer un cadre éthique de référence pour le développement des systèmes d'intelligence artificielle et des systèmes autonomes.**

Souhaitant dépasser la recherche de la performance technologique en soi, ou le succès commercial, l'IEEE **visé à ce que ces systèmes se comportent d'une manière bénéfique pour l'humanité** et que leur développement **contribue au bien-être** de celle-ci. L'approche est, d'une part, de produire un document évolutif rédigé collectivement et, d'autre part, de proposer des standards qui pourraient devenir des standards industriels. La **première version du document** d'IEEE « *Conception conforme à l'éthique : une vision pour fixer comme priorité le bien-être humain avec l'intelligence artificielle et les systèmes autonomes* » a été publiée le **13 décembre 2016**, avec l'idée d'une **discussion d'ici l'été 2017** et la diffusion d'une **deuxième version consolidée prévue à l'automne 2017.**

Plusieurs événements, réunions et téléconférences sont programmés afin de discuter du texte, y compris en Europe, par exemple à Bruxelles le 11 avril 2017, et surtout un événement marquant, qui se tiendra sous la forme d'une plénière les 5 et 6 juin à Austin (Texas), à l'issue duquel sera réalisée la deuxième version du texte. C'est la communication de ce texte qui est prévue à l'automne 2017. Enfin, une troisième plénière aura lieu en Asie en 2018 (Japon ou Chine).

Le document publié le 13 décembre 2016 aborde les **différents aspects de la création et du développement d'algorithmes et de l'intelligence artificielle concernés par des questionnements éthiques et**

---

propose des **recommandations** pour chacun d'entre eux. Il en formule huit, que vos rapporteurs récapitulent ici.

### **1. Les principes généraux de la recherche en intelligence artificielle**

Le développement de l'intelligence artificielle doit être encadré par un respect des principes fondamentaux des droits humains, de responsabilité, de transparence, d'éducation et de connaissance.

### **2. Les valeurs « programmées » dans les systèmes autonomes**

Les valeurs morales à intégrer aux algorithmes des systèmes autonomes ne peuvent être universelles, et, sans tomber dans le relativisme, doivent davantage s'adapter aux communautés d'utilisateurs concernées et aux tâches qui lui sont confiées. Il est important de veiller, dès la conception des algorithmes, à ce que la multiplicité de valeurs ne les fasse pas entrer en conflit les unes avec les autres et ne désavantage aucun groupe d'utilisateurs. Cela implique donc qu'une architecture de calcul exigeante des valeurs et normes éthiques doit être respectée.

### **3. La méthodologie de recherche et de conception éthiques**

Il est essentiel que la méthodologie de recherche et de conception d'algorithmes et de systèmes autonomes comble de nombreux manques. Au-delà de son enseignement actuellement absent des programmes d'études en ingénierie, l'éthique doit être intégrée dans de nombreux domaines. Les pratiques industrielles doivent être davantage marquées par une culture éthique et la communauté « tech » doit s'emparer des sujets éthiques et assumer sa responsabilité éthique. De même, du fait du mode de fonctionnement et de prise de décision des algorithmes, il est nécessaire d'inclure des composants de type « boîtes noires » afin d'enregistrer les informations aidant à l'analyse des processus de décision et d'action des systèmes autonomes.

### **4. La sécurité**

Les comportements imprévus ou involontaires de systèmes d'intelligence artificielle représentent potentiellement un danger grandissant. Il est de fait essentiel de renforcer la sécurité de l'utilisation de systèmes d'intelligence qui, en devenant de plus en plus capables, peuvent devenir dangereux. Les chercheurs et concepteurs de systèmes de plus en plus autonomes devront se confronter à un ensemble complexe de défis de sécurité sur le plan technologique ainsi que sur le plan éthique.

### **5. La protection des données à caractère personnel**

L'un des principaux dilemmes éthiques concernant le développement de l'intelligence artificielle concerne l'asymétrie de données (*data asymmetry*), entre ceux qui les produisent et ceux qui les agrègent, les traitent, les manipulent et les vendent. La protection des données à caractère personnel doit de fait être organisée en considération de différents facteurs : comment est défini et est identifié le caractère « personnel » d'une donnée ;

---

comment définir le consentement d'accès à des données à caractère personnel ; les conditions d'accès et de traitement de ces données ; etc.

### **6. Les considérations juridiques**

L'utilisation de systèmes autonomes soulève de nombreuses questions d'un point de vue juridique. Des exigences de responsabilisation, de transparence et de vérifiabilité des actions des robots sont essentielles et les dispositifs existants doivent être améliorés. À titre d'exemple, la transparence des systèmes autonomes permet d'assurer qu'une intelligence artificielle respecte les droits individuels et, utilisée par une administration, qu'elle ne porte pas atteinte aux droits des citoyens et peut recueillir leur confiance. En outre, il est nécessaire d'adapter le cadre juridique concernant la responsabilité des préjudices et dommages causés par un système autonome, ainsi que concernant l'intégrité et la protection des données à caractère personnel.

### **7. La défense et les « robots tueurs »**

L'utilisation d'armes létales autonomes, également appelées « robots tueurs » revêt un caractère risqué, en cela que leurs actions peuvent être altérées et devenir un danger non-maîtrisable, en cela que la surveillance humaine en est exclue. Ces « robots tueurs », à l'instar des drones militaires, sont en outre critiqués, et la légitimation de leur développement pourrait potentiellement créer des précédents géopolitiquement dangereux à moyen terme, notamment en termes de prolifération de ces armes, d'abus d'utilisation et d'escalade rapide des conflits. En outre, l'absence de standards de conception ne permet pas d'adopter des règles éthiques clairement définies.

### **8. Les problèmes économiques et humanitaires**

L'objectif de ce rapport sur le versant économique et social est d'identifier les principaux moteurs de l'écosystème mondial de la technologie humaine et de prendre en compte les ramifications économiques et humanitaires, afin de suggérer des opportunités-clés de solutions qui pourraient être mises en œuvre en débloquent les points critiques de tension. Les systèmes autonomes et, plus largement, l'intelligence artificielle, souffrent d'une mauvaise image auprès du grand public, du fait d'une interprétation erronée par de nombreuses œuvres de culture populaire alarmistes sur les capacités d'un système d'intelligence artificielle « forte ».

Le phénomène de robotisation et de développement de l'intelligence artificielle n'est généralement pas perçu uniquement dans un contexte de marché. Si toute politique publique sur l'intelligence artificielle peut potentiellement ralentir l'innovation, il demeure que de nombreuses craintes, notamment concernant l'emploi, émergent, en cela que les changements technologiques se produisent trop rapidement pour les méthodes existantes de formation de la main-d'œuvre. En outre, l'accès aux technologies d'intelligence artificielle n'est pas équitablement réparti, entraînant un manque de compréhension des informations par une certaine partie de la

---

population ; en cela, l'avènement de l'intelligence artificielle et des systèmes autonomes peut exacerber les différences économiques et structurelles entre les pays développés et les pays en développement.

En outre, l'IEEE élabore des **standards**, dans une démarche parallèle à la discussion du document « Conception conforme à l'éthique : une vision pour fixer comme priorité le bien-être humain avec l'intelligence artificielle et les systèmes autonomes ». **Trois propositions de standards industriels** ont ainsi été proposées et d'autres sont à venir. Ils sont en cours de rédaction dans des groupes de travail ouverts, au sein, de la IEEE Standard association :

- Standard pour un processus tenant compte des considérations éthiques dans la conception des systèmes

- Standard sur la prise en compte de la protection de la vie privée dans les systèmes et logiciels utilisant des données personnelles.

- Standard sur les niveaux de transparence mesurables pour le test de systèmes et l'évaluation de leur niveau de conformité.

Une telle rédaction de standards prend au moins un an et d'autres idées de standard pourraient être proposées et soumises à la discussion.

Selon l'animateur du comité d'IEEE ayant produit le document « *Conception conforme à l'éthique : une vision pour fixer comme priorité le bien-être humain avec l'intelligence artificielle et les systèmes autonomes* » et qui prépare le second rapport, Raja Chatila, directeur de l'Institut des systèmes intelligents et de robotique (ISIR), auditionné à plusieurs reprises par vos rapporteurs, deux aspects concernant l'éthique devraient être particulièrement abordés : d'une part, les **methodologies de conception éthique de systèmes autonomes**, de manière à ce que ceux-ci tiennent compte des valeurs éthiques humaines (par exemple respect de la vie humaine, des droits humains) et de manière à ce que les algorithmes qui les régissent soient transparents, explicables, traçables, et, d'autre part, **l'éthique des machines, c'est à dire comment les décisions prises par une machine peuvent intégrer un raisonnement éthique.**

#### **4. Une sensibilisation insuffisante du grand public à ces questions et un besoin de partage en temps réel de la culture scientifique et de ses enjeux éthiques**

Vos rapporteurs constatent **une sensibilisation insuffisante du grand public aux questions posées par l'intelligence artificielle et les systèmes autonomes.** Les **traitements médiatiques** de ces questions restent le plus souvent **sensationnalistes, voire alarmistes**, alors qu'une information objective serait souhaitable.

La **vision déjà tronquée du grand public**, sous l'effet des œuvres de fiction, et en particulier du **cinéma**, n'est pas améliorée par la lecture de la



---

plupart des articles disponibles sur l'intelligence artificielle dans nos journaux et magazines.

Vos rapporteurs veulent affirmer avec force le **besoin de partage en temps réel de la culture scientifique et de ses enjeux éthiques**.

Votre rapporteure, Dominique Gillot, **présidente du Conseil national de la culture scientifique, technique et industrielle (CNCSTI)**, entend rappeler que ce Conseil, placé auprès du ministre chargé de la Culture et du ministre chargé de la Recherche, « *participe à l'élaboration d'une politique nationale en matière de développement de la culture scientifique, technique et industrielle, en cohérence avec les grandes orientations de la stratégie nationale de recherche* ».

Réuni pour la première fois le 24 novembre 2015<sup>1</sup>, ses membres, réfléchissent aux **actions à conduire** et sur l'articulation entre le niveau national et le niveau régional. Parmi les thèmes prioritaires de réflexion et de travail figurent **l'utilisation des technologies et notamment du numérique dans la médiation scientifique**, les entreprises et l'innovation, les filles et la science, l'appui de la recherche aux décisions publiques, l'après COP 21. Votre rapporteure, Dominique Gillot, souligne **l'importance de la visibilité du débat public autour de ces questions**.

Les activités d'**éducation populaire ou les ateliers citoyens** sont **d'autres pistes pour sensibiliser le grand public aux questions posées par l'intelligence artificielle**. Vos rapporteurs ne se satisfont pas des traitements sensationnalistes ou alarmistes de ces questions par les médias. Les informations contenues dans le présent rapport devraient être accessibles à tous et transmises par le biais d'activités d'éducation populaire ou par des ateliers citoyens.

Vos rapporteurs observent, en outre, que non seulement le numérique doit être transformé par l'éthique, mais que **le numérique est, par lui-même, un facteur d'évolution des règles éthiques appliquées par les chercheurs**, surtout au cours des dernières années. Le contrôle par les pairs se fait de manière **plus décentralisée et collaborative**.

L'exemple de *pubpeer*, site de discussion en ligne d'articles scientifiques<sup>2</sup> peut être cité de manière significative. Les réseaux sociaux de chercheurs<sup>3</sup> sont une autre illustration.

---

<sup>1</sup> Cette réunion a été suivie par la remise du prix « Le goût des sciences », prix qui a pour objectif de valoriser le travail des chercheurs et des éditeurs, d'encourager les vocations scientifiques et d'affirmer l'importance de la culture scientifique au sein de la culture générale contemporaine.

<sup>2</sup> Cf. <https://pubpeer.com/>

<sup>3</sup> Cf. <https://www.researchgate.net>

---

### III. LES QUESTIONS TECHNOLOGIQUES ET SCIENTIFIQUES QUI SE POSENT EN MATIÈRE D'INTELLIGENCE ARTIFICIELLE

#### A. LES SUJETS D'INTERROGATION LIÉS AUX ALGORITHMES UTILISÉS PAR LES TECHNOLOGIES D'INTELLIGENCE ARTIFICIELLE

##### 1. Les questions de sécurité et de robustesse

Concernant la sécurité au sens de la **sécurité numérique**, l'OPECST a récemment rendu un rapport sur les risques et la sécurité numérique<sup>1</sup>, ce point fera donc l'objet de courts développements dans le présent rapport. Vos rapporteurs renvoient à ce rapport qui traite d'un point de vue général de la sécurité numérique.

Un exemple fameux de problématique de sécurité pour les systèmes d'intelligence artificielle est celui du risque de **piratage d'un drone ou d'une voiture autonome. Les cas existent et doivent donc être résolus.** La sécurité de ces systèmes emporte des conséquences en terme de vie humaine. Une piste peut résider dans le fait de ne pas être perpétuellement connecté afin de prévenir le risque de piratage. Selon John Krafcik, le président de Waymo, ex-Google *Self-Driving Car Project* (branche d'Alphabet pour la conduite autonome) une connexion Internet permanente embarquée n'est pas nécessaire pour un véhicule sans chauffeur et, au contraire, un accès Internet discontinu dans une voiture autonome serait un gage de sécurité. Afin d'assurer la sécurité des passagers d'une voiture autonome face au risque de piratage, il s'agit de **ne pas dépendre de la connectivité embarquée**, comme l'accès à Internet. John Krafcik explique ainsi : *« la cybersécurité est quelque chose que nous prenons très sérieusement (...) quand nous évoquons nos voitures autonomes, ce n'est pas seulement pour dire qu'il n'y a pas de conducteur humain, mais aussi qu'il n'y a pas de connexion cloud en continu depuis le véhicule ».*

Les connexions au réseau seront donc gérées avec parcimonie, lorsque la voiture en a besoin. Ce ne sera pas une connexion en continu, susceptible d'être piratée pour s'introduire dans le véhicule. Waymo met ainsi en circulation des véhicules autonomes Chrysler Pacifica à Mountain View (Californie), où se trouve le siège de Google, visité par vos rapporteurs, et à Phoenix (Arizona).

Au-delà du piratage, qui est intentionnel, se pose le problème de la **perte de contrôle des systèmes d'intelligence artificielle.** Des recherches sont à conduire dans ce sens, pour relever le défi de sécurité et de robustesse des technologies. Depuis 2016, Thierry Berthier, chaire de cyberdéfense de Saint-Cyr à l'université de Limoges, Jean-Gabriel Ganascia, UPMC-LIP6, et Olivier Kempf, IRIS, conduisent une étude sur un scénario concret **de crise militaire suite à une dérive de systèmes d'intelligence artificielle**, dans le cas d'une hypothèse « faible » de dérive malveillante dans le sens où

---

<sup>1</sup> Rapport « Sécurité numérique et risques : enjeux et chances pour les entreprises », n° 271 (2014-2015).

---

l'intelligence artificielle impliquée n'a **pas de volonté de nuisance** ni de « métacompréhension » de son environnement ou de sa propre activité. Ils ne font intervenir que des capacités et fonctionnalités de l'intelligence artificielle existantes ou en cours de développement, notamment dans les récents programmes initiés par la Darpa, l'agence américaine pour les projets de recherche avancée de défense, visitée par vos rapporteurs. La complexité des systèmes et des processus d'apprentissage pourrait conduire à des situations critiques. Le risque naîtrait ainsi de **l'association de choix humains et de mécanismes numériques**, une série d'éléments mis bout à bout pourraient en effet devenir potentiellement dangereux sans que chacun de ces mécanismes pris individuellement ne le soit. Un article devrait paraître en 2017 dans la revue de la Défense Nationale suite à cette étude.

En robotique il est **nécessaire de toujours pouvoir arrêter un système, la question peut se poser pour une intelligence artificielle**, qu'il s'agisse d'un système informatique ou de son incarnation dans un robot. La réversibilité du fonctionnement de l'intelligence artificielle est essentielle, elle évoque un peu le Golem de Prague qui se tourne vers son maître, le prophète Jérémie, et lui dit : « *défais moi !* ». Vos rapporteurs rappellent qu'en 2016, **Google a également posé la question** du manque de contrôle potentiel d'agents apprenants qui pourraient apprendre à empêcher leur interruption dans une tâche. C'est dans ce sens que **la firme développe l'idée d'un « bouton rouge » permettant la désactivation des intelligences artificielles**<sup>1</sup>.

Pour la CERNA, **la question du débrayage de certaines fonctions autonomes**, voire de la **mise hors service** du robot par l'utilisateur, **est centrale**. Elle se demande ainsi : « *quand et comment l'utilisateur peut-il éteindre des fonctions du robot, voire le robot lui-même ? Le robot peut-il ou doit-il empêcher ces extinctions, dans quelles circonstances et sur quelles bases objectives ?* ».

Des recherches complémentaires sont nécessaires sur ce sujet. Pour paraphraser Raymond Aron<sup>2</sup>, **l'enjeu est donc, face à une paix improbable avec les machines, de rendre la guerre impossible**.

## **2. Les biais et les problèmes posés par les données nécessaires aux algorithmes d'apprentissage automatique**

Les biais sont **l'un des plus gros problèmes posés par les algorithmes d'apprentissage automatique**, ou pour être plus rigoureux, posés **par les données nécessaires aux algorithmes**. La question concerne en effet plus les données que les algorithmes eux-mêmes. Les impacts se font ressentir après le traitement, mais **les biais sont introduits en amont dès le stade des jeux de données**.

---

<sup>1</sup> Morgane Tual, « Pourquoi Google a conçu un bouton rouge pour désactiver des intelligences artificielles », *Le Monde*, 7 juin 2016.

<sup>2</sup> Il utilisait l'expression de « Paix impossible, guerre improbable ».

---

En effet, les algorithmes d'apprentissage automatique et en particulier d'apprentissage profond vont **reproduire les biais des données qu'ils traitent**, en particulier toutes les discriminations connues dans nos sociétés. Les données peuvent inclure toute sorte de biais. Selon vos rapporteurs, cette difficulté ne doit jamais être négligée.

Outre relever le défi de l'apprentissage non-supervisé, il convient donc d'**être vigilant sur ces biais**, qui de surcroît sont **souvent invisibles** sauf si des efforts de recherche sont entrepris, ainsi que l'ont expliqué plusieurs spécialistes à vos rapporteurs. Les algorithmes ne détectent pas les biais, ils sont bêtes, comme a pu le dire un chercheur.

Le second rapport de la CERNA, en cours de publication, traite de ce point.

Ces biais peuvent aussi être mis en relation avec les problèmes de **loyauté des plateformes**, ainsi que de **l'utilisation des systèmes d'intelligence artificielle à des fins non éthiques, partisans et sans diversification des choix**, comme l'affirme votre rapporteure Dominique Gillot.

Vos rapporteurs jugent que la **gouvernance des algorithmes et des prédictions qu'ils opèrent est nécessaire**.

### **3. Le phénomène de « boîtes noires » des algorithmes de *deep learning* appelle un effort de recherche fondamentale vers leur transparence**

Les **connaissances existantes sur les systèmes d'intelligence artificielle** montrent que nous ne disposons d'**aucune explication théorique satisfaisante des raisons pour lesquelles les algorithmes de *deep learning***, à savoir des réseaux de neurones en couches multiples, fonctionnent, ou, pour être plus précis, donnent, dans un certain nombre de domaines, d'excellents résultats.

**Ce problème d'opacité reste entièrement à résoudre**. On parle ici de phénomènes de « boîtes noires », mais elles n'ont rien à voir avec les boîtes noires des avions, qui sont des enregistreurs numériques.

Le défi à relever est donc celui de l'objectif d'**explicabilité des algorithmes de *deep learning***. Il s'agit là aussi d'une autre question qui peut se rattacher à la question générale de la gouvernance des algorithmes.

Comme il a été vu, le Conseil général de l'économie (CGE) a rendu le 15/12/2016 un rapport au ministre de l'économie et des finances portant sur les « Modalités de régulation des algorithmes de traitement des contenus »<sup>1</sup>. Bien que les algorithmes de traitement des contenus sont inséparables des données qu'ils traitent et des plateformes qui les utilisent pour proposer un service et alors qu'il y a de **nombreux travaux sur la protection des données**

---

<sup>1</sup> <http://www.economie.gouv.fr/cge/modalites-regulation-des-algorithmes-traitement-des-contenus>

---

**et sur la loyauté des plateformes, il y en a encore peu sur les algorithmes eux-mêmes.** Ce rapport souligne donc qu'il faut **développer la capacité à tester et contrôler les algorithmes**<sup>1</sup>.

Vos rapporteurs rappellent à nouveau ici que l'INRIA avec sa **plateforme « Transalgo »**, placée sous la direction de Nozha Boujemaa, développe en 2017 de manière utile une **plateforme scientifique d'évaluation de la responsabilité et de la transparence des algorithmes**<sup>2</sup>, afin de répondre aux préoccupations exprimées d'**explicabilité des algorithmes**. Une telle démarche, en lien avec le *Data Transparency Lab* du MIT va dans le bon sens mais gagnerait à voir sa force de frappe être **démultipliée par la mobilisation de plusieurs équipes de recherche**. L'INRIA ne peut rester la seule structure en France à conduire un tel projet de ce type. Le CNNum, la CERNA et la CNIL sont associés à la plateforme « Transalgo » mais il s'agit plus de partager des expériences que de conduire des recherches. L'INRIA explique ainsi que *« le CNNum joindra ses forces avec celles d'Inria dans TransAlgo dans le respect des missions de chacun. Il prendra en charge le recensement et l'objectivisation de la situation actuelle de certaines pratiques des plateformes à travers un dispositif contributif (citoyens et professionnels). Les données des différentes sources de régulation européennes ou internationales viendront enrichir également le centre de ressources. Pour faire remonter des cas d'usages bien réels, nous avons prévu de collaborer avec des think tank comme la FING (Fondation internet nouvelle génération), ou des associations de consommateurs comme Que-Choisir, en plus de la CERNA (Commission de réflexion sur l'éthique de la recherche en sciences et technologies du numérique d'Allistene). Nous allons travailler également à partir des remontées d'expression de besoins qui viendront du CNNum, de la Direction Générale de la Concurrence, de la Consommation et de la Répression des Fraudes (DGCCRF), de l'Autorité française de contrôle en matière de protection des données personnelles (CNIL), afin d'identifier les problèmes les plus observés par le citoyen, les industriels, les autorités de régulation »*.

#### **4. La question des bulles d'information dites « bulles de filtres »**

L'**information ciblée** tout comme la **publicité personnalisée** ou la **logique de construction des « fils d'actualité » des réseaux sociaux**, à l'instar de celui de Facebook, sont autant d'exemples de réalités déjà manifestes d'usage des systèmes d'intelligence artificielle, qui sont **de nature**

---

<sup>1</sup> Ses auteurs proposent cinq pistes d'action qui ont pour objet **la montée en compétence et le développement de l'expertise des pouvoirs publics**, mais aussi d'appeler au **développement de bonnes pratiques** dans les différents secteurs économiques. Ils soulignent aussi qu'il faut **préserver une image positive des technologies** utilisées pour concevoir ou opérer des algorithmes. C'est essentiel pour continuer à attirer les jeunes générations de françaises et de français dans des filières de formation exigeantes (mathématiques, ingénieurs ou data scientists) où la France est aujourd'hui bien placée.

<sup>2</sup> Cf. <https://www.inria.fr/actualite/actualites-inria/transalgo>



---

à **changer notre rapport au monde, aux autres et à la connaissance en orientant, voire en manipulant, notre perception de la réalité.**

La réflexion inquiète présentée dans le livre *Filter bubbles* par Eli Pariser, Président de MoveOn et cofondateur de Avaaz, illustre ce fait : **les algorithmes intelligents sélectionnent le contenu d'informations et créent par là des « bulles de filtres » qui se multiplient et transforment le rapport de l'individu au monde.**

Ce sujet mérite une **vigilance accrue des pouvoirs publics.** Pour vos rapporteurs, **l'enfermement, qu'il soit politique, idéologique ou cognitif, doit être combattu.**

La question va bien plus loin que les critiques formulées à l'encontre des **fausses informations**, ou *Fake news*, amplifiées par les réseaux sociaux. Sur ce dernier point, la recherche est assez bien avancée et comme l'a indiqué à plusieurs reprises Yann LeCun, directeur de la recherche en intelligence artificielle de Facebook, l'intelligence artificielle peut être utilisée pour limiter les flux de fausses informations. Des **outils de vérification** (*fact checking*) sont ainsi mis en place par plusieurs plateformes, à commencer par Facebook.

## **B. LES SUJETS D'INTERROGATION LIÉS À LA « SINGULARITÉ », À LA « CONVERGENCE NBIC » ET AU « TRANSHUMANISME »**

### **1. La « singularité », point de passage de l'IA faible à l'IA forte peut, à long terme, constituer un risque**

La rupture dite de la « **singularité technologique** » appelée aussi **simplement singularité**, est le nom que des écrivains et des chercheurs en intelligence artificielle ont donné au **passage de l'IA faible à l'IA forte**<sup>1</sup>. La **singularité représente un tournant hypothétique supérieur dans l'évolution technologique**, dont l'intelligence artificielle serait le ressort principal.

De nombreuses œuvres de science-fiction ont décrit ce tournant, qui a été une source d'inspiration très riche pour le cinéma. Les films

---

<sup>1</sup> Dès 1965, Irving John Good a décrit, en précurseur, la singularité, mais sans la nommer, et ce de la manière suivante : « Mettons qu'une machine supra-intelligente soit une machine capable dans tous les domaines d'activités intellectuelles de grandement surpasser un humain, aussi brillant soit-il. Comme la conception de telles machines est l'une de ces activités intellectuelles, une machine supra-intelligente pourrait concevoir des machines encore meilleures ; il y aurait alors sans conteste une explosion d'intelligence, et l'intelligence humaine serait très vite dépassée. Ainsi, l'invention de la première machine supra-intelligente est la dernière invention que l'homme ait besoin de réaliser ». Vernor Vinge a commencé à parler de la singularité dans les années 1980 et a formulé ses idées dans un premier article paru en 1993 intitulé « Technological Singularity ». Il y posait l'hypothèse que dans un délai de trente ans, l'humanité aurait les moyens de créer une intelligence surhumaine mettant un terme à l'ère humaine.

---

« Terminator », « Matrix », « Transcendance » sont des exemples de la « singularité technologique », qui est donc bien plus qu'une simple **hostilité de l'intelligence artificielle**, également souvent au cœur de l'intrigue des œuvres de science-fiction.

Les progrès en matière d'intelligence artificielle, en particulier avec le *deep learning*, sont parfois interprétés comme de « bons » augures de la « singularité » mais rien ne permet de garantir la capacité à créer au cours des prochaines décennies une super-intelligence dépassant l'ensemble des capacités humaines. Par exemple, en s'appuyant sur la loi de Moore, **Ray Kurzweil** prédit que les machines rivalisant avec l'intelligence humaine arriveront **d'ici 2020** et qu'elles le **dépasseront en 2045**<sup>1</sup>. D'après Raja Chatila, directeur de recherche à l'Institut des systèmes intelligents et de robotiques (Isir), nous en sommes, aujourd'hui, encore loin car « *pour être intelligente comme un humain, une machine devra d'abord avoir la perception d'elle-même, ressentir son environnement, traiter des informations à flux continu et en tirer du sens pour ensuite agir (...) Ce n'est pas la technique qui fait défaut à la machine, mais le sens de ses actions, et l'intégration de la notion de concept* ».

Les réflexions de Jean-Gabriel Ganascia et de Max Dauchet devant vos rapporteurs vont également dans ce sens. Cette super-intelligence peut paraître fondée, mais, ainsi que le soulignent Hubert et Stuart Dreyfus, « *cela fait un demi-siècle, depuis que les ordinateurs sont apparus au monde, que l'on promet de bientôt les programmer pour les rendre intelligents et que l'on promet aussi, ou plutôt que l'on a peur, qu'ils apprennent bientôt à nous comprendre nous-mêmes comme des ordinateurs. En 1947, Alan Turing prédisait qu'il existerait un ordinateur intelligent d'ici la fin du siècle. Maintenant que le millénaire est dépassé de trois ans, il est temps de faire une évaluation rétrospective des tentatives faites pour programmer des ordinateurs intelligents comme HAL dans le film 2001, l'Odyssée de l'espace* »<sup>2</sup>. Ce jugement sévère est tempéré par Jean-Gabriel Ganascia selon lequel **on attend toujours l'intelligence artificielle au tournant parce qu'elle déçoit les espérances qu'elles suscitent** : « *HAL, l'ordinateur intelligent du film de Stanley Kubrick 2001, l'Odyssée de l'espace, ne voit pas encore le jour, même si le millénaire est déjà passé. Traduction automatique, compréhension du langage naturelle, vision, démonstration de théorème, résolution de problèmes, robotique... l'histoire récente accumule les échecs. Rien de vraiment tangible n'advient dans ce secteur de la technologie... Autant de lieux communs bien répandus, que l'on retrouve depuis longtemps* ». L'IA forte est une notion implicite chez Marvin Minsky dès la fondation de l'intelligence artificielle mais elle correspond surtout à une invention conceptuelle des années 1980. L'explication donnée par l'un des pères fondateurs de l'intelligence

---

<sup>1</sup> Les prédictions du futurologue peuvent être rappelées. Les ordinateurs atteindront une capacité de traitement comparable au cerveau humain en 2020. En 2022 les États-Unis et l'Europe adopteront des lois réglementant les relations entre les humains et le robot, l'activité des robots, leurs droits, leurs devoirs et autres restrictions seront fixés. En 2031, l'impression des organes humains par des imprimantes 3D sera possible. La circulation de véhicules autonomes sur les routes devient la norme en 2033. Immortalité de l'homme en 2042 et, enfin, la singularité en 2045.

<sup>2</sup> « From Socrates to Expert Systems : The Limits and Dangers of Calculative Rationality », 2004.



---

artificielle au cours de la dernière décennie, Marvin Minsky, est intéressante et pose la question de la place de la recherche publique : « *pourquoi n'avons-nous pas eu HAL en 2001 ? Parce que des problèmes centraux, comme le raisonnement de culture générale, sont négligés, la plupart des chercheurs se concentrent sur des aspects tels que des applications commerciales des réseaux neuronaux ou des algorithmes génétiques* »<sup>1</sup>.

Le principal propagateur de ce risque existentiel est Raymond Kurzweil, sa fonction de directeur de la recherche de Google laissant perplexe. L'IA forte n'est *a priori* pas pour tout de suite mais d'après Stuart Armstrong, chercheur au Futur of Humanity Institute, Oxford, dirigé par Nick Bostrom<sup>2</sup> « *il y a des risques pour que cela arrive plus tôt que prévu* ».

Il convient de noter qu'une étude publiée en octobre 2016 par Google, montre que, programmées pour protéger la confidentialité de communications, deux IA peuvent « *découvrir des formes de chiffrement et déchiffrement, sans qu'on leur ait enseigné des algorithmes spécifiques pour ce faire* », selon les termes des chercheurs. Autrement dit, des formes de communication ignorées des concepteurs des algorithmes. Le constat est d'ailleurs identique avec le système de traduction automatique du même Google, qui intègre désormais des algorithmes de *deep learning*. Une autre étude montre que ce système est capable d'établir des connexions entre des concepts et des mots qui ne sont pas formellement liés. Pour les chercheurs à l'origine de l'étude, la conséquence de ce constat serait claire : le système a développé son propre langage interne, une *interlingua*. Ils précisent toutefois que « *les réseaux neuronaux sont complexes et les interactions difficiles à décrire* ».

## 2. Un prophétisme dystopique indémontrable scientifiquement

Nier la possibilité d'une IA forte n'a pas de sens, toutefois se prononcer sur son imminence ou sur le calendrier précis de son avènement semble tout aussi peu raisonnable, car c'est **indémontrable scientifiquement**. Outre le calendrier incertain de la singularité comme en témoignent les limites de la loi de Moore, il convient d'observer que les théories de la singularité s'apparentent à un prophétisme dystopique. Pour le sociologue Dominique Cardon, la tentation de l'IA forte est anthropomorphiste. Certains sont en effet tentés de plaquer sur les futures intelligences artificielles des modes de raisonnement spécifiques à l'intelligence humaine.

Ce catastrophisme oublie également le **caractère irréductible de l'intelligence humaine au calcul**. Il évacue la place des émotions, de l'intelligence corporelle. Non seulement l'avènement d'une super intelligence à long terme n'est pas certaine mais sa menace à court terme relève du pur fantasme. Pour Mustafa Suleyman, cofondateur de Deep Mind,

---

<sup>1</sup> Marvin Minsky, « *It's 2001. Where Is HAL?* », Dr. Dobb's Technetcast, 2001.

<sup>2</sup> Vos rapporteurs les ont rencontré tous les deux, le second est auteur de « *Superintelligence: Paths, Dangers, Strategy* », 2014.

---

qui est à l'origine d'AlphaGo, il est impossible prévoir de prévoir la date d'arrivée de l'IA forte.

Il s'agit donc de fantasmes, et notamment de **fantasmes sur la capacité des algorithmes à devenir conscients**, à savoir dotés de capacités réflexives leur permettant de se représenter à soi-même. Selon Gérard Sabah « *la bonne question à se poser ce n'est pas de savoir si un système peut être conscient, mais de savoir ce que l'homme attribue à sa conscience* » et pour David Sadek, directeur de la recherche de l'Institut Mines-Télécom, prudent, « *on ne sait pas plus se représenter mentalement le fonctionnement d'une machine inconnue que celui d'un autre être humain* ».

Au-delà des incertitudes sur la capacité d'une machine à avoir une conscience, qui selon certains philosophes, serait une caractéristique réservée aux êtres vivants, se pose la question d'identifier ce que peut faire une technologie d'intelligence artificielle, à savoir calculer. La calculabilité renvoie à la **thèse de Church** (du nom du mathématicien Alonzo Church), qui concerne la définition de la notion de calculabilité : dans une forme dite « physique », elle affirme que la notion physique de la **calculabilité**, définie comme étant tout traitement systématique réalisable par un processus physique ou mécanique, peut être exprimée par un ensemble de règles de calcul. Cette thèse a exercé une influence puissante en informatique, on peut penser à l'article d'Alan Turing de 1936 et à son modèle mécanique de calculabilité. Dans le prolongement de cette thèse figure le « **computationnalisme** », qui est une théorie fonctionnaliste en philosophie de l'esprit qui, pour des raisons méthodologiques, conçoit l'esprit comme un système de traitement de l'information et compare la pensée à un calcul (ou en anglais, computation) et, plus précisément, à l'application d'un système de règles.

À la fin des années 1980, le computationnalisme a été concurrencé par un nouveau modèle cognitif, le connexionnisme, qui vise à montrer qu'on peut expliquer le langage de la pensée sans faire appel à un raisonnement gouverné par un système de règles, comme le fait le computationnalisme.

Le computationnalisme a surtout été la **cible de diverses critiques**, en particulier de John Searle, Hubert Dreyfus, ou Roger Penrose, qui tournaient toutes autour de la réduction de la pensée et/ou de la compréhension à la simple application d'un système de règles. Vos rapporteurs sont sensibles à ces critiques et en particulier à l'illusion de compréhension, telle qu'elle ressort de **l'expérience de la chambre chinoise** mise en évidence par John Searle<sup>1</sup>. Cette expérience vise à montrer qu'une

---

<sup>1</sup> Il pose l'hypothèse d'une personne qui n'a aucune connaissance du chinois enfermée dans une chambre à laquelle est mise à disposition un catalogue de règles permettant de répondre à des phrases en chinois, selon des règles de syntaxe. Cette personne enfermée dans la chambre reçoit donc des phrases écrites en chinois et, en appliquant les règles à sa disposition, produit d'autres phrases en chinois qui constituent en fait des réponses à des questions posées par un vrai sinophone situé à l'extérieur de la chambre. Du point de vue du locuteur qui pose les questions, la personne enfermée dans la chambre se comporte comme un individu qui parlerait vraiment chinois. Mais, en

---

**intelligence artificielle ne peut être qu'une intelligence artificielle faible et ne peut que simuler une conscience**, plutôt que de posséder des authentiques états mentaux de conscience et d'intentionnalité. Elle vise à montrer également que le test de Turing est insuffisant pour déterminer si une intelligence artificielle possède ou non une intelligence comparable à celle de l'homme.

Vos rapporteurs s'interrogent plus largement sur l'idée de **machines qui gagnent contre l'homme** : a-t-on raison de dire, que ce soit aux échecs, au jeu de Go, ou en janvier 2017 au poker, que la machine a battu l'homme et qu'elle est plus forte que lui ?

Ils partagent l'opinion de Laurence Devillers, selon laquelle cette **formulation est excessive et revient à essentialiser la machine**. Cette réification nous fait oublier que **l'intelligence des algorithmes c'est aussi et peut-être avant tout l'intelligence de leurs développeurs**. Elle affirme ainsi que cette victoire au jeu de Go invoque « *une comparaison est nulle et non avenue : on compare un humain avec une machine ; mais derrière la machine se trouvent cent ingénieurs au travail* ».

En effet, AlphaGo est peut-être le **meilleur joueur de Go de tous les temps**, mais il n'est **pas en mesure de parler** ou de **distinguer un chat d'un chien**, ce dont serait, quant à lui, capable n'importe quel joueur de Go humain débutant.

L'écrivain futuriste et entrepreneur, Jerry Kaplan, fait valoir que « *le terme même d'intelligence artificielle est trompeur. Le fait que l'on puisse programmer une machine pour jouer aux échecs, au Go, à Jeopardy ou pour conduire une voiture ne signifie pas pour autant qu'elle soit intelligente ! Autrefois, les calculs étaient effectués à la main, par des humains très intelligents et portant une grande attention au détail. Aujourd'hui, n'importe quelle calculatrice achetée en supermarché peut faire bien mieux que ces brillants cerveaux de jadis. Ces calculatrices sont-elles pour autant intelligentes ? Je ne le crois pas. Au fil du temps, nous découvrons de nouvelles techniques permettant de résoudre des problèmes bien précis, à l'aide de l'automatisation. Cela ne signifie pas pour autant que nous soyons en train de construire une super-intelligence en passe de prendre le pouvoir à notre place* ». Les défis que l'intelligence artificielle devra relever sont d'un autre ordre selon David Sadek puisque « *l'intelligence artificielle, c'est avant tout la didactique des machines, c'est-à-dire apprendre aux machines ce que les humains savent déjà faire (...) quand on parle de vision artificielle, par exemple lorsque Google apprend à identifier un chat sur des photos, ce n'est pas parce qu'un algorithme apprend à repérer un chat qu'il sait ce que c'est. De même, le test de Turing ne sera peut-être pas le meilleur critère pour évaluer les IA à l'avenir* ».

Ces observations conduisent à relativiser les récents progrès de l'intelligence artificielle et, en particulier, à contester le fantasme de l'intelligence artificielle forte car elles récusent la pertinence d'une comparaison avec l'intelligence humaine.

---

*l'occurrence, cette dernière n'a aucune compréhension de la signification des phrases en chinois qu'elle crée. Elle ne fait que suivre des règles prédéterminées.*

---

Selon Raja Chatila, la question initiale de Turing devrait conduire à l'interrogation suivante : une machine peut-elle avoir une *faculté de conscience d'elle-même* ? Car malgré toutes les recherches en robotique et intelligence artificielle, les **résultats, aussi significatifs soient-ils, restent le plus souvent applicables dans des contextes restreints et bien délimités**. Ainsi, la perception ne permet pas à un robot de *comprendre* son environnement, c'est à dire d'élaborer une connaissance suffisamment générale et opératoire sur celui-ci (d'où la nécessité d'étudier la notion d'affordance), la prise de décision reste limitée à des problèmes relativement simples et bien modélisés. Les **principes fondamentaux, qui permettraient aux robots d'interpréter leur environnement, restent largement incompris** : comprendre leurs propres actions et leurs effets, prendre des initiatives, exhiber des comportements exploratoires, acquérir de nouvelles connaissances et de nouvelles capacités... Les clés pour permettre la réalisation de ces fonctions cognitives pourraient être le **méta-raisonnement** et la **capacité d'auto-évaluation**, deux mécanismes réflexifs.

Yann LeCun, dans un article de juillet 2016, se demande « *pourquoi croire à un moment de rupture où les machines seront supérieures à l'homme ? Elles vont simplement devenir de plus en plus intelligentes et de plus en plus faciles à utiliser. Cela amplifiera notre propre intelligence !* » et il estime que « *beaucoup des scénarios catastrophes (en intelligence artificielle) sont élaborés par des personnes qui ne connaissent pas les limites actuelles du domaine. Or les spécialistes disent qu'ils sont loin de la réalité* ».

De même Rob High, directeur technique du projet Watson d'IBM estime qu'il est « *trop tôt pour employer le terme intelligence artificielle, mieux vaut parler d'outils capable d'élargir les capacités cognitives humaines* ».

Lors de sa rencontre avec vos rapporteurs au siège de Google, Greg Corrado, directeur de la recherche en intelligence artificielle chez Google, a lui aussi fait valoir qu'il était plus juste de parler d'**intelligence augmentée** plutôt que d'intelligence artificielle.

Pour Jean-Claude Heudin, **l'intelligence artificielle ne remplace pas l'homme mais augmente son intelligence**, en formant une sorte de « troisième hémisphère ».

Cette idée de **complémentarité hommes-machines et d'intelligence augmentée a convaincu vos rapporteurs**. Selon François Taddéi, directeur du Centre de recherche interdisciplinaire, « *les intelligences humaine et artificielle coévoluent. Mais ce sont encore les combinaisons homme/machine qui sont les plus performantes : on le voit aux échecs, où une équipe homme/machine est capable de battre et l'homme et la machine* ».

En 1996, puis à nouveau en 1997, le champion d'échecs Garry Kasparov était battu par le système d'intelligence artificielle Deep Blue créé par IBM : de nos jours, **n'importe quel programme gratuit d'échecs téléchargeable sur Internet peut battre non seulement les plus grands champions d'échecs mais aussi Deep Blue**. Mais l'homme et la machine,

---

**l'homme-centaure, sont toujours plus forts que toutes les machines.** Les joueurs d'échecs et de Go l'ont expérimenté.

Ainsi que vos rapporteurs ont eu l'occasion de le rappeler précédemment dans le présent rapport, une victoire équivalente d'une intelligence artificielle au jeu de Go semblait impossible, tant ce jeu exige une subtilité et une complexité propres à l'intelligence humaine. Pourtant, le 15 mars 2016, le système d'intelligence artificielle AlphaGo créé par l'entreprise britannique DeepMind, rachetée en 2014 par Google, a battu le champion de Go, Lee Sedol, avec un score final de 4 à 1, marquant donc l'histoire des progrès en intelligence artificielle. Selon l'académicien des sciences Gérard Sabah, **il est difficile et risqué d'établir des prévisions détaillées sur l'avenir à court, moyen et long terme de l'intelligence artificielle.**

Il a toutefois pu émettre quelques hypothèses générales fondées sur l'observation du passé :

- l'intelligence artificielle continuera à faire émerger de **nouvelles techniques de calcul** ;
- certaines de ces nouvelles techniques trouveront des **applications utiles**, mais avec des délais variables, pouvant aller jusqu'à deux décennies ;
- certains de ces nouveaux développements conduiront à la **formation et à la rupture de nouveaux domaines de l'informatique** ;
- la méthodologie des techniques de l'intelligence artificielle, maintenant bien établie par des analyses théoriques et des tests empiriques, permettra d'obtenir des **produits robustes et fiables**.

### **3. Les questions posées par la « convergence NBIC »**

La « convergence NBIC » est un thème issu du rapport particulièrement volumineux de Mihail C. Roco and William Sims Bainbridge à la National Science Foundation (États-Unis) en 2003, intitulé *Converging Technologies for Improving Human Performance*<sup>1</sup>. Il s'agit des convergences entre les nanotechnologies, les biotechnologies, les technologies de l'information et les sciences cognitives. **Ce projet ambitieux de fertilisation croisée n'a pas produit de grands résultats à ce stade mais les progrès en intelligence artificielle, en génomique, en sciences cognitives et en neurosciences reposent la question aujourd'hui.**

Ce projet ambitieux d'un enrichissement mutuel de ces champs de recherche mérite donc que vos rapporteurs le mentionnent. Notre collègue Jean-Yves Le Déaut, président de l'OPECST, conduit d'ailleurs un travail à ce sujet pour l'Assemblée parlementaire du Conseil de l'Europe, en tant que rapporteur pour la science et la technologie. Son projet de **rapport sur la**

---

<sup>1</sup> [http://www.wtec.org/ConvergingTechnologies/Report/NBIC\\_report.pdf](http://www.wtec.org/ConvergingTechnologies/Report/NBIC_report.pdf)

---

**convergence technologique, l'intelligence artificielle et les droits de l'homme** contiendra des pistes pour que cette convergence soit respectueuse des droits humains.

Il pourrait appeler, en particulier, à la vigilance en matière de renforcement de la transparence concernant : les opérations de traitements automatisés visant à collecter, manier et utiliser les données à caractère personnel ; l'information du public sur la valeur des données qu'il produit, le consentement sur leur utilisation et la fixation de la durée du temps de conservation de ces données ; l'information de toute personne sur le traitement de données dont elle est la source ainsi que sur les modèles mathématiques, statistiques qui permettent le profilage ; la conception et l'utilisation de logiciels persuasifs et d'algorithmes respectant pleinement la dignité et les droits fondamentaux de tous les utilisateurs, en particulier les plus vulnérables, dont les personnes âgées et les personnes handicapées.

#### **4. La tentation du « transhumanisme »**

**La crainte d'une intelligence artificielle qui échapperait au contrôle de l'homme est une des angoisses majeures face à l'essor de ces technologies**, comme il a été vu. En cela, la prospective en intelligence artificielle aboutit souvent à des **scénarios de dystopie technologique**. Ce pessimisme n'est cependant pas partagé par l'ensemble des futurologues puisque, pour certains, les progrès de l'intelligence artificielle permettront de protéger la vie humaine, par la sécurité routière et la médecine évidemment, voire d'offrir une **opportunité historique pour concrétiser l'utopie transhumaniste**.

Le transhumanisme est un **mouvement philosophique prédisant et travaillant à une amélioration de la nature de l'homme grâce aux sciences et aux évolutions technologiques**.

Le chercheur et prêtre jésuite Pierre Teilhard de Chardin aurait été, l'un des premiers à utiliser le terme transhumain, en 1951<sup>1</sup>. Julian Huxley<sup>2</sup> crée lui le concept de transhumanisme en 1957. Bien que le terme soit un label recouvrant des mouvements très différents, il renvoie au dépassement des souffrances humaines grâce aux découvertes scientifiques. L'amélioration de nos capacités par l'intermédiaire des sciences et des technologies disponibles est au cœur de ce mouvement. Pour les transhumanistes, le progrès scientifique doit être orienté vers cet objectif. **Le transhumanisme est multidisciplinaire dans la mesure où il agrège, pour parvenir à ses fins, l'ensemble des sciences et des connaissances**. Pour les

---

<sup>1</sup> Sa première occurrence serait la conclusion d'un article intitulé « Du préhumain à l'ultra-humain », paru au sein de l'Almanach des Sciences de 1951.

<sup>2</sup> Frère de l'écrivain Aldous Huxley, Julian Huxley est un biologiste britannique, théoricien de l'eugénisme, premier directeur de l'UNESCO, fondateur du WWF et auteur connu pour ses livres de vulgarisation scientifique.

---

transhumanistes, l'homme « augmenté » pourrait, demain, devenir virtuellement immortel.

Le transhumanisme s'apparente à une religion. Son principal prophète est Raymond Kurzweil, déjà cité s'agissant de la singularité.

**Le projet transhumaniste de mort de la mort et de fin de la souffrance n'empêche pas l'adhésion de vos rapporteurs. Il s'apparente à une négation de la nature humaine. Or pour vos rapporteurs, l'intelligence artificielle n'est pas un acte de foi et ne doit pas le devenir.**

Selon Raja Chatila, « *derrière ces discours, nous avons des vues de l'esprit qui n'ont rien d'opérationnelles, elles sont en réalité des idéologies, qu'on cherche à imposer pour gommer les différences entre l'humain et le non-humain* ». Jean-Gabriel Ganascia relève que ce projet possède un versant cybernétique et un versant plus biologique. Luc Ferry, dans son essai *La Révolution transhumaniste*, souligne ces différentes sensibilités au sein de la « famille » transhumaniste.

Pour vos rapporteurs, il s'agit de chimères ou d'écrans de fumée qui empêchent de se poser les vraies questions pertinentes. Selon eux, il est essentiel de **savoir anticiper les problèmes potentiels posés par l'intelligence artificielle**. À court terme, ces problèmes risquent d'être ignorés et pris à tort pour de la science-fiction. Il convient en effet de **distinguer les craintes issues de certaines fictions cinématographiques, telles que Terminator, Matrix ou Transcendance, des problèmes réels qui risquent de survenir plus ou moins rapidement.**



---

## TROISIÈME PARTIE : LES PROPOSITIONS DE VOS RAPPORTEURS

### I. POUR UNE INTELLIGENCE ARTIFICIELLE MAÎTRISÉE

#### 1. Proposition n° 1 : **Se garder d'une contrainte juridique trop forte sur la recherche en intelligence artificielle, qui - en tout état de cause - gagnerait à être, autant que possible, européenne, voire internationale, plutôt que nationale**

Vos rapporteurs appellent à **ne pas céder à la tentation de proposer et de mettre en place des mesures** dans un cadre trop étroitement national, qui décourageraient la recherche en intelligence artificielle, mais plutôt à soutenir une forme de régulation internationale. Il convient de **se garder d'une contrainte juridique trop forte** sur la recherche dans ce domaine, qui gagnerait à être, en tout état de cause, instaurée à l'échelle européenne, voire internationale, mais pas uniquement nationale.

Le professeur de droit Albert De Lapradelle affirmait, de manière provocatrice, que *« ce ne sont pas les philosophes avec leurs théories, ni les juristes avec leurs formules, mais les ingénieurs avec leurs inventions qui font le droit et surtout le progrès du droit »*. Vos rapporteurs estiment que **c'est surtout au législateur de réfléchir au droit souhaitable en matière d'intelligence artificielle et de robotique**, au terme d'un **débat public éclairé avec toutes les parties prenantes, du citoyen ordinaire à l'expert scientifique, en passant par l'entrepreneur et le technicien**.

Il faudra savoir **adapter la législation aux nouveaux risques** posés par l'intelligence artificielle et la robotique, mais il faudra aussi **veiller à ne pas freiner le développement de ces technologies**. Des mesures trop contraignantes auraient pour effet **d'augmenter les coûts de commercialisation et de mise en conformité**, ou de poser de nouveaux freins légaux à l'innovation et à la rentabilité de ces secteurs d'activité. Un **équilibre subtil est à trouver et le présent rapport entend contribuer à la réflexion collective en la matière**.

**La peur ne doit pas paralyser** : vos rapporteurs ne veulent pas tomber dans la solution de facilité qui consisterait à faire un usage extensif du principe de précaution et donc à limiter, *a priori*, la recherche en intelligence artificielle. Une telle démarche serait contraire à l'esprit scientifique, que vos rapporteurs revendiquent. Elle serait également **préjudiciable à l'intérêt national** : sans rivaliser directement avec les États-Unis, la Chine ou le Japon en la matière, la France dispose d'**atouts considérables en matière de recherche en intelligence artificielle et ne doit pas perdre cet avantage comparatif, au risque de se placer hors-jeu dans la compétition internationale qui s'est engagée**.

---

## 2. Proposition n° 2 : Favoriser des algorithmes et des robots sûrs, transparents et justes, et prévoir une charte de l'intelligence artificielle et de la robotique

Il faut favoriser **des algorithmes et des robots** qui soient à la fois **sûrs, transparents et justes**. Il faut aussi prévoir une **charte de l'intelligence artificielle et de la robotique**, qui nécessitera **du temps et une concertation internationale**.

Les considérations de sécurité et de robustesse évoquées dans le présent rapport amènent à conclure qu'il est pertinent de prévoir des mécanismes d'arrêt d'urgence. **L'interruption d'un système d'intelligence artificielle ou d'un robot doit toujours être possible**. Des « boutons rouges », réels ou virtuels, doivent permettre la désactivation des systèmes. Une telle sécurité est souhaitable mais elle peut être difficile à mettre en œuvre.

Il faut rechercher la **transparence des algorithmes** contre les boîtes noires, les biais et les discriminations.

Il convient de prévoir aussi des mécanismes de traçabilité, du type **journaux de bord** ou *loggings*<sup>1</sup>, ces « mémoires internes » de type enregistreurs numériques des avions (appelées aussi « boîtes noires », mais qui n'ont rien à voir avec le qualificatif de « boîtes noires » appliqué aux algorithmes de *deep learning*). Pour pouvoir expliquer un éventuel dysfonctionnement, il faut toujours pouvoir **remonter aux derniers circuits de décision et leurs motifs**. Comme il a été dit précédemment, la **transparence des algorithmes** de *deep learning* n'existe pas, à ce stade, et reste à construire. Il s'agit d'un **axe prioritaire pour la recherche**. L'initiative TransAlgo de l'INRIA, avec le concours du Conseil national du numérique (CNNum), va dans ce sens et mérite d'être applaudie. Mais elle doit se prolonger et prendre davantage d'ampleur, pour dépasser le strict cadre de l'INRIA. La question de la transparence des algorithmes ne se confond pas avec la question de leur loyauté, mais elle doit aussi être posée. Il faut noter qu'à la différence des méthodes d'apprentissage automatique profond tels que les algorithmes de *deep learning*, **le traitement et l'analyse de données effectués par les réseaux bayésiens sont des systèmes transparents**. Il ne revient pas aux pouvoirs publics de choisir entre telle ou telle technologie mais, **à performance égale, les technologies transparentes doivent être privilégiées**.

Pour Raja Chatila, directeur de l'Institut de Systèmes Intelligents et de Robotique, « *le bouton rouge, c'est bien mais c'est déjà trop tard, il faut tout*

---

<sup>1</sup> En informatique, ces journaux, historiques des événements ou *loggings* désignent « l'enregistrement séquentiel dans un fichier ou une base de données de tous les événements affectant un processus particulier (application, activité d'un réseau informatique...). Le journal (en anglais log file ou plus simplement log), désigne alors le fichier contenant ces enregistrements. Généralement datés et classés par ordre chronologique, ces derniers permettent d'analyser pas à pas l'activité interne du processus et ses interactions avec son environnement » (source : wikipédia).

---

*prévoir pour ne pas en arriver là. L'état du système doit être constamment observé et il faut pouvoir détecter toute déviance avant l'arrivée des problèmes* ». Il faut porter, selon lui, une grande attention aux **signes précurseurs**, d'autant plus que les systèmes régis par des technologies d'intelligence artificielle peuvent avoir de vastes implications, qui seront de plus en plus stratégiques, à l'image de l'alimentation d'une grande ville en électricité. Il rappelle que tout robot peut virtuellement être dangereux, puisque qu'il s'agit d'un objet puissant en mouvement : ses mouvements peuvent devenir violents et son alimentation électrique peut impliquer des risques d'électrocution. Les dispositifs de vigilance mis en œuvre doivent permettre de relever ce défi.

La question de **la faisabilité de ces préconisations exigeantes, visant à disposer d'algorithmes et de robots qui soient à la fois sûrs, transparents et justes, se pose**. Il n'est pas sûr qu'elles puissent être satisfaites, mais ce sont des **objectifs que nous devons nous fixer**, quoiqu'il en soit.

La **charte de l'intelligence artificielle et de la robotique** proclamerait ces objectifs et viserait à codifier les bonnes pratiques. Elle devrait être **internationale**, dans toute la mesure du possible, ou à défaut européenne. Elle ne se prononcerait **pas en faveur de la création d'une personnalité morale des robots**, mais proposerait des règles relatives aux interactions homme-machine et poserait des limites en matière d'imitation du vivant, pour les robots androïdes comme pour les agents conversationnels. Le citoyen devrait toujours savoir s'il est en présence d'une machine ou d'un humain.

Cette charte pourrait, en outre, proposer d'interdire les robots tueurs. Toutefois, ce débat aurait davantage sa place dans le **cadre international des conventions sur les armes et du traité sur la non-prolifération des armes nucléaires (TNP)**.

### **3. Proposition n° 3 : Former à l'éthique de l'intelligence artificielle et de la robotique dans certains cursus spécialisés de l'enseignement supérieur**

Nous devons offrir de façon systématique des **formations à l'éthique de l'intelligence artificielle et de la robotique** dans les **cursus de l'enseignement supérieur** qui traitent de l'intelligence artificielle et de la robotique. Google Deep Mind le propose, d'ailleurs, pour tous les masters britanniques spécialisés. La CERNA pourrait être associée à ce projet.

Les formes de cette offre de formations devront être innovantes, **mobiliser les étudiants, à travers des ateliers, des débats et des mises en situation**. Il conviendra d'éviter la forme du module de cours théorique, validé après le suivi d'un simple cours magistral.

### **4. Proposition n° 4 : Confier à un institut national de l'éthique de l'intelligence artificielle et de la robotique un rôle d'animation**

---

## du débat public sur les principes éthiques qui doivent encadrer ces technologies

Au-delà de la nouvelle mission de la CNIL, il faut **confier à un institut national de l'éthique de l'intelligence artificielle et de la robotique un rôle d'animation du débat public sur les principes éthiques qui doivent encadrer ces technologies**. Il peut s'agir d'un institut ou d'une commission nationale. La création d'un comité national d'éthique de la robotique est proposée par le **Symop**, et celle d'un comité national d'éthique de l'intelligence artificielle a été évoquée devant vos rapporteurs par **Max Dauchet et Laurence Devillers**.

Les expériences de la CNIL, de la CERNA et du COMETS seront à considérer. Il n'est pas certain qu'il faille **séparer la recherche en intelligence artificielle de la réflexion éthique sur l'intelligence artificielle**, mais il est certain que **les chercheurs spécialisés en intelligence artificielle et en robotique ne doivent pas être les seuls à participer à cette réflexion**.

L'institut devra s'ouvrir aux SHS, aux associations et aux ONG. Les pouvoirs publics ne devront pas être les seuls à en assurer le financement. Les entreprises privées, qui se donnent pour objectif **d'informer et d'éduquer dans le domaine des ces technologies et d'accroître leurs effets bénéfiques pour la société** (à l'image du « *partnership on AI* »), pourraient participer à son financement..

Cet institut national de l'éthique de l'intelligence artificielle et de la robotique pourra s'intéresser aux problématiques « d'explicabilité » évoquées plus haut, d'autant plus que le projet **Trans-algo pose le problème de sa restriction à l'INRIA**. La démarche ne doit pas être réservée à une seule structure de recherche, mais **plusieurs équipes doivent y travailler parallèlement et un institut national pourrait impulser les projets, coordonner les recherches, animer le débat public et faire des propositions aux autorités publiques**.

### 5. Proposition n° 5 : Accompagner les transformations du marché du travail sous l'effet de l'intelligence artificielle et de la robotique en menant une politique de formation continue ambitieuse visant à s'adapter aux exigences de requalification et d'amélioration des compétences

Les transformations du marché du travail sous l'effet de l'intelligence artificielle et de la robotique imposent de mener **une politique de formation continue ambitieuse** visant à s'adapter aux exigences de requalification et d'amélioration des compétences. Il s'agit à la fois **d'accompagner les transformations du marché du travail, d'adapter le système éducatif** et de développer les **formations professionnelles idoines** pour garantir une souplesse suffisante dans la reconversion des travailleurs.

---

Les métiers du futur devront relever les défis technologiques posés par l'intelligence artificielle et la robotique, sachant que **les hommes bénéficieront de nouvelles activités liées à la diffusion de ces technologies** : concevoir, entraîner, éduquer, surveiller, réparer les systèmes et les machines. **De nouveaux métiers deviendront envisageables**, et ils apporteront autant de **nouvelles opportunités d'emplois**. Parmi ces métiers qui vont émerger, certains sont encore inconnus, mais d'autres peuvent d'ores et déjà être identifiés : **concepteur, entraîneur, éducateur, vérificateur ou, encore, évaluateur des systèmes d'intelligence artificielle et des robots**.

## II. POUR UNE INTELLIGENCE ARTIFICIELLE UTILE, AU SERVICE DE L'HOMME ET DES VALEURS HUMANISTES

### 1. Proposition n° 6 : Redonner une place essentielle à la recherche fondamentale et revaloriser la place de la recherche publique par rapport à la recherche privée tout en encourageant leur coopération

Il faut parvenir à **redonner une place essentielle à la recherche fondamentale** et à **revaloriser le rôle de la recherche publique par rapport à la recherche privée**, tout en encourageant leur coopération.

Seule la recherche fondamentale peut permettre de **répondre aux problèmes d'explicabilité des algorithmes et de biais dans les données**. Nous en avons besoin car, des enjeux de maîtrise des technologies à ceux du du financement de la recherche publique, tout se tient.

La recherche fondamentale pose, par ailleurs, la question du mode de financement des projets : il faut **favoriser la recherche transversale et ne surtout pas reproduire l'hyper-spécialisation entre sous-domaines de l'intelligence artificielle**. Ce serait une erreur de financer d'un côté la perception, d'un autre, la vision, encore d'un autre la prise de décisions, l'apprentissage machine, la robotique, la relation homme-machine... Il faut **mobiliser les équipes de chercheurs autour de grands projets nationaux structurants** et résister aux sirènes de l'hyper-spécialisation entre sous-domaines de l'intelligence artificielle. D'après les informations recueillies par vos rapporteurs, le Commissariat général à l'investissement (CGI), Bpifrance et l'initiative du Gouvernement France IA orienteraient le futur PIA dans le sens de cette hyper-spécialisation, ce qui est de nature à les inquiéter.

Le mode de financement des projets pourrait gagner à **s'inscrire dans des temps plus longs que 3, 4 ou 5 ans seulement**, de manière à porter les projets à leur pleine maturation. C'est le cas de la NSF aux États-Unis.

Des projets de **grandes bases de données labellisées**, nécessaires à l'apprentissage machine, pourraient être lancés, par exemple autour de la langue française, du marché de l'emploi ou des données de santé, sous condition d'anonymisation. Ces initiatives pourraient se faire en lien avec la

---

CNAM, l'APHP, la BNF, l'INA ou, encore, avec Pôle emploi. Dans son intervention lors de la journée « Entreprises françaises et intelligence artificielle » organisée par le MEDEF et l'AFIA le 23 janvier 2017, Yves Caseau, animateur du groupe de travail sur l'IA de l'Académie des Technologies, a fourni les premières recommandations de son groupe, la première consistant à la **collecte en France de jeux de données massifs**<sup>1</sup>, car ces jeux de données sont essentiels au développement de l'intelligence artificielle par l'apprentissage

Quant au **niveau des investissements requis, vos rapporteurs ne se sont pas avancés à réaliser leur propre chiffrage**. Selon Bertrand Braunschweig, directeur du centre de l'Inria de Saclay, l'effort financier nécessaire serait de l'ordre de **10 milliards d'euros**, financements publics et privés compris.

De nombreuses coopérations public-privé existent et fonctionnent et il est loisible de s'en inspirer (Microsoft-Inria, CNRS avec RHODIA...)

Jean-Gabriel Ganascia pose la question de l'articulation entre recherche publique et recherche privée de la manière suivante : *« depuis une trentaine d'année, un accent très fort a été mis au plan européen et au plan national, sur les projets collaboratifs entre l'université et l'industrie. Cela a certainement eu des aspects très bénéfiques, mais ce mode de coopération présente aussi des limites. En effet, le caractère très administratif du montage des projets qui conduit, en particulier au plan européen, à solliciter des sociétés spécialisées pour le montage des projets, le taux d'acceptation des projets (moins de 8%, parfois de 2% pour certaines actions), les procédures d'évaluation très rigides, conduisent à une recherche chère, conventionnelle et peu innovante, tendant à stériliser les équipes. À cela, il faut ajouter qu'en l'absence de post-évaluation sérieuse, on n'est pas en mesure de tirer parti des résultats de projets financés, ce qui conduit à un gaspillage des ressources. Il faudrait encourager des partenariats bilatéraux entre une équipe privée et un laboratoire public, avec des actions plus légères. Il est à noter que des actions à plus long terme, comme les Labex, ou des financements de bourses de type « CIFRE » complétés par des contrats conclus grâce au crédit d'impôt recherche, peuvent avoir des effets très positifs ».*

Il est important de **produire de l'intelligence artificielle « à la française »**, en intégrant notamment les sciences humaines et sociales. Les réponses possibles à la question de la vulnérabilité de la société aux bouleversements des innovations technologiques doit pouvoir s'appuyer sur les sciences humaines.

---

<sup>1</sup> Les trois autres recommandations sont :

- commencer à utiliser des réseaux neuronaux afin de résoudre des problèmes de classification externe (exemple : TensorFlow outil en Open source développé par Google) ;
- maîtriser la technologie d'automatisation (robotic process automation) en augmentant le volume d'investissement dans la robotique de pointe et de robotique à haute performance ;
- implémenter les premiers chatbots d'assistance client sur des périmètres fonctionnels simples.



---

## **2. Proposition n° 7 : Encourager la constitution de champions européens en intelligence artificielle et en robotique**

Il faut **encourager la constitution de champions européens en intelligence artificielle et en robotique**, sur le modèle d'Airbus, ce qui implique de poser la question du droit de la concurrence dans l'UE.

Sans verser directement dans le nationalisme industriel appliqué à l'intelligence artificielle, il faut réfléchir aux formes de protection qui pourraient être instituées, car les laboratoires français sont pillés de leurs chercheurs par des multinationales nord-américaines et chinoises. **La question du soutien à la création et à la croissance des entreprises et start-ups françaises de ce secteur doit être étudiée**, afin de faire **émerger une industrie française de l'intelligence artificielle** qui développera des produits innovants exportables. Là-aussi, des protections pourraient être utiles.

Comme l'affirment justement Thierry de Montbrial et Thomas Gomart, dans un ouvrage qu'ils viennent de publier, nous devons défendre « *notre intérêt national* » et vos rapporteurs y ajoutent *notre intérêt européen*.

## **3. Proposition n° 8 : Orienter les investissements dans la recherche en intelligence artificielle vers l'utilité sociale des découvertes**

Il faut orienter les investissements dans la recherche en intelligence artificielle vers l'utilité sociale des découvertes, en encourageant les applications à impact sociétal bénéfique : bien-être, santé, dépendance, handicap, infrastructures civiles, gestion des catastrophes. L'IA ne doit pas s'intéresser qu'au *trading* à haute fréquence THF.

Il convient de saisir, dans notre intérêt national, les opportunités ouvertes par les technologies d'intelligence artificielle. Là aussi, les entreprises privées, qui se donnent pour objectif d'informer et d'éduquer dans le domaine de ces technologies et d'accroître leurs effets bénéfiques pour la société, pourraient participer à l'effort collectif nécessaire.

## **4. Proposition n° 9 : Élargir l'offre de cursus et de modules de formation aux technologies d'intelligence artificielle dans l'enseignement supérieur et créer - en France - au moins un pôle d'excellence international et interdisciplinaire en intelligence artificielle et en robotique**

Il est nécessaire **d'élargir l'offre de cursus et de modules de formation aux technologies d'intelligence artificielle dans l'enseignement supérieur et de créer en France au moins un pôle d'excellence international et interdisciplinaire en intelligence artificielle et en robotique**. Il s'agit de renforcer ces formations, alors que des besoins considérables apparaîtront bientôt pour le développement de l'intelligence artificielle. Des cycles de



---

formation longs apparaissent donc nécessaires, mais des cycles de formation plus courts et plus professionnalisants le sont aussi.

Il sera possible de s'appuyer sur l'expérience des 138 cours spécifiques en IA, ou techniques en relation avec l'IA, dispensés dans l'enseignement supérieur français chaque année et sur les 16 masters existants qui, de près ou de loin, sont spécialisés en intelligence artificielle en France, à l'image du Master MVA-MATH de l'ENS Cachan, du Master en sciences cognitives (Cog master) de l'ENS Ulm, des Masters ANDROIDE (AgeNts Distribués, Robotique, Recherche opérationnelle, Interaction, Decision) et DAC (Master Données, Apprentissage et Connaissances de l'Université Pierre et Marie Curie Paris 6 (UPMC), du Master en informatique spécialité intelligence artificielle de Université Paris-Descartes ou, encore du Master « Intelligence Artificielle et Reconnaissance des Formes » de l'Université Paul Sabatier de Toulouse.

En outre, il est nécessaire de créer en France au moins un **pôle d'excellence international et interdisciplinaire** en intelligence artificielle et en robotique. Jusqu'à deux ou trois pôles pourraient voir le jour, en s'appuyant sur l'excellence de l'INRIA, du LAS de Toulouse, de l'ENS, de l'institut Mines Télécom... Mais, à défaut d'une création pure et simple, il paraît **urgent d'encourager la coordination et d'accroître la cohérence des instituts et des centres existants, et de leurs équipes de recherches.**

Vos rapporteurs ont visité au Royaume-Uni, en Suisse et aux États-Unis des pôles d'excellence à visibilité internationale et à vocation interdisciplinaire. Ils jugent indispensables de s'inspirer de ces pôles pour structurer la recherche française, qui est déjà d'excellent niveau. Et comme l'a fait remarquer le jeune chercheur Jill-Jënn Vie à vos rapporteurs, le cas de son laboratoire de recherche Riken en intelligence artificielle à Tokyo montre qu'il est possible de développer l'excellence à l'échelle de tout un pays, ici le Japon, et que ce rayonnement est utile à tous les chercheurs concernés.

**5. Proposition n° 10 : Structurer et mobiliser la communauté française de la recherche en intelligence artificielle en organisant davantage de concours primés à dimension nationale, destinés à dynamiser la recherche en intelligence artificielle, par exemple autour du traitement de grandes bases de données nationales labellisées**

L'initiative « France IA » est une étape importante dans la **mobilisation de la communauté française de la recherche en intelligence artificielle**. Il faut continuer et structurer encore davantage celle-ci, par exemple en organisant davantage de **concours primés à dimension nationale**, destinés à dynamiser la recherche en intelligence artificielle, par exemple autour du traitement de grandes bases de données nationales labellisées.

---

Les **projets de la DARPA** reposent souvent sur ce modèle, qui crée une saine émulation au sein des équipes. Un **travail avec l'ANR** peut être envisagé pour définir une offre française de grands **concours primés en IA**.

Le traitement de grandes bases de données nationales labellisées nécessitera de mettre à disposition des **jeux de données massifs**, ce que le Gouvernement pourrait favoriser comme il a été vu dans la proposition n°1.

### **6. Proposition n° 11 : Assurer une meilleure prise en compte de la diversité et de la place des femmes dans la recherche en intelligence artificielle**

**La place des femmes et la question des minorités dans la recherche en intelligence artificielle sont des défis qu'il convient de relever.**

Diversifier le profil des chercheurs en intelligence artificielle et travailler sur le thème de la diversité dans ce secteur apparaissent nécessaires, vos rapporteurs ayant noté la **sous-représentation des minorités** dans ce secteur. La réponse qui leur a été faite, selon laquelle ce serait le cas plus globalement pour tout le secteur de l'informatique et pour les filières scientifiques en général, ne suffit pas à justifier la situation présente.

Ils notent, d'ailleurs, que le Gouvernement a signé, le 31 janvier 2017, un plan sectoriel « mixité dans les métiers du numérique » avec une quinzaine de structures telles que Cap digital, le Syndicat national du jeu vidéo, TECH-IN France ou Syntec numérique. Des plans similaires ont été lancés auparavant dans le secteur du bâtiment ou les transports. Il est vrai qu'on ne dénombre que 28 % de femmes dans le secteur du numérique contre 48 % dans le reste de l'économie, et que derrière ce pourcentage se cache une réalité encore moins admissible : les femmes sont surtout présentes dans les emplois de secrétariat du secteur du numérique, mais sont particulièrement sous-représentées dans les métiers de techniciens, d'informaticiens ou d'ingénieurs.

Là-aussi, les entreprises privées, qui se donnent pour objectif d'informer et d'éduquer dans le domaine de ces technologies et d'accroître leurs effets bénéfiques pour la société, pourraient participer à l'effort collectif nécessaire.

## **III. POUR UNE INTELLIGENCE ARTIFICIELLE DÉMYSTIFIÉE**

### **1. Proposition n° 12 : Organiser des formations à l'informatique dans l'enseignement primaire et secondaire faisant une place à l'intelligence artificielle et à la robotique**

Toute formation à l'informatique doit s'accompagner d'une consolidation de l'apprentissage des mathématiques. Des **formations spécifiques à l'informatique** sont dispensées dans l'enseignement primaire

et secondaire, mais ces enseignements sont le plus souvent **facultatifs**<sup>1</sup> et restent **insuffisants**. Ils ne font pas toujours, en outre, de **place à l'intelligence artificielle et à la robotique**. Il s'agit d'une insuffisance à laquelle il convient de remédier en urgence. La **formation des enseignants** est aussi une priorité.

Il peut être noté qu'un **plan numérique pour l'éducation** a été lancé par le Président de la République en mai 2015<sup>2</sup>, à l'issue de la concertation nationale sur le numérique éducatif, et que le conseil supérieur des programmes (CSP) travaille depuis 2015 à un **projet de programme pour un enseignement d'exploration d'informatique et de création numérique**<sup>3</sup>.

## **2. Proposition n° 13 : Former et sensibiliser le grand public à l'intelligence artificielle par des campagnes de communication, l'organisation d'un salon international de l'intelligence artificielle et de la robotique et la diffusion d'émissions de télévision pédagogiques**

Il importe de **former et de sensibiliser le grand public à l'intelligence artificielle par des campagnes de communication, par l'organisation d'un salon international de l'intelligence artificielle et de la robotique et par la diffusion d'émissions de télévision pédagogiques**. Le

---

<sup>1</sup> Le programme d'enseignement facultatif d'informatique et création numérique pour les classes de première des séries générales et les classes terminales des séries ES et L peut être lu ici : [http://www.education.gouv.fr/pid285/bulletin\\_officiel.html?cid\\_bo=104657](http://www.education.gouv.fr/pid285/bulletin_officiel.html?cid_bo=104657) . Il peut être rapproché de celui d'enseignement de spécialité d'informatique et sciences du numérique de la série scientifique : [http://www.education.gouv.fr/pid25535/bulletin\\_officiel.html?cid\\_bo=57572](http://www.education.gouv.fr/pid25535/bulletin_officiel.html?cid_bo=57572)

<sup>2</sup> Il s'agit de mieux préparer les élèves à être acteur du monde de demain en développant des méthodes d'apprentissages innovantes pour favoriser la réussite scolaire et développer l'autonomie ; en formant des citoyens responsables et autonomes à l'ère du numérique ; et en préparant les élèves aux emplois digitaux de demain. Sa mise en œuvre repose sur quatre piliers : la formation, les ressources, l'équipement et l'innovation. Pour aider les enseignants à faire évoluer leur pratique pédagogique en intégrant harmonieusement les outils numériques à leurs cours, un programme de formation à la fois initiale et continue est mis en place sur l'ensemble du territoire :

- une formation de trois jours par an dédiée au numérique à destination des enseignants et chefs d'établissement de collège ;

- des formations mises en place au niveau de l'établissement pour une meilleure prise en main des outils numériques ;

- des formations à distance pour tous les enseignants et les professeurs stagiaires via la plateforme de formation M@gistère;

- le développement de cours en ligne (Moocs) pour les enseignants et les professeurs stagiaires sur le portail France université numérique (FUN-Mooc).

Il s'agit de retenir trois axes de formation : la maîtrise des outils numériques pour une meilleure prise en main des outils par les enseignants ; les usages du numérique dans les disciplines pour développer de nouvelles méthodes d'enseignement ; et la culture numérique et l'éducation aux médias et à l'information pour transmettre aux enseignants les bases essentielles liées à l'usage d'internet et des réseaux sociaux. Enfin, une plateforme en ligne nationale (Myriaé) est mise à disposition des enseignants en vue de présenter toutes les ressources pédagogiques numériques, gratuites ou payantes, produites par les éditeurs privés ou publics. Cf. <http://ecolenumerique.education.gouv.fr/plan-numerique-pour-l-education/>

<sup>3</sup> Cf. <http://www.education.gouv.fr/cid89179/projet-de-programme-pour-un-enseignement-d-exploration-d-informatique-et-de-creation-numerique.html>

---

CNSTI est un organe utile en la matière. Il convient de faire savoir que les progrès en intelligence artificielle sont d'abord bénéfiques à la société, et que les risques éventuels doivent être anticipés et peuvent être maîtrisés. Enfin, l'avènement d'une intelligence artificielle forte reste peu probable. Nous nous orientons plus vraisemblablement vers de nouvelles formes d'intelligence et de nouvelles complémentarités entre l'homme et la machine, qui conduiront à une « intelligence augmentée ».

Un **salon international de l'intelligence artificielle et de la robotique est à organiser en France**, en s'inspirant de *VivaTech* et de l'initiative *Innorobo*<sup>1</sup> portée par Catherine Simon, organisatrice du salon français de la robotique. Il n'existe pas de salon européen du type du *Consumer Electronics Show (CES)* de Las Vegas, qui est un salon consacré à l'innovation technologique en électronique<sup>2</sup> orienté vers le grand public dès sa création en 1967, puisqu'il est organisé par la *Consumer Technology Association*. Le grand événement professionnel européen en la matière est, par défaut, la foire de Hanovre, le plus vaste salon au monde pour la technologie industrielle, qui réunit, selon les années, entre 7 et 13 salons phares internationaux au même endroit : *Industrial Automation*, *MDA (Motion, Drive & Automation)*, *Digital Factory*, *ComVac*, *Industrial Supply*, *Energy*, *Power Plant Technology*, *Wind*, *MobiliTec*, *CoilTechnica*, *SurfaceTechnology*, *IndustrialGreenTec*, *Research & Technology*. Un autre salon européen est *Automatica*, salon international qui se tient aussi en Allemagne, à Munich, et regroupe tous les professionnels du secteur de l'automatisme industriel et de la robotique.

**Les réseaux sociaux ou la télévision pourraient être des supports pour des émissions de partage de la connaissance.** Les émissions de télévision du type de « la faute à l'algo »<sup>3</sup>, diffusée par la chaîne No Life, ou

---

<sup>1</sup> *Innorobo est une entreprise d'Impact Consulting spécialisée dans le business développement par l'innovation et un expert international des marchés robotiques mondiaux, qui promeut une approche humaine des technologies robotiques. Les actions d'Innorobo s'expriment selon 3 axes : son événement international, ses « Ressources » et sa communauté. Tout au long de l'année, Innorobo interagit avec un réseau de plus de 3 500 organisations robotiques dans le monde et 20 000 leaders et décideurs, tous acteurs de l'innovation ouverte, qui voient les technologies, produits et services robotiques non seulement comme des opportunités de croissance et de compétitivité par l'innovation, mais aussi comme une source de progrès pour l'Humain. Innorobo est fermement convaincu que les objectifs économiques peuvent servir et être alignés avec une plus grande cause, celle d'une humanité durable. Cf [www.innorobo.com](http://www.innorobo.com)*

<sup>2</sup> Voir le site officiel du salon : <http://www.ces.tech/>

<sup>3</sup> *L'émission La Faute à l'algo est écrite et réalisée par Michel Blockelet et Jill-Jënn Vie en collaboration avec la chaîne Nolife, qui en assure la production. Présentée par Frédéric Hosteing, elle retrace de façon pédagogique divers moments de nos vies où les algorithmes ont échappé à notre contrôle, et où des bugs ont eu des répercussions parfois insolites, parfois désastreuses sur notre économie ou nos libertés. L'émission, dont la musique est composée par un algorithme, se déroule en 2098, et son synopsis est le suivant : « 2098. Les Algorithmes ont pris le contrôle de notre société. Mais comment en sommes-nous arrivés là ? Voici quelques vidéos du futur pour prendre conscience du rôle grandissant des algorithmes dans nos vies, à l'origine de notre déchéance ». L'émission a été diffusée d'octobre 2015 à décembre 2016, à raison de 23 épisodes de 6 minutes, portant sur des sujets allant de la transparence des algorithmes jusqu'aux monnaies virtuelles, en passant par l'ubérisation, cf. <http://fautealgo.fr>*

---

de certains épisodes de « data-gueule », diffusés par la chaîne France 4, en sont des exemples intéressants.

Une mobilisation à cette fin de communication des entreprises et des chercheurs est indispensable. Il faut **faire preuve de pédagogie, expliquer que l'intelligence artificielle est complémentaire de l'homme, qu'elle ne le concurrence pas**. En dépit des métaphores anthropomorphiques et de la réification de l'IA, il s'agit simplement de sciences et de technologies du traitement automatique de l'information, qui impliquent davantage de progrès que de risques, même si ces derniers doivent être maîtrisés.

Il faut se saisir du *partnership on AI* pour associer les GAFAMI voire les « GAFAMITIS »<sup>1</sup> à ce travail pédagogique. Là aussi, le **coût du financement pourrait être partagé avec les entreprises privées**, qui se donnent pour objectif d'informer et d'éduquer dans le domaine de ces technologies. Pour votre rapporteur Claude de Ganay, il est déplorable que les grand-messes d'Apple ne s'accompagnent jamais de présentations pédagogiques sur les technologies d'intelligence artificielle.

### **3. Proposition n° 14 : Former et sensibiliser le grand public aux conséquences pratiques de l'intelligence artificielle et de la robotisation**

Il s'agit de former et de sensibiliser le grand public aux conséquences pratiques de l'intelligence artificielle et de la robotisation. Avec un effort de formation continue visant l'amélioration continue des compétences, il s'agit de **permettre aux travailleurs et au grand public d'envisager de manière positive les transitions à venir**.

Selon votre rapporteur Claude de Ganay, il ne faut pas uniquement s'adapter aux exigences de requalification et d'amélioration des compétences, mais redéfinir un projet de société partagé en débattant des **conséquences pratiques de l'intelligence artificielle et de la robotisation**.

Là encore, le **coût du financement pourrait être partagé** avec les entreprises privées, qui se donnent pour objectif d'informer et d'éduquer dans le domaine de ces technologies.

---

<sup>1</sup> Google, Apple, Facebook, Amazon, Microsoft, IBM, Twitter, Intel et Salesforce. Ces entreprises américaines représentent la pointe de la recherche en IA et de ses applications.

---

#### 4. Proposition n° 15 : Être vigilant sur les usages spectaculaires et alarmistes du concept d'intelligence artificielle et de représentation des robots

Vos rapporteurs appellent à la **vigilance sur les usages spectaculaires et alarmistes du concept d'intelligence artificielle et de représentation des robots**. Il s'agit d'éviter les dérapages, dans le respect de la liberté de création et de la liberté d'expression. La **vérification des publicités** serait un premier pas vers une plus grande maîtrise de la communication médiatique sur le sujet. Votre rapporteur Claude de Ganay invite à ne pas limiter la représentation courante de l'intelligence artificielle aux robots qui nous concurrenceraient ou nous menaceraient en tant qu'êtres humains. Face aux scénarios catastrophistes, il faut **sensibiliser les écoles de journalisme** à l'intérêt d'une présentation équilibrée et nuancée de l'intelligence artificielle, et à leur responsabilité en la matière. La loi pour une République numérique est un pas utile dans la direction d'une société qui perçoit la révolution numérique de façon positive.





---

## CONCLUSION

Les propositions du présent rapport devront être **remises en débat au fur et à mesure des nouvelles découvertes scientifiques**, de leurs transferts et de leurs usages. Vos rapporteurs tiennent à ce que le point d'équilibre qu'ils ont cherché à atteindre dans le présent rapport puisse évoluer, en fonction des évolutions du contexte résultant du jeu de ces variables.

Vos rapporteurs appellent à la **poursuite des travaux de l'OPECST sur les enjeux de l'intelligence artificielle en 2017 et 2018**. Ce suivi pourra prendre la forme d'une veille générale des rapporteurs, d'une incitation à la reprise de leurs propositions par le Gouvernement, ainsi que d'un approfondissement de leur travail, le cas échéant en ciblant leurs investigations plus particulièrement sur certaines dimensions ou sur certains secteurs. **Le suivi par l'OPECST d'un sujet aussi important et mouvant apparait indispensable.**

Ils proposent, en outre, la **poursuite du plan national pour l'intelligence artificielle** annoncé en janvier 2017, puis précisé de manière plus détaillée à la fin du mois de mars 2017.

Ils forment le vœu que ce plan connaisse de francs succès, de manière moins contrastée que le « plan Calcul » lancé en 1966 ou que le plan « Informatique pour tous » lancé en 1985. Au cours de leurs investigations sur l'intelligence artificielle, vos rapporteurs ont eu à l'esprit le **rapport sur l'informatisation de la société** publié en 1978 : il préconisait de manière audacieuse d'associer les télécommunications et l'informatique grâce à la connexion de terminaux informatiques permettant la visualisation et l'échange, à travers les réseaux de télécommunication, de données stockées dans des ordinateurs. Ce rapport remis au Président de la République Valéry Giscard d'Estaing en décembre 1977 par Simon Nora et Alain Minc a inventé le concept de télématique et proposait le lancement du réseau Minitel, exactement quinze ans après qu'un chercheur du Massachusetts Institute of Technology (MIT) ait rédigé les premiers textes décrivant les interactions sociales rendues possibles par l'intermédiaire d'un réseau d'ordinateurs<sup>1</sup>. La popularisation d'Internet dans les années 1990 a éclipsé la télématique, mais les inspirations des deux projets étaient proches. **La stratégie nationale pour l'intelligence artificielle ne devra pas se tromper de cible, mais bien définir des objectifs réalistes garantissant des résultats effectifs.**

---

<sup>1</sup> En juillet 1962, Joseph Carl Robnett Licklider entamait en effet cette réflexion en vue de faciliter les communications entre chercheurs de la Defense Advanced Research Projects Agency (DARPA) du ministère américain de la Défense. Quelques mois plus tard, en octobre 1962, il devint le premier chef du programme de recherche en informatique de la DARPA. La rigueur scientifique exige de se référer également aux travaux de Leonard Kleinrock, lui aussi chercheur au MIT, et qui publia en 1961 le premier texte théorique sur la « commutation de paquets ».

---

Ni quête vaine, ni projet de remplacement de l'homme par la machine, **l'intelligence artificielle représente une chance à saisir pour nos sociétés et nos économies. La France doit relever ce défi.**

Les **progrès** en intelligence artificielle sont d'abord et avant tout **bénéfiques**. Ils comportent aussi des **risques**, qu'il serait malhonnête de nier. Mais ces risques peuvent et doivent être **identifiés, anticipés et maîtrisés**.

L'avènement d'une **super-intelligence** ne fait pas partie de ces risques à court et moyen terme. A **long terme, la réalité de cette menace n'est pas certaine**. Quant à son **imminence à court ou moyen terme**, prophétisée par plusieurs figures médiatiques, elle **relève du pur fantasme** aux yeux de vos rapporteurs. Le présent rapport se veut une première contribution à un travail indispensable d'identification, d'anticipation et de maîtrise des risques réels. Ce travail de démystification et d'objectivation doit être **collectif, interdisciplinaire et international**.

Afin de prévenir de futures désillusions, il est nécessaire d'assurer un **suivi continu de ces technologies et de leurs usages**, en sachant que les cycles d'espoirs et de déceptions qui jalonnent l'histoire de l'intelligence artificielle invitent à ne pas avoir d'attentes irréalistes à l'égard de ces technologies dans un avenir proche.

Les propositions du présent rapport veulent aller dans ce sens.

Nous nous prononçons, au final, **pour une intelligence artificielle maîtrisée, utile et démystifiée**: maîtrisée, parce que ces technologies devront être les plus sûres, les plus transparentes et les plus justes possibles ; utile parce qu'elles doivent, dans le respect des valeurs humanistes, profiter à tous au terme d'un large débat public ; démystifiée, enfin, parce que les difficultés d'acceptabilité sociale de l'intelligence artificielle résultent largement de visions catastrophistes sans fondement.

Plutôt qu'une hypothétique confrontation dans le futur entre les hommes et les machines, qui relève de la science-fiction dystopique, nous croyons au bel avenir de la complémentarité homme-machine. La conviction de vos rapporteurs est que nous allons bien **plus vers une intelligence humaine augmentée que vers une intelligence artificielle concurrençant l'homme**.

## **ANNEXES**

### **ANNEXE 1 : SAISINE DE L'OFFICE**

Lettre de saisine de l'Office par la commission des affaires économiques du Sénat.

R É P U B L I Q U E F R A N Ç A I S E



COMMISSION  
DES AFFAIRES  
ÉCONOMIQUES

LE PRÉSIDENT

Monsieur Bruno SIDO  
Premier Vice-Président de l'Office  
Parlementaire d'évaluation des choix  
scientifiques et technologiques  
Sénateur de la Haute-Marne  
Casier de la Poste  
Palais du Luxembourg

Paris, le 29 février 2016

Réf: 1330

Monsieur le Président,

Le concept d'intelligence artificielle s'est développé après la Seconde guerre Mondiale avec l'apparition des premiers ordinateurs et des logiciels à-même de les faire fonctionner.

Il connaît actuellement une évolution très rapide qui touche tous les domaines de la vie économique et sociale (production industrielle, médecine, construction, vie en société...) bien au-delà de la seule bureautique professionnelle ou personnelle qui en forme toujours l'image principale.

La crainte que des machines « intelligentes » puissent un jour échapper à leurs concepteurs pour provoquer risques et dangers est inhérente à l'intelligence artificielle et pose le problème, jamais résolu, du point au-delà duquel la complexité n'est plus maîtrisable.

Les enjeux économiques mais aussi sociétaux de ce sujet sont considérables. C'est pourquoi il paraît utile à la Commission des affaires économiques que l'Office parlementaire d'évaluation des choix scientifiques et technologiques éclaire la représentation nationale sur les choix des politiques publiques à opérer en ce domaine.

Je vous prie d'agréer, Monsieur le Président, l'expression de ma considération distinguée.

Jean Claude LENOIR

---

## ANNEXE 2 : RÉUNION DE L'OPECST DU 14 MARS 2017 : ADOPTION DU RAPPORT

**M. Jean-Yves Le Déaut, député, président.** – Nous aurons en cette fin de la quatorzième législature examiné cinq rapports au cours des deux mois de février et mars. Le 22 février dernier, nous avons adopté le rapport d'évaluation de la stratégie nationale de recherche (SNR), ainsi que son volet spécifique sur l'énergie. Aujourd'hui même, au déjeuner de l'Association nationale de la recherche et de la technologie (ANRT), le ministre de l'Enseignement supérieur et de la recherche, devant le Premier ministre, m'a remercié de la qualité de ce travail – le Comité opérationnel (ComOp), comme diverses autres instances du ministère, examinera notre rapport – et m'a indiqué que nos analyses, pourtant sévères, seraient une aide dans les arbitrages du ministère, si bien que nos critiques feront avancer les choses. Le 8 mars, nous avons examiné le rapport d'évaluation du quatrième plan national de gestion des matières et déchets radioactifs ; enfin, notre collègue Catherine Procaccia et moi rédigeons actuellement le rapport sur les nouvelles biotechnologies.

Merci à nos deux rapporteurs pour leur travail d'une grande qualité, sur un sujet majeur. La convergence technologique – nanotechnologies, biologie, informatique, sciences cognitives (NBIC) – a des conséquences fortes, avec la numérisation et la robotisation de la société. L'interface entre l'homme et la machine est transformé, et de l'homme soigné, réparé par la machine, on envisage à présent l'homme augmenté ! Cela ne va pas sans susciter des interrogations pour les droits de l'homme. Je rédige actuellement un rapport pour l'Assemblée parlementaire du Conseil de l'Europe sur ce sujet. Il est possible d'espionner, grâce à l'informatique et à la puissance de calcul, tout individu et tout comportement de la vie en société. Cela pose question...

Je souhaite rendre hommage à Jean-Claude Etienne, qui fut député, puis sénateur, et notre vice-président de l'OPECST. Il s'est éteint samedi dernier à l'âge de 75 ans. Il était venu encore l'an dernier nous rendre visite, à l'occasion du trentième anniversaire de notre Office. Il en fut un membre éminent, en plus d'être professeur agrégé de rhumatologie et professeur à l'université de Reims.

**M. Bruno Sido, sénateur, vice-président.** – Je m'associe à cet hommage, d'autant que je fus son vice-président au conseil général de mon département et que je le connaissais bien. J'ai eu plaisir à travailler avec lui. Son parcours fut exceptionnel : d'abord agrégé de mathématiques, il suivit ensuite les traces de son frère médecin pour pouvoir – aimait-il à dire – acheter lui aussi une Peugeot 203... C'est ainsi qu'il devint médecin. Humaniste, il voyait dans la politique un moyen, comme la médecine, d'accompagner l'homme dans ses difficultés et de trouver des solutions.

---

J'ai eu l'honneur de lui succéder ici comme vice-président, comme il me l'avait proposé. Ses obsèques auront lieu jeudi.

*(Mmes et MM les parlementaires se lèvent et observent une minute de silence.)*

**Mme Dominique Gillot, sénatrice, rapporteure.** – Monsieur le président a mentionné nos critiques sur la SNR. Nous n'avons certes pas été complaisants sur sa mise en œuvre. Il est vrai aussi que le pilotage est difficile en pareille phase de mutation, quand ce qui a été décidé il y a trois ans est déjà à revoir... D'où l'intérêt des focus que l'OPECST publie.

Je m'associe bien sûr aux propos tenus sur Jean-Claude Etienne : agrégés de sciences ou non, nous cherchons, comme parlementaires, à comprendre le monde et à éclairer nos décisions. Ce qui m'amène naturellement à la présentation de notre rapport...

L'OPECST a été saisi le 29 février 2016, par la commission des affaires économiques du Sénat, d'une étude sur l'intelligence artificielle (IA). Nous sommes fiers d'en avoir été, M. Claude de Ganay et moi-même, les rapporteurs. Un bouleversement pourrait transformer profondément nos sociétés : les technologies d'intelligence artificielle. Elles pourront apporter dans notre futur des progrès dans de nombreux domaines, or elles ne font pas l'objet d'une analyse sereine et objective. L'intelligence artificielle suscite en effet enthousiasme, espoir et intérêt mais aussi méfiance, incrédulité ou oppositions.

L'irruption de l'intelligence artificielle au cœur du débat public remonte à un peu plus de deux ans, après la diffusion d'une lettre d'avertissement sur les dangers potentiels de l'intelligence artificielle, publiée en janvier 2015, qui a recueilli plus de 5 000 signatures en un an. Elle a été lancée pour alerter l'opinion publique et insister sur l'urgence de définir des règles éthiques, afin de cadrer la recherche.

Aucun argument sérieux ne venait étayer cette première mise en garde quant au risque présumé de dérive malveillante ! Pourtant, cette alerte a contribué à renforcer les peurs et les angoisses face aux technologies d'intelligence artificielle.

Notons que 2016 a fait figure d'année de l'intelligence artificielle : chaire d'informatique du Collège de France attribuée à Yann LeCun, victoire du système d'intelligence artificielle AlphaGo créé par DeepMind sur le champion de go, Lee Sedol, et ainsi de suite, tout au long de l'année : les initiatives en matière d'intelligence artificielle se sont multipliées à un rythme effréné. Impossible d'en faire l'inventaire !

Après l'irruption de l'intelligence artificielle dans le débat public en 2015, 2016 et le premier trimestre 2017 ont été jalonnés de nombreux rapports sur l'intelligence artificielle, émanant du Parlement européen, de la Maison blanche, de la Chambre des communes, de l'Association mondiale des ingénieurs électriciens et électroniciens, de la Commission de réflexion sur l'éthique de la recherche en sciences et technologies du numérique de

l'alliance du numérique (CERNA), de l'INRIA, de l'Institut Mines-Télécom, du Club informatique des grandes entreprises françaises (CIGREF), du Syndicat des machines et technologies de production (SYMOP), de l'Association française pour l'intelligence artificielle (AFIA), de l'Association française contre l'intelligence artificielle (AFCIA), etc. Des conférences d'envergure nationale ou internationale ont aussi été organisées sur le sujet par les Nations Unies, l'OCDE, la Fondation pour le futur de la vie, le Medef, l'AFIA entre autres. Enfin, l'initiative « France IA » lancée par le Gouvernement en janvier 2017 s'est accompagnée de l'annonce d'un plan national pour l'intelligence artificielle, dont nous attendons le détail d'ici la fin du mois.

Devant cet emballement, alors que les progrès se font à une vitesse exponentielle et reposent de plus en plus sur un financement privé aux moyens considérables, il est indispensable que la réflexion soit conduite de manière sereine et rationnelle, afin de mettre en avant la réalité des connaissances, les opportunités tout autant que les risques, afin aussi de rassurer le public et de démystifier les représentations biaisées. Comme le disait Marie Curie, « *dans la vie, rien n'est à craindre, tout est à comprendre* ».

Les progrès en intelligence artificielle posent des questions auxquelles toute la société doit être sensibilisée : quels sont les opportunités et les risques qui se dessinent ? La France et l'Europe sont-elles dans une position satisfaisante dans la course mondiale ? Quelles places respectives pour la recherche publique et la recherche privée ? Quelle coopération entre celles-ci ? Quelles priorités pour les investissements dans la recherche en intelligence artificielle ? Quels principes éthiques, juridiques et politiques doivent encadrer ces technologies ? La régulation doit-elle se placer au niveau national, européen ou international ?

Le débat public ne peut pas s'engager sereinement dans l'ignorance des technologies mises en œuvre, des méthodes scientifiques et des principes de l'intelligence artificielle. Nous avons donc entendu faire l'état de la recherche et des usages des technologies d'intelligence artificielle, en constante évolution. Nous nous sommes interrogés sur la façon d'assurer le respect de règles éthiques dans la recherche en IA et au-delà, parce que « *science sans conscience n'est que ruine de l'âme* », ainsi que l'affirmait Rabelais.

Suite à l'adoption de l'étude de faisabilité le 28 juin 2016, nos auditions et déplacements ont commencé en septembre 2016 ; tous deux renouvelables, nous avons dû interrompre nos investigations le mois dernier. Soit une période utile d'environ six mois : nous avons donc dû préciser un champ d'investigations, en ayant le souci d'optimiser la plus-value relative du rapport, répondre à la saisine de la commission des affaires économiques du Sénat et faire mieux connaître l'intelligence artificielle. Les enjeux sont tout autant scientifiques et technologiques que politiques, philosophiques, éthiques, juridiques, éducatifs, médicaux, militaires ou, encore, économiques. Nous avons dû choisir.



---

Les aspects scientifiques et technologiques constituant le cœur de métier de l'OPECST, c'est la recherche publique et privée en intelligence artificielle qui a été retenue, tout comme les enjeux philosophiques, éthiques, politiques, juridiques et éducatifs, car ils soulèvent des questions essentielles - y répondre devrait aider à dépasser les peurs et les inquiétudes pour engager un débat public plus serein et mieux étayé.

Les enjeux financiers, économiques et industriels n'ont pas été écartés, mais sont mis au second plan car ils correspondent moins directement à la plus-value spécifique de l'OPECST. Enfin, les usages de l'intelligence artificielle pour la défense, les technologies militaires et la médecine ont été écartés.

Nous avons mis l'accent sur les enjeux éthiques, car ils permettent d'aborder les sujets de manière transversale. La méthode de travail a été fondée sur des auditions et des déplacements en France et à l'étranger, présentés en annexe du rapport. Nous avons aussi eu une journée de tables rondes.

Le rapport contient une histoire et même une « préhistoire » assez détaillée de l'intelligence artificielle et des technologies rattachées. L'intelligence artificielle a fêté l'année dernière son soixantième anniversaire, puisqu'elle a été inventée en tant que discipline et en tant que concept en 1956 lors d'une école d'été à Dartmouth. La conférence affirme que « *chaque aspect de l'apprentissage ou toute autre caractéristique de l'intelligence peut être si précisément décrit qu'une machine peut être conçue pour le simuler* ». Le projet n'est pas de construire une machine rivalisant avec l'homme mais de simuler telle ou telle tâche que l'on réserve à l'intelligence humaine. Face à l'emballement des prises de position médiatisées, il n'est pas inutile de le rappeler...

Le concept a fait l'objet d'un débat. Le choix du nom a sans doute été motivé par une quête de visibilité de ce nouveau champ de recherche. « Intelligence artificielle » a pu apparaître plus séduisant que « sciences et technologies du traitement de l'information ». mais l'anthropomorphisme essentialiste qui s'est exprimé dans ce choix n'a sans doute pas contribué à apaiser les peurs suscitées par le projet prométhéen de construction d'une machine rivalisant avec l'intelligence humaine.

L'intelligence artificielle repose sur l'utilisation d'algorithmes, suites finies et non ambiguës d'opérations ou d'instructions permettant, à l'aide d'entrées, de résoudre un problème ou d'obtenir un résultat, ces sorties étant réalisées selon un certain rendement. Les algorithmes peuvent, en effet, servir à calculer, à gérer des informations, à analyser des données, à communiquer, à commander un robot, à fabriquer des biens ou, encore, à modéliser et simuler - comme le font certains outils de météorologie, de sismologie, d'océanographie, de planétologie, d'urbanisme,...

L'informatique traite plutôt de questions résolues par des algorithmes connus, alors que l'on applique le label d'« intelligence

artificielle » à des applications permettant plutôt de résoudre des problèmes moins évidents pour lesquels aucun algorithme satisfaisant n'existe encore.

Le paradoxe résultant de cette définition est le suivant : dès que le problème a été résolu par une technologie dite d'intelligence artificielle, l'activité correspondante n'est plus considérée comme une preuve d'intelligence de la machine. Les cas connus de résolutions de problèmes d'algèbre ou de capacité à jouer à des jeux (des jeux d'échecs ou de Go par exemple) illustrent ce phénomène. Nick Bostrom explique ainsi que « *beaucoup d'intelligence artificielle de pointe a filtré dans des applications générales, sans y être officiellement rattachée car dès que quelque chose devient suffisamment utile et commun, on lui retire l'étiquette d'intelligence artificielle* ».

Les progrès en matière d'intelligence artificielle étant tangibles depuis les années cinquante, les frontières de l'intelligence artificielle sont donc sans cesse repoussées et ce qui était appelé intelligence artificielle hier n'est donc plus nécessairement considéré comme tel aujourd'hui.

Dès l'origine, l'intelligence artificielle est une étiquette. Ce label recouvre en réalité des technologies diverses, qui traduisent la variété des formes d'intelligence en général : elles vont de formes explicites (systèmes experts et raisonnements logiques et symboliques) à des formes plus implicites (réseaux bayésiens et surtout réseaux de neurones et *deep learning*). Nous avons voulu retracer dans le rapport, de manière inédite, la richesse et la diversité de ces technologies.

De manière caricaturale, on pourrait résumer les technologies d'intelligence artificielle à un champ de recherche où cohabitent deux grands types d'approches : les approches symboliques et les approches connexionnistes.

Nous notons que « l'âge d'or de l'IA » qui court de 1956 au début des années soixante-dix, est marqué par les approches symboliques et les raisonnements logiques, qui sont de nombreux types et sont tous décrits dans le rapport. Cet âge d'or a été suivi d'un premier « hiver de l'intelligence artificielle » dans la décennie soixante-dix : les financements sont revus à la baisse, suite à divers rapports assez critiques, les prédictions exagérément optimistes des débuts ne se réalisant pas et les techniques ne fonctionnant que dans des cas simples.

Ce constat témoigne du caractère cyclique des investissements en intelligence artificielle selon une boucle « espoirs-déceptions », et l'enthousiasme se renouvelle dans les années quatre-vingt autour des systèmes experts, de leurs usages et de l'ingénierie des connaissances. Suit un nouvel hiver de l'intelligence artificielle dans les années quatre-vingt-dix.

Pour autant, des découvertes scientifiques sont faites dans la période. Après la renaissance de l'intérêt pour les réseaux de neurones artificiels avec de nouveaux modèles théoriques de calculs, les années quatre-vingt-dix voient se développer la programmation génétique ainsi que les systèmes multi-agents ou l'intelligence artificielle distribuée.

---

De très nombreux autres domaines et technologies d'intelligence artificielle peuvent être ajoutés à ceux déjà mentionnés : les machines à vecteur de support (SVM), l'apprentissage machine dont l'apprentissage par renforcement, la programmation par contraintes, les raisonnements à partir de cas, les logiques de description, les algorithmes génétiques, la recherche dans les espaces d'états, la planification, les ontologies, les logiques de description... tous ces exemples analysés de manière détaillée dans le rapport visent à illustrer la variété et la richesse qui se cache derrière le label « intelligence artificielle » : les technologies d'intelligence artificielle sont en fait quasi-innombrables ; surtout, les chercheurs, tels des artisans, hybrident des solutions inédites au cas par cas, en fonction de leur tour de main personnel.

Le tableau académique international des domaines de l'intelligence artificielle retient cinq domaines : traitement du langage naturel, vision, apprentissage automatique, systèmes multi-agents, robotique. Nous renvoyons au rapport pour plus de détails. C'est une histoire passionnante !

Faisons un focus sur l'apprentissage machine, au cœur des débats actuels. La difficulté liée aux algorithmes classiques réside dans le fait que l'ensemble des comportements possibles d'un système, compte tenu de toutes les entrées possibles, devient rapidement trop complexe à décrire. Cette explosion combinatoire justifie de confier à des programmes le soin d'ajuster un modèle adaptatif permettant de gérer cette complexité et de l'utiliser de manière opérationnelle en prenant en compte l'évolution de la base des informations pour lesquelles les comportements en réponse ont été validés. C'est ce que l'on appelle l'apprentissage automatique ou *machine learning*, qui permet d'apprendre et d'améliorer le système d'analyse ou de réponse. En ce sens, on peut dire que ces types particuliers d'algorithmes apprennent.

Un apprentissage est dit « supervisé » lorsque le réseau est forcé à converger vers un état final précis, en même temps qu'un motif lui est présenté. À l'inverse, lors d'un apprentissage « non-supervisé », le réseau est laissé libre de converger vers n'importe quel état final lorsqu'un motif ou un élément lui est présenté.

Entre ces deux extrêmes, l'apprentissage automatique ou *machine learning* peut être semi-supervisé ou partiellement supervisé. C'est le cas dans de nombreuses applications.

L'apprentissage automatique peut lui-même reposer sur plusieurs méthodes : l'apprentissage par renforcement, l'apprentissage par transfert, ou, encore, l'apprentissage profond, qui est le plus en pointe aujourd'hui. Le « *deep learning* » rencontre un succès particulièrement remarquables dans la présente décennie. Pourtant cette méthode est ancienne. Son essor doit beaucoup à l'émergence récente de données massives ou *big data*, et à l'accélération de la vitesse de calcul des processeurs, mais son histoire remonte aux années quarante : les « réseaux de neurones artificiels » sont imaginés dès cette époque.

---

Un réseau de neurones artificiels est la modélisation d'un ensemble d'éléments interconnectés, chacun ayant des entrées et des sorties numériques. Le comportement d'un neurone artificiel dépend de la somme pondérée de ses valeurs d'entrée. Si cette somme dépasse un certain seuil, la sortie prend une valeur positive, sinon elle reste nulle. Un réseau peut comporter une couche d'entrée (les données), une de sortie (les résultats), et une ou plusieurs couches intermédiaires.

Cet apprentissage permet d'ajuster les poids synaptiques afin que les correspondances entre les entrées et les sorties soient les meilleures possible. Il s'agit donc de combiner de nombreuses fonctions simples pour former des fonctions complexes et d'apprendre les liens entre ces fonctions simples à partir d'exemples étiquetés.

Il ne s'agit en aucun cas de réseaux de neurones de synthèse, ce n'est qu'une image, sans doute malheureuse car elle entretient une forme de confusion, en lien avec la notion d'intelligence artificielle. L'analogie avec le fonctionnement du cerveau humain repose sur le fait que les fonctions simples rappellent le rôle joué par les neurones, tandis que les connexions rappellent les synapses. Certains chercheurs préfèrent ainsi parler de neurones électroniques et de synapses électroniques.

Outre les réseaux multi-couches, d'importantes découvertes en apprentissage profond remontent aux années quatre-vingt, telles que la rétropropagation du gradient. L'idée générale de la rétropropagation consiste à rétropropager l'erreur commise par un neurone à ses synapses et aux neurones qui y sont reliés. Il s'agit en effet de faire converger l'algorithme de manière itérative vers une configuration optimisée des poids synaptiques.

En apprentissage profond, qui repose donc sur des réseaux de neurones profonds (*deep neural networks*), les réseaux de neurones artificiels peuvent donc être à apprentissage supervisé ou non (ils sont le plus souvent supervisés, comme dans le cas du Perceptron), avec ou sans rétropropagation (*back propagation*) et on peut distinguer les technologies selon la manière particulière d'organiser les neurones en réseau : les réseaux peuvent être en couches, tel les architectures profondes ou multicouches (plusieurs dizaines ou centaines de couches), dans lesquels chaque neurone d'une couche est connecté à tous les neurones de la couche précédente et de la couche suivante (c'est la structure la plus fréquente) ; les réseaux peuvent être totalement interconnectés (« réseaux de Hopfield » et « machines de Boltzmann ») ; les réseaux peuvent permettre de prendre en compte le contexte tel une mémoire, avec le cas des réseaux neuronaux récurrents ; enfin, les réseaux peuvent se chevaucher, un peu comme dans le calcul matriciel, à l'instar des réseaux neuronaux à convolution.

Nous ne disposons d'aucune explication théorique des raisons pour lesquelles les réseaux de neurones fonctionnent aussi bien, c'est-à-dire donnent, dans un certain nombre de domaines, d'excellents résultats. La

---

technologie devance donc la science en la matière : c'est à la recherche d'éclaircir ce sujet...

Les technologies disponibles en intelligence artificielle peuvent se combiner entre elles : les combinaisons et les hybridations sont quasi-systématiques, le programme AlphaGo de Google-DeepMind a ainsi appris à jouer au jeu de go par une méthode de *deep learning* couplée à un apprentissage par renforcement et à une optimisation selon la méthode Monte-Carlo, qui repose sur le hasard.

De plus en plus, les outils d'intelligence artificielle sont utilisés conjointement. Par exemple, les systèmes experts sont utilisés avec le raisonnement par analogie, éventuellement dans le cadre de systèmes multi-agents. De même, les SVM et l'apprentissage par renforcement se combinent très efficacement avec l'apprentissage profond des réseaux de neurones. Le *deep learning*, peut aussi s'enrichir de logiques floues ou d'algorithmes génétiques.

Derrière le concept d'intelligence artificielle, ce sont des technologies très variées qui donnent lieu à des applications spécifiques pour des tâches toujours très spécialisées. Les applications sectorielles présentes ou futures sont d'envergure considérable, que l'on pense par exemple aux transports, à l'aéronautique, à l'énergie, à l'environnement, à l'agriculture, au commerce, à la finance, à la défense, à la sécurité, à la sécurité informatique, à la communication, à l'éducation, aux loisirs, à la santé, à la dépendance ou au handicap.

Il s'agit d'autant de jalons d'applications sectorielles, dont le rapport retrace les possibilités, nous y renvoyons donc. Le potentiel de ces technologies est immense et ouvre de manière transversale un espace d'opportunités inédit : nos économies peuvent en bénéficier car les champs d'application sont et seront de plus en plus nombreux. Ces technologies sont non seulement en évolution constante, mais leurs combinaisons ouvrent de nouvelles perspectives.

Selon Stéphane Mallat, professeur à l'École normale supérieure, il s'agit d'« une rupture non seulement technologique, mais aussi scientifique ». Traditionnellement, les modèles sont construits par les chercheurs eux-mêmes à partir de données d'observation, en n'utilisant guère plus de dix variables alors que « les algorithmes d'apprentissage sélectionnent seuls le modèle optimal pour décrire un phénomène à partir d'une masse de données » et avec une complexité inatteignable pour nos cerveaux humains, puisque cela peut représenter jusqu'à plusieurs millions de variables, contre une dizaine pour un laboratoire humain. Alors que le principe de base de la méthode scientifique réside dans le fait que les modèles ou les théories sont classiquement construits par les chercheurs à partir des observations, le *deep learning* change la donne en assistant et amplifiant l'expertise scientifique dans la construction des modèles.

Denis Girou, directeur de l'Institut du développement et des ressources en informatique scientifique au CNRS estime que « la science a pu

construire des modèles de plus en plus complexes grâce à l'augmentation de la puissance de calcul des outils informatiques, au point que la simulation numérique est désormais considérée comme le troisième pilier de la science après la théorie et l'expérience ».

Selon Yann LeCun, le défi scientifique auquel les chercheurs doivent s'atteler c'est celui de l'apprentissage non supervisé. Dans sa leçon inaugurale au Collège de France, il estime ainsi que *« tant que le problème de l'apprentissage non-supervisé ne sera pas résolu, nous n'aurons pas de machines vraiment intelligentes. C'est une question fondamentale scientifique et mathématique, pas une question de technologie. Résoudre ce problème pourra prendre de nombreuses années ou plusieurs décennies. À la vérité, nous n'en savons rien »*.

L'intelligence artificielle, qui agit sur la base de ce qu'elle sait, devra donc relever le défi d'agir sans savoir, puisque comme l'affirmait le biologiste, psychologue et épistémologue Jean Piaget *« L'intelligence, ça n'est pas ce que l'on sait, mais ce que l'on fait quand on ne sait pas »*. J'y insiste, ce que sait l'intelligence artificielle, c'est l'homme qui le lui a appris.

**M. Claude de Ganay, député, rapporteur.** – Je vais vous parler quant à moi des caractéristiques et des enjeux de la recherche en l'intelligence artificielle.

La recherche privée tient une place prépondérante, y compris sur le plan de la recherche fondamentale. Cette recherche est dominée aujourd'hui par les entreprises américaines et peut-être, demain, par les entreprises chinoises. Des enseignants-chercheurs parmi les plus brillants ont été recrutés par ces grandes entreprises : Yann LeCun (Facebook), Andrew Ng (Baidu, après Google), Geoffrey Hinton (Google), Fei Fei Li (Google), Rob Fergus (Facebook), Nando de Freitas (Google)...

Les entreprises américaines dominant donc, mais la recherche et les entreprises chinoises montent en puissance. La Chine a ainsi pris la tête des publications en *deep learning* depuis trois ans. L'entreprise Baidu a développé le principal moteur de recherche chinois, site le plus consulté en Chine et le cinquième plus consulté au niveau mondial : indexant près d'un milliard de pages, l'entreprise dispose d'un flux de données permettant d'envisager des applications dans de nombreux domaines. Ses résultats algorithmiques sont impressionnants, malgré son existence récente. Le système de reconnaissance d'image de Baidu a ainsi battu celui de Google depuis 2015. Le recrutement du chercheur de Stanford Andrew Ng par Baidu en 2014 en tant que responsable de l'intelligence artificielle alors qu'il en était le responsable chez Google est emblématique. De même, en 2017, Baidu débauche Qi Lu, au poste de numéro deux, alors qu'il était auparavant vice-président chez Microsoft et directeur des projets Bing, Skype et Microsoft Office et auparavant directeur de la recherche de Yahoo.

Le 13<sup>ème</sup> plan quinquennal chinois comprend une liste de quinze « grands projets » qui structurent les priorités scientifiques avec des investissements de plusieurs milliards d'euros. Ce plan vise à dynamiser la

---

recherche chinoise en IA et à concurrencer les États-Unis. Parmi ces projets, ceux en lien avec l'IA représentent un montant de 100 milliards de yuans sur trois ans.

Autre caractéristique, l'interdisciplinarité est indispensable en intelligence artificielle, alors que la discipline demeure éclatée en une cinquantaine de sous-domaines de recherche, un tableau les décrit dans notre rapport.

Par ailleurs, la recherche en intelligence artificielle est soumise à une contrainte d'acceptabilité sociale assez forte, notamment sous l'effet de représentations catastrophistes, comme en témoignent différents sondages d'opinion, eux aussi rappelés dans le rapport.

Plusieurs interventions médiatiques et pétitions ont cherché en 2015 à interpeler l'opinion à propos des risques qui seraient inhérents à l'intelligence artificielle. L'existence d'une association française contre l'intelligence artificielle (AFCIA) est révélatrice d'un certain climat d'angoisse puisque la France serait le seul pays où une telle association existerait. L'AFCIA juge « illégitime et dangereuse la recherche scientifique visant à créer des organismes à intelligence artificielle supra-humaine » et considère que le seul moyen « d'éviter un avenir funeste pour l'humanité est d'obtenir l'interdiction légale de la recherche en intelligence artificielle à l'échelle mondiale ». Se définissant comme association de lobbying, elle vise à obtenir cette interdiction auprès des pouvoirs publics. Jacques Attali s'est, à la fin de l'année 2016, prononcé pour un moratoire sur les technologies d'intelligence artificielle, ce qui nous a beaucoup surpris.

Dernière caractéristique : la multiplication des initiatives visant la prise en compte de principes éthiques dans la recherche et les usages de l'intelligence artificielle. Cela vaut pour la recherche publique, comme pour la recherche privée, en Europe comme en Amérique.

Concernant la recherche française en IA, notre pays dispose d'importants atouts à faire valoir, même si la communauté française de l'intelligence artificielle est encore insuffisamment organisée, connue et visible. La reconnaissance internationale des travaux des chercheurs français doit beaucoup à des organismes comme l'INRIA, le CNRS, le CEA, différentes universités et grandes écoles, par exemple l'ENS et Mines-Télécom, qui produisent des travaux à visibilité internationale. Nous décrivons dans le détail ces structures et leurs laboratoires, à l'excellence reconnue.

La France dispose d'un réseau de chercheurs très compétent et d'un tissu de *start-up* très dynamique : 240 d'entre elles sont spécialisées en intelligence artificielle. Ce tissu de *start-up*, encouragé par l'initiative FrenchTech, est très riche. Selon l'investisseur en IA Paul Strachman, « *La France est l'un des écosystèmes les plus vibrants en ce qui concerne l'intelligence artificielle. Malheureusement, cela n'est pas très su en-dehors de la France. Et parfois même en-dedans* ».



---

Pour Mark Zuckerberg, le président de Facebook, « *la France dispose de l'une des communautés de chercheurs en intelligence artificielle la plus forte du monde* ». De même Mike Schroepfer, le directeur technique de Facebook, estimait en 2015 que Paris avait « *la plus grande concentration de toute l'Europe en matière d'intelligence artificielle* ». Facebook a préféré Paris à Londres pour ouvrir en juin 2015 un laboratoire de recherche consacré à l'intelligence artificielle ; ces recherches sont pilotées par Yann LeCun au niveau mondial.

Le bon niveau de nos étudiants a également souvent été cité. Mais nous nous inquiétons d'un phénomène de rachat de *start-up* et de fuite des cerveaux voire de pillage de nos talents, lié aux conditions attractives offertes à l'étranger. Lors de son audition, Stéphane Mallat a fait valoir que depuis plusieurs années la totalité des étudiants issus des masters spécialisés de l'ENS quittaient la France aussitôt leur formation achevée. Il faut permettre à ces jeunes génies, qui sont autant de chercheurs et d'entrepreneurs en devenir, de disposer d'opportunités en France et permettre aux *start-ups* de se développer sans être rachetées par les géants américains, chinois ou japonais dès qu'elles présentent un profil viable.

Par ailleurs, nous relevons que la communauté française de l'intelligence artificielle se constitue surtout en-dehors des institutions, à travers les *meetups*. Le principal d'entre eux, le *Paris machine learning meetup* regroupe 5 205 membres. Quant à l'AFIA, comprenant environ 300 membres, elle semble assez fermée sur elle-même. Elle aurait tout intérêt à transcender ses propres limites pour relever le défi d'une intelligence artificielle française ouverte, visible et conquérante.

Au total, on voit une sous-estimation des atouts considérables de la France, mais il existe un risque de « décrochage » par rapport à la recherche internationale la plus avancée en intelligence artificielle.

Sur les impacts sociaux et économiques potentiels de l'IA, et les enjeux liés à ces questions, nous avons perçu les signes avant-coureurs de l'évolution vers une économie globalisée de « plateformes ».

On parle des « GAFA », parfois des « GAFAMI », mais il serait plus juste de parler des « GAFAMITIS » (Google, Apple, Facebook, Amazon, Microsoft, IBM, Twitter, Intel et Salesforce), des « NATU » (Netflix, Airbnb, Tesla et Uber) et des « BATX » (l'expression désignant les géants chinois du numérique, Baidu, Alibaba, Tencent et Xiaomi). Ces exemples emblématiques des bouleversements en cours sont les prémises de la place dominante et monopolistique occupée par quelques entreprises dans ce futur contexte général. Chacune de ces entreprises est entrée, selon un modèle « *the winner takes it all* » (« le vainqueur prend tout »), dans une course pour acquérir une position de pointe dans les technologies d'intelligence artificielle afin de tirer profit de la position dominante qui en résultera : l'accroissement significatif des investissements dans la recherche en intelligence artificielle pourrait bien conduire à une concentration horizontale progressive des grandes entreprises, voire au monopole de ces plateformes dominant une économie globalisée. On assiste à une montée en

---

puissance significative dans les acquisitions, un tableau les décrivant figure dans le rapport.

S'agissant des bouleversements annoncés dans le marché du travail, les pronostics sont très contrastés, allant de 9 à 47 % de disparition d'emplois. Pour le Conseil d'orientation pour l'emploi, moins de 10 % des emplois existants français apparaissent menacés par l'automatisation et la numérisation et la moitié des emplois existants est susceptible d'évoluer de façon significative. Nous pensons, quant à nous, que les études sous-estiment les évolutions de contenu des métiers et les créations d'emplois. Le solde global reste inconnu mais nous avons la conviction d'une future coopération hommes-machines heureuse.

L'éducation peut être le levier et le bénéficiaire des avancées en intelligence artificielle. La relation émetteur-récepteur est transformée et modifie tant la pédagogie que les principes d'évaluation. Les moyens de prédire la réussite des élèves et d'optimiser les enseignements seront précisés par les systèmes d'intelligence artificielle qui permettront la différenciation des méthodes et des contenus enseignés, la personnalisation devant être adaptée à la diversité des élèves. Les nouvelles technologies ne seront pas en compétition avec les enseignants, elles leur seront complémentaires. Les cours en ligne ouverts et massifs, ou MOOC (*massive open online course*) seront, de ce point de vue, des ressources utilisables pour appliquer ces nouvelles méthodes pédagogiques innovantes et permettre aux jeunes générations d'accéder dans des conditions optimales à la connaissance.

Nous sommes convaincus de la possibilité imminente d'une révolution bénéfique de notre cadre de vie et de l'aide aux personnes. Des changements profonds sont à venir dans la connaissance et dans le contrôle de notre environnement et de la santé des populations. Les *smart grids*, systèmes d'économie d'énergie par une consommation optimisée, et les *smart cities* (villes intelligentes) seront les expressions des bénéfices que nous pouvons tirer de l'intelligence artificielle. Et cela se traduira évidemment en matière de transports, de sécurité, de santé, de dépendance et de handicap. Notre cadre de vie, la qualité de nos vies seront améliorés par l'usage massif de technologies d'intelligence artificielle.

J'ai été surpris par des propos d'un de mes concitoyens, grand-maître de la confrérie « bérouttes et traditions » de Cernoy-en-Berry, qui m'a expliqué tous les bienfaits que les robots et les systèmes d'intelligence artificielle pourraient avoir pour la ruralité, en particulier pour les personnes âgées, isolées ou dépendantes. Le cas des voitures autonomes a été évoqué mais d'autres applications utiles vont émerger.

En matière de handicap, nous allons vers des progrès majeurs, avec les prothèses intelligentes, des exosquelettes robotisés ou avec des systèmes capables de voir des images et d'en décrire le contenu pour des malvoyants.

Les agents conversationnels ou *bots*, les robots de service, les agents d'assistance, d'aide à la mobilité vont progressivement cohabiter avec nous.

---

Cela nécessitera une grande vigilance. L'éducation et la prévention sont indispensables dans ce contexte de cohabitation croissante. Et il convient d'apporter une grande attention aux logiques d'empathie et aux aspects émotionnels.

J'en viens aux questions éthiques et juridiques posées par les progrès en intelligence artificielle. Nous avons dressé le bilan des initiatives existantes et présenté toutes les propositions qui sont mises sur la table et je renvoie sur ce point à notre rapport. Il y a aussi les deux rapports issus des institutions de l'Union européenne, Parlement européen et Comité économique et social européen (CESE), les trois rapports de la Maison blanche, le rapport de la Chambre des communes du Royaume-Uni, le groupe de travail de la *Royal Society*, les initiatives chinoises et japonaises qui accordent une place contrastée aux questions éthiques... La stratégie du Gouvernement pour l'intelligence artificielle arrive hélas un peu tard pour être intégrée dans les stratégies nationales destinées au monde de la recherche. Les éléments seront communiqués d'ici le 29 mars par le Gouvernement, nous enrichirons notre rapport en conséquence.

Nous invitons à dépasser les « lois d'Asimov » pour faire un point sur le droit de la robotique. Reconnaître une personnalité juridique des robots est une des pistes innovantes qui parcourent le débat public sur la robotique, mais nous ne sommes pas convaincus de l'intérêt de reconnaître une personnalité juridique aux robots, ce sujet n'est pas une question qui mérite d'être posée à ce stade.

S'agissant des autres aspects juridiques de l'intelligence artificielle et de la robotique, il sera loisible de conduire une réflexion et de faire de la prospective concernant la conception, la propriété intellectuelle et l'autorisation de commercialisation. Pour différents spécialistes, il n'y a pas d'urgence à combler un vide juridique béant... car il n'y a pas de vide juridique béant. Les rapports parus sur le sujet, notamment dans le monde anglo-saxon, vont dans le même sens et ne recommandent pas de mesures législatives. La protection des données personnelles et de la vie privée méritera peut-être, en revanche, d'être renforcée dans l'avenir, en s'adaptant aux nouvelles innovations. À ce stade, le droit est suffisamment protecteur.

S'agissant des voitures autonomes nous avons conduit des analyses présentées dans le rapport, le laboratoire « Moral machine » du MIT que nous avons visité travaille notamment sur les dilemmes éthiques. Les résultats provisoires des tests conduisent à identifier différents facteurs de choix : le nombre de tués (on préfère la solution qui réduit le nombre de morts), le fait de sacrifier en priorité des personnes qui transgressent les règles (exemple du voleur), le fait de sacrifier en priorité un animal contre un humain, le fait de sacrifier en priorité une personne plus âgée face à une personne plus jeune et *a fortiori* un enfant, le fait de sacrifier en priorité un homme face à une femme... Ce dernier point, soit dit en passant, mériterait un débat !

---

Sur les régimes de responsabilité, nous notons que quatre régimes pourraient trouver à s'appliquer aux accidents causés par des robots : le régime de responsabilité du fait des produits défectueux, celui de la responsabilité du fait des animaux, celui de la responsabilité du fait d'autrui, ou, encore, celui, traditionnel, de la responsabilité du fait des choses, mais qui ne s'applique que de façon résiduelle par rapport au régime de responsabilité du fait des produits défectueux.

On pourrait envisager de mettre en place une « chaîne de responsabilité », ou responsabilité en cascade. Dans la mesure où trois ou quatre acteurs sont en présence (le producteur de la partie physique du robot, le concepteur de l'intelligence artificielle, l'utilisateur et s'il est distinct de ce dernier, le propriétaire), il est possible d'imaginer que chacun puisse supporter une part de responsabilité selon les circonstances dans lesquelles est survenu le dommage. Il sera en tout cas important d'identifier des pistes d'avenir qui ne fassent pas courir le risque de déresponsabiliser les acteurs du secteur, à commencer par les industriels de la robotique.

En outre, il conviendrait de réfléchir à la possibilité d'instituer des systèmes d'assurance spécifiques. Mais la Fédération française de l'assurance estime qu'il est encore trop tôt pour répondre à la question.

Le rapport présente le cadre national de la réflexion sur les enjeux éthiques de l'intelligence artificielle, avec la CERNA par exemple, qui joue en la matière un rôle majeur, elle a d'ailleurs produit deux rapports, dont nous rendons compte.

Le rôle de la CERNA est réinterrogé suite aux nouvelles missions dévolues à la CNIL. Aux termes de la loi pour une République numérique, la CNIL, notre autorité de contrôle en matière de protection des données personnelles, est chargée de conduire une réflexion sur les questions d'éthique liées au numérique et aux algorithmes. La CNIL a choisi d'y répondre par l'organisation en 2017 d'un cycle de débats intitulé « Les algorithmes en débat ». À l'automne 2017, la CNIL rendra publique la synthèse des échanges afin d'établir une « *cartographie de l'état du débat public* » et un « *panorama des défis et enjeux* ». L'INRIA, à travers le projet « Transalgo », développe en 2017 de manière utile une plateforme d'évaluation de la transparence des algorithmes, afin de répondre aux préoccupations d'explicabilité des algorithmes. L'articulation et la complémentarité entre le travail de la CERNA, de l'INRIA et de la CNIL sont à rechercher.

Nous avons par ailleurs présenté dans le rapport les expériences de réflexion non gouvernementales sur les enjeux éthiques de l'intelligence artificielle, aux États-Unis et au Royaume-Uni, qui sont particulièrement nombreuses et se sont multipliées de façon impressionnante dans la période récente.

L'une des principales initiatives est l'Institut du futur de la vie ou « *Future of Life Institute* » (FLI), situé près du MIT et de Harvard, fondé en mars 2014. Il est à l'origine en janvier 2015 de la lettre d'avertissement sur les

dangers potentiels de l'intelligence artificielle. Le FLI, que nous avons visité en janvier 2017, se donne pour mission de « catalyser et soutenir la recherche et les initiatives visant la sauvegarde de la vie et proposant une vision optimiste de l'avenir ». Il s'agit de « tirer le meilleur profit des nouvelles technologies et de prévenir les risques potentiels pour l'humanité du développement de l'intelligence artificielle ». Lors d'un colloque à New-York sur les défis posés par l'émergence de l'intelligence artificielle, organisé le 14 octobre 2015 par l'Institut de recherche sur la criminalité et la justice des Nations Unies (Unicri), Max Tegmark était invité avec un autre expert à s'exprimer devant quelques 130 délégués. Tous deux ont clairement souligné les risques liés à l'intelligence artificielle et appelé à la mise en place d'une réflexion solide sur l'éthique de l'intelligence artificielle. Le second expert était Nick Bostrom, le fondateur du *Future of humanity Institute* (FHI) en 2005 à l'Université d'Oxford ; il a également fondé, dès 2004, un *Institute for ethics and emerging technologies* (IIEET), proche du mouvement transhumaniste.

De manière similaire au *Future of humanity Institute*, ont été créées plusieurs structures qui travaillent ensemble en réseau, au sein de l'Université de Cambridge : un *Centre for the Study of Existential Risks* (CSER) créé en 2012, un *Leverhulme Centre for the Future of Intelligence* créé en 2016 et, au sein de l'Université de Berkeley, un *Machine Intelligence Research Institute* (MIRI).

Les anciens dirigeants de Paypal, Elon Musk (actuellement patron de Tesla et SpaceX) et Sam Altman ont fondé, le 11 décembre 2015, la « fondation *Open AI* » dans le but de promouvoir l'intelligence artificielle éthique et ouverte. Nous avons visité cette fondation basée dans la Silicon Valley et rencontré ses responsables. Le dernier exemple, peut-être le plus significatif, est le « *Partnership on AI* » formé en septembre 2016 par Google, Microsoft, Facebook, IBM et Amazon afin de réfléchir de manière collective. Vos rapporteurs se sont réjouis du fait qu'Apple a rejoint cette initiative le 26 janvier 2017.

Le rapport « *L'Intelligence artificielle et la vie en 2030* » publié en septembre 2016 par l'Université Stanford dévoile les résultats de l'étude « *One Hundred Year Study of Artificial Intelligence* », un projet universitaire débuté en 2014 et initié par Eric Horvitz, chercheur au laboratoire Microsoft Research.

Dans la foulée et puisqu'il n'existait aucun guide commun de bonnes pratiques dans le domaine de l'intelligence artificielle, plusieurs spécialistes de l'intelligence artificielle et de la robotique se sont réunis lors de la conférence « *Beneficial AI 2017* » organisée par le *Future of Life Institute*. La conférence s'est tenue à Asilomar, en Californie du 5 au 8 janvier 2017. Les participants ont adopté vingt-trois principes baptisés « Les 23 principes d'Asilomar », guide de référence pour l'encadrement éthique du développement de l'intelligence artificielle.

Nous nous interrogeons sur les objectifs précis des GAFAMI et d'Elon Musk à travers ces nombreuses initiatives, qui donnent une place trop

---

grande au risque de l'émergence d'une IA forte qui dominerait et pourrait faire s'éteindre l'espèce humaine. La volonté de ces nouveaux géants pourrait-elle être de se dédouaner ou de créer un nuage de fumée pour ne pas parler des vrais problèmes éthiques posés à court terme par les technologies d'intelligence artificielle, telles que l'usage des données ou le respect de la vie privée ? Vos rapporteurs n'ont pas tranché et laissent aux auteurs de ces initiatives le bénéfice du doute.

Nous soulignons enfin l'important travail en cours sur les enjeux éthiques au sein de l'association mondiale des ingénieurs électriciens et électroniciens (*Institute of Electrical and Electronics Engineers* ou IEEE), qui regroupe plus de 400 000 membres. Son initiative mondiale pour « les considérations éthiques dans l'intelligence artificielle et les systèmes autonomes » a pour objectif de proposer un cadre éthique pour les systèmes d'intelligence artificielle et des systèmes autonomes. Une première version du document a été publiée en décembre 2016, avec l'idée d'une discussion d'ici l'été 2017 et la diffusion d'une deuxième version consolidée à l'automne 2017.

Alors que toutes ces initiatives sur l'éthique sont menées, nous constatons une sensibilisation insuffisante du grand public aux questions posées par l'intelligence artificielle et les systèmes autonomes. Les traitements médiatiques de ces questions restent le plus souvent sensationnalistes voire alarmistes, alors qu'une information objective serait souhaitable. La vision déjà tronquée du grand public, sous l'effet des œuvres de fiction, et en particulier du cinéma, n'est pas améliorée par la lecture de la plupart des articles disponibles sur l'intelligence artificielle dans nos journaux et magazines.

**Mme Dominique Gillot, sénatrice, rapporteure.** – J'en viens aux questions technologiques et scientifiques qui se posent en matière d'intelligence artificielle. Il y a d'abord les sujets d'interrogation liés aux algorithmes utilisés par les technologies d'AI. Le rapport contient des développements sur les questions de sécurité et de robustesse et conclut sur la nécessité de toujours pouvoir arrêter un système d'intelligence artificielle, qu'il s'agisse d'un système informatique ou de son incarnation dans un robot. En 2016, Google a également posé la question du risque de perte de contrôle et c'est dans ce sens que la firme développe l'idée d'un « bouton rouge » permettant la désactivation des intelligences artificielles. La CERNA a aussi cette recommandation. Des recherches complémentaires sont nécessaires car en IA cela peut être compliqué. Pour paraphraser Raymond Aron, qui utilisait l'expression de « Paix impossible, guerre improbable » l'enjeu est donc, face à une paix improbable avec les machines, de rendre la guerre impossible.

Les biais sont l'un des plus gros problèmes posés par les algorithmes d'apprentissage automatique, ou pour être plus rigoureux, posés par les données nécessaires aux algorithmes. La question concerne en effet plus les données que les algorithmes eux-mêmes. Les impacts se font ressortir après

le traitement, mais les biais, eux, sont introduits en amont dès le stade des jeux de données. En effet, les algorithmes d'apprentissage automatique et en particulier d'apprentissage profond vont reproduire, en particulier si les données ne sont pas corrigées, toutes les discriminations connues dans nos sociétés. Il convient donc d'être vigilant sur ces biais, souvent invisibles sans recherches. Le second rapport de la CERNA traite de ce point ainsi que l'initiative « Transalgo » de l'INRIA.

La gouvernance des algorithmes et des prédictions qu'ils opèrent est nécessaire. Le phénomène de « boîtes noires » des algorithmes de *deep learning* appelle un effort de recherche fondamentale pour accroître leur transparence : nous ne disposons d'aucune explication théorique satisfaisante des raisons pour lesquelles les algorithmes de *deep learning* donnent, dans un certain nombre de domaines, d'excellents résultats. Ce problème d'opacité reste entièrement à résoudre. On parle ici de phénomènes de « boîtes noires », mais elles n'ont rien à voir avec les boîtes noires des avions, qui sont des enregistreurs numériques. Le défi à relever est donc celui de l'objectif d'explicabilité des algorithmes de *deep learning*. L'initiative Transalgo de l'INRIA va dans ce sens, afin de répondre aux préoccupations exprimées. Une telle démarche va dans la bonne direction mais gagnerait à voir sa force de frappe être démultipliée par la mobilisation de plusieurs équipes de recherche. L'INRIA ne peut rester la seule structure en France à conduire un tel projet.

Enfin, les algorithmes sélectionnent le contenu des informations dont nous disposons, ce qui pose la question des bulles d'information dites « bulles de filtres » (*filter bubbles*) : l'information ciblée tout comme la publicité personnalisée ou la logique de construction des « fils d'actualité » des réseaux sociaux, à l'instar de celui de Facebook, sont autant d'exemples de réalités déjà manifestes d'usage des systèmes d'intelligence artificielle, qui sont de nature à changer notre rapport au monde, aux autres et à la connaissance en orientant, voire en manipulant, notre perception de la réalité.

Ce sujet mérite une vigilance accrue des pouvoirs publics. Pour nous, l'enfermement, qu'il soit politique, idéologique ou cognitif, doit être combattu. La question va bien plus loin que les critiques formulées à l'encontre des fausses informations ou *fake news*. Sur ce dernier point, la recherche est assez bien avancée et comme l'a indiqué Yann LeCun, l'intelligence artificielle peut être utilisée pour limiter les flux de fausses informations. Des outils de vérification sont ainsi mis en place par plusieurs plateformes, à commencer par Facebook.

J'en viens aux interrogations liées à la singularité, à la convergence NBIC et au transhumanisme.

La rupture dite de la « singularité technologique » appelée aussi simplement singularité, est le nom que des écrivains et des chercheurs en intelligence artificielle ont donné au passage de l'IA faible à l'IA forte. La singularité représente un tournant hypothétique dans l'évolution



---

technologique, dont l'intelligence artificielle serait le ressort principal. De nombreuses œuvres de science-fiction ont décrit ce tournant, qui a été une source d'inspiration très riche pour le cinéma. Les films *Terminator*, *Matrix* ou *Transcendance* sont des exemples de la singularité technologique qui, au-delà de la simple hostilité de l'intelligence artificielle, est souvent au cœur de l'intrigue des œuvres de science-fiction.

Les progrès en matière d'intelligence artificielle, en particulier avec le *deep learning*, sont parfois interprétés comme de « bons » augures de la « singularité » mais rien ne permet de garantir la capacité à créer au cours des prochaines décennies une super-intelligence dépassant l'ensemble des capacités humaines. Par exemple, en s'appuyant sur la loi de Moore, Ray Kurzweil prédit dans un prophétisme dystopique que les machines rivalisant avec l'intelligence humaine arriveraient d'ici 2020 et qu'elles le dépasseraient en 2045.

Nous en sommes aujourd'hui encore très loin et il n'est pas sûr que nous y arrivions un jour. AlphaGo est peut-être le meilleur joueur de Go de tous les temps, mais il n'est pas en mesure de parler ou de distinguer un chat d'un chien, ce dont serait capable n'importe quel joueur de Go humain débutant. Pour le sociologue Dominique Cardon, la tentation de l'IA forte est anthropomorphiste. Certains sont en effet tentés de plaquer sur les futures intelligences artificielles des modes de raisonnement spécifiques à l'intelligence humaine.

L'écrivain et entrepreneur futuriste Jerry Kaplan fait valoir que « *le terme même d'intelligence artificielle est trompeur. Le fait que l'on puisse programmer une machine pour jouer aux échecs, au Go, à Jeopardy ou pour conduire une voiture ne signifie pas pour autant qu'elle soit intelligente ! Aujourd'hui, n'importe quelle calculatrice achetée en supermarché peut faire bien mieux que les plus brillants cerveaux. Ces calculatrices sont-elles pour autant intelligentes ? Je ne le crois pas. Au fil du temps, nous découvrons de nouvelles techniques permettant de résoudre des problèmes bien précis, à l'aide de l'automatisation. Cela ne signifie pas pour autant que nous soyons en train de construire une super-intelligence en passe de prendre le pouvoir à notre place* ».

Ces observations conduisent à relativiser les récents progrès de l'intelligence artificielle et en particulier à contester le fantasme de l'intelligence artificielle forte car elles récusent la pertinence d'une comparaison avec l'intelligence humaine.

Ce catastrophisme oublie également le caractère irréductible de l'intelligence humaine au calcul. Il évacue la place des émotions, celle de l'intelligence corporelle.

Non seulement l'avènement d'une super intelligence à long terme n'est pas certaine mais la menace à court terme relève du pur fantasme. Il s'agit de fantasmes sur la capacité des algorithmes à devenir conscients, autrement dit dotés de capacités réflexives les rendant capables de se représenter à eux-mêmes.

Pour nous, nier la possibilité d'une IA forte n'a pas de sens, toutefois se prononcer sur son imminence ou sur le calendrier précis de son avènement semble tout aussi peu raisonnable, car c'est indémontrable scientifiquement.

Yann LeCun estime que *« beaucoup des scénarios catastrophes (en intelligence artificielle) sont élaborés par des personnes qui ne connaissent pas les limites actuelles du domaine. Or les spécialistes disent qu'ils sont loin de la réalité »*.

De même Rob High, directeur technique du projet Watson d'IBM, estime qu'il est *« trop tôt pour employer le terme intelligence artificielle, mieux vaut parler d'outils capables d'élargir les capacités cognitives humaines »*.

Greg Corrado, directeur de la recherche en intelligence artificielle chez Google, nous a expliqué qu'il était plus juste de parler d'intelligence augmentée plutôt que d'intelligence artificielle. Pour Jean-Claude Heudin, l'intelligence artificielle ne remplace pas l'homme mais augmente son intelligence, en formant une sorte de *« troisième hémisphère »*.

Cette idée de complémentarité hommes-machines et d'intelligence augmentée nous a convaincus. François Taddéi explique lui que *« les intelligences humaine et artificielle coévoluent. Mais ce sont encore les combinaisons homme-machine qui sont les plus performantes : on le voit aux échecs, où une équipe homme-machine est capable de battre et l'homme et la machine »*. L'homme et la machine, les hommes-centaures, sont toujours plus forts que toutes les machines.

Pour ce qui concerne la *« convergence NBIC »*, convergences entre les nanotechnologies, les biotechnologies, les technologies de l'information et les sciences cognitives, thème issu du rapport de MM. Roco et Bainbridge à la *National Science Foundation* (États-Unis) en 2003, ce projet ambitieux de fertilisation croisée n'a pas produit de grands résultats à ce stade mais les progrès en intelligence artificielle, en génomique, en sciences cognitives et en neurosciences reposent la question aujourd'hui.

Notre président Jean-Yves Le Déaut conduit un travail à ce sujet pour l'Assemblée parlementaire du Conseil de l'Europe, en tant que rapporteur pour la science et la technologie, afin que cette convergence soit respectueuse des droits humains.

Nous avons vu que la prospective en intelligence artificielle aboutit souvent à des scénarios de dystopie technologique mais ce pessimisme n'est pas partagé par l'ensemble des futurologues puisque, pour certains, les progrès de l'intelligence artificielle permettront de protéger et prolonger la vie humaine, mais aussi d'offrir une opportunité historique pour concrétiser l'utopie transhumaniste. Le transhumanisme est un mouvement philosophique qui s'apparente à une religion, prédisant et travaillant à une amélioration de la nature de l'homme grâce aux sciences et aux évolutions technologiques. Pour les transhumanistes, l'homme *« augmenté »* pourrait devenir immortel. Inutile de préciser que je n'y crois pas du tout...

---

Ce projet transhumaniste de mort de la mort et de fin de la souffrance n'empêche pas l'adhésion de vos rapporteurs. Il s'apparente à une négation de la nature humaine. Pour nous, l'intelligence artificielle n'est pas un acte de foi et ne doit pas le devenir.

Selon Raja Chatila *« derrière ces discours, nous avons des vues de l'esprit qui n'ont rien d'opérationnelles, elles sont en réalité des idéologies, qu'on cherche à imposer pour gommer les différences entre l'humain et le non-humain »*.

Il s'agit de chimères qui empêchent de se poser les vraies questions pertinentes. Il est essentiel de savoir anticiper les problèmes potentiels posés par l'intelligence artificielle. À court terme, ces problèmes risquent d'être ignorés et pris à tort pour de la science-fiction. Il convient en effet de distinguer les craintes issues de certaines fictions cinématographiques des problèmes réels qui risquent de survenir plus ou moins rapidement.

J'en arrive donc à nos quinze recommandations. Nous sommes pour une IA maîtrisée, objet de nos cinq premières propositions. Tout d'abord, proposition n° 1 : se garder d'une contrainte juridique trop forte sur la recherche en intelligence artificielle, qui – en tout état de cause – gagnerait à être, autant que possible, européenne, voire internationale, plutôt que nationale

Proposition n° 2 : favoriser des algorithmes et des robots sûrs, transparents et justes et prévoir une charte de l'intelligence artificielle et de la robotique. Il faut rechercher la transparence des algorithmes contre les boîtes noires, les biais et les discriminations. Il convient de prévoir aussi des mécanismes de traçabilité, de type enregistreurs numériques des avions. Une charte de l'intelligence artificielle et de la robotique la plus internationale possible, européenne à défaut, proclamerait ces objectifs éthiques et viserait à codifier les bonnes pratiques. Elle proposerait des règles sur les interactions homme-machine, en posant des limites en matière d'imitation du vivant, pour les robots androïdes comme pour les agents conversationnels.

Proposition n° 3 : former à l'éthique de l'intelligence artificielle et de la robotique dans les cursus spécialisés de l'enseignement supérieur qui traitent de l'intelligence artificielle et de la robotique.

Proposition n° 4 : confier à un institut national de l'éthique de l'intelligence artificielle et de la robotique un rôle d'animation du débat public sur les principes éthiques qui doivent encadrer ces technologies. Au-delà de la nouvelle mission de la CNIL, cet institut national de l'éthique de l'intelligence artificielle et de la robotique pourra s'intéresser aux problématiques d'explicabilité vues plus haut. La démarche ne doit pas être réservée à une seule structure de recherche, plusieurs équipes doivent y travailler parallèlement et un institut national pourrait impulser les projets, coordonner les recherches, animer le débat public et faire des propositions aux autorités. Les pouvoirs publics ne devront pas être les seuls à le financer. Les entreprises privées, qui se donnent pour objectif d'informer et d'éduquer sur ces technologies et d'accroître leurs effets bénéfiques pour la société (à

---

l'image du « *partnership on AI* »), pourraient participer au financement de l'institut.

Proposition n° 5 : accompagner les transformations du marché du travail sous l'effet de l'intelligence artificielle et de la robotique en menant une politique de formation continue ambitieuse visant à s'adapter aux exigences de requalification et d'amélioration des compétences. Je propose à titre personnel de réfléchir à un nouveau mode de financement de notre système de protection sociale, qui serait un complément des cotisations existantes et qui pourrait consister en un prélèvement de cotisations sociales sur les agents autonomes, dans la mesure où ils remplacent des emplois occupés par des êtres humains. C'est le rapport de Mady Delvaux qui m'a inspiré cette idée, d'autant que la proposition de taxer les robots a fait son apparition dans la campagne pour les élections présidentielles. Je précise que mon co-rapporteur est contre toute taxe spécifique sur l'intelligence artificielle et les robots.

**M. Claude de Ganay, député, rapporteur.** – C'est vrai ! Un mécanisme de ce type constituerait selon moi un mauvais signal et découragerait la recherche, l'innovation et l'activité économique. La TVA et l'impôt sur les sociétés (IS) s'appliquent déjà à ces activités – et quand ils ne le font pas il faudra y veiller. C'est notre seul point de désaccord – le mot est trop fort – et nous présentons quinze propositions totalement communes. Dominique Gillot ne propose pas une telle taxe, elle l'évoque comme sujet de réflexion.

Je poursuis avec notre deuxième série de propositions, pour une intelligence artificielle utile, au service de l'homme et des valeurs humanistes.

Notre proposition n° 6 : redonner une place essentielle à la recherche fondamentale et revaloriser la place de la recherche publique par rapport à la recherche privée tout en encourageant leur coopération. Seule la recherche fondamentale peut répondre aux problèmes d'explicabilité des algorithmes et de biais dans les données. Nous en avons besoin. Des enjeux de maîtrise des technologies aux enjeux de financement de la recherche publique, tout se tient.

La recherche fondamentale pose par ailleurs la question du mode de financement des projets : il faut favoriser la recherche transversale et ne surtout pas reproduire l'hyperspécialisation entre sous-domaines de l'intelligence artificielle. Il est en cela nécessaire de mobiliser les équipes de chercheurs autour de grands projets nationaux structurants. Des projets de grandes bases de données labellisées, nécessaires à l'apprentissage machine, pourraient être lancés, par exemple autour de la langue française, du marché de l'emploi ou des données de santé, sous condition d'anonymisation.

Proposition n° 7 : encourager la constitution de champions européens en intelligence artificielle et en robotique, un peu sur le modèle d'Airbus. Sans verser dans le nationalisme industriel, il faut réfléchir aux

---

protections qui pourraient être instituées : les laboratoires français sont pillés de leurs chercheurs par les multinationales nord-américaines et chinoises.

Proposition n° 8 : orienter les investissements dans la recherche en intelligence artificielle vers l'utilité sociale des découvertes, à savoir des applications à impact sociétal bénéfique : bien-être, santé, dépendance, handicap, infrastructures civiles, gestion des catastrophes.

Proposition n° 9 : élargir l'offre de cursus et de modules de formation aux technologies d'intelligence artificielle dans l'enseignement supérieur français et créer dans notre pays au moins un pôle d'excellence international et interdisciplinaire en intelligence artificielle et en robotique. Il peut s'agir d'un pôle, de deux ou de trois, en s'appuyant sur l'excellence de l'INRIA, du LAS de Toulouse, de l'ENS, de l'Institut Mines-Télécom... Mais à défaut d'une création pure et simple, il sera urgent d'encourager la coordination et d'accroître la cohérence des instituts, des centres et des équipes de recherches.

Proposition n° 10 : structurer et mobiliser la communauté française de la recherche en intelligence artificielle en organisant davantage de concours primés à dimension nationale, destinés à dynamiser la recherche en intelligence artificielle. « France IA » est une étape importante dans la mobilisation de la communauté française de la recherche en intelligence artificielle. Il faut continuer et la structurer encore davantage, par exemple à travers l'organisation de concours. Le traitement de grandes bases de données nationales labellisées pourrait être l'objet d'un de ces concours. Un travail avec l'Agence nationale de recherche (ANR) peut être envisagé pour définir une offre française de grands concours primés en IA.

Proposition n° 11 : assurer une meilleure prise en compte de la diversité et de la place des femmes dans la recherche en intelligence artificielle. La place des femmes et la question des minorités dans la recherche en intelligence artificielle sont des défis qu'il convient de relever.

J'en arrive à la troisième et dernière série de propositions, pour une intelligence artificielle démystifiée.

Proposition n° 12 : organiser des formations à l'informatique dans l'enseignement primaire et secondaire faisant une place à l'intelligence artificielle et à la robotique. Il s'agit d'aller un peu plus loin que l'offre actuelle.

Proposition n° 13 : former et sensibiliser le grand public à l'intelligence artificielle par des campagnes de communication, l'organisation d'un salon international de l'intelligence artificielle et de la robotique et la diffusion d'émissions de télévision pédagogiques

Un salon international de l'intelligence artificielle et de la robotique est à organiser en France, en s'inspirant de *VivaTech* et de l'initiative *Innoro* portée par Catherine Simon, organisatrice du Salon français de la robotique. Ce salon pourrait être le pendant européen du CES (*Consumer electronics Show*) organisé à Las Vegas et sans équivalent en Europe.

Les réseaux sociaux ou la télévision pourraient être des supports pour des émissions de partage de la connaissance. Les émissions de télévision du type « la faute à l'algo », diffusées par la chaîne No Life et certains épisodes de « data-gueule », diffusés par la chaîne France 4 sont des exemples intéressants.

Il faut se saisir du « *partnership on AI* » pour associer les entreprises à ce travail pédagogique. Là aussi, le coût du financement pourrait être partagé avec les entreprises privées, qui se donnent pour objectif d'informer et d'éduquer sur ces technologies. Je trouve déplorable que les grand-messes d'Apple ne s'accompagnent jamais de présentations pédagogiques sur les technologies d'intelligence artificielle.

Proposition n° 14 : former et sensibiliser le grand public aux conséquences pratiques de l'intelligence artificielle et de la robotisation, il s'agit, en complément de l'offre de formation continue visant l'amélioration continue des compétences, de permettre aux travailleurs et au grand public d'envisager de manière positive les transitions à venir en termes de conséquences pratiques de l'intelligence artificielle et de la robotisation. Là encore, le coût du financement pourrait être partagé avec les entreprises privées, qui se donnent pour objectif d'informer et d'éduquer sur ces technologies.

Quinzième et dernière proposition : être vigilant sur les usages spectaculaires et alarmistes du concept d'intelligence artificielle et de représentations des robots. Il s'agit d'éviter les dérapages, mais dans le respect de la liberté de création et de la liberté d'expression. La vérification des publicités serait un premier pas vers une plus grande maîtrise de la communication médiatique sur le sujet. Face aux scénarios catastrophistes, il faut sensibiliser les écoles de journalisme.

Pour conclure, j'indique que nos propositions devront être remises en débat à proportion des découvertes scientifiques, de leurs transferts et de leurs usages. Le point d'équilibre issu du rapport doit pouvoir évoluer, en fonction du contexte qui résultera du jeu de ces variables.

Nous faisons une proposition à nos collègues, en plus de nos 15 propositions, une sorte de proposition zéro comme aurait dit Asimov : la poursuite des travaux de l'OPECST sur les enjeux de l'intelligence artificielle en 2017 et 2018. Outre une veille générale des rapporteurs sur le sujet, il s'agira d'un suivi de la reprise de leurs propositions par le Gouvernement ainsi que d'un approfondissement de leur travail. Un suivi du sujet par l'OPECST apparaît indispensable.

Nous appelons par ailleurs à la poursuite du plan national pour l'intelligence artificielle, annoncé en janvier 2017 par le Gouvernement et qui sera précisé de manière plus détaillé à la fin du mois.

Ni quête vaine, ni projet de remplacement de l'homme par la machine, l'intelligence artificielle représente une chance à saisir pour nos sociétés et nos économies. La France doit relever ce défi. Les progrès en

---

intelligence artificielle sont d'abord et avant tout bénéfiques. Ils comportent aussi des risques, il serait malhonnête de le nier. Mais ces risques peuvent et doivent être identifiés, anticipés et maîtrisés.

L'avènement d'une super-intelligence ne fait pas partie de ces risques à court et moyen terme. Et, à long terme, la réalité de cette menace n'est pas certaine, quant à son imminence à court ou moyen terme, prophétisée par plusieurs figures médiatiques, elle relève du pur fantasme. Le présent rapport se veut une première contribution au travail indispensable d'identification, d'anticipations et de maîtrise des risques réels, travail de démystification et d'objectivation qui doit être collectif, interdisciplinaire et international.

**Mme Dominique Gillot, sénatrice, rapporteure.** – J'insiste à mon tour sur la nécessité pour notre Office de poursuivre ces travaux en 2017 et 2018. La meilleure façon de prévenir tout risque de futures désillusions est de suivre en continu l'évolution de ces technologies et de leurs usages, sachant que les cycles d'espoirs et de déceptions qui jalonnent l'histoire de l'intelligence artificielle invitent à ne pas faire preuve d'attentes irréalistes. Les propositions du rapport vont dans ce sens. Nous nous prononçons pour une intelligence artificielle maîtrisée, utile et démystifiée. Maîtrisée, parce que ces technologies devront être les plus sûres, les plus transparentes et les plus justes possibles. Utile, parce qu'elles doivent, dans le respect des valeurs humanistes, profiter à tous au terme d'un large débat public. Et démystifiée, parce que les difficultés d'acceptabilité sociale de l'intelligence artificielle résultent largement de visions catastrophistes erronées, propagées par des ignorants.

Plutôt qu'une hypothétique future confrontation entre les hommes et les machines, qui relève de la science-fiction dystopique, nous croyons au bel avenir de la complémentarité entre l'homme et la machine. C'est, au final, bien plus vers une intelligence humaine augmentée que vers une intelligence artificielle concurrençant l'homme que nous allons.

**M. Jean-Yves Le Déaut, président.** – Merci pour cet excellent rapport. Avant de poursuivre les travaux durant les années à venir, il faudra d'abord être réélus, et que l'Office soit saisi, puisqu'il ne peut s'autosaisir...

**Mme Dominique Gillot, rapporteure.** – Bien entendu ! C'est un vœu que j'exprimais là.

**M. Bruno Sido, vice-président.** – Ce rapport est passionnant. Je suis d'accord avec toutes vos propositions, sauf la cinquième, qui n'a selon moi pas sa place ici, notre office ne traitant pas de questions sociales ou fiscales. Il faudra renvoyer ces points aux commissions permanentes concernées. Quant à la proposition n°11, je ne sais pas qu'en science, le sexisme ou le refus de la diversité aient cours : les Américains par exemple, prennent les meilleurs dans le monde entier, et s'en vantent !

**M. Claude de Ganay, rapporteur.** – Il y a plus d'hommes que de femmes dans la recherche et dans les formations scientifiques : il faut que les

femmes rejoignent ces rangs. Comment favoriser ce mouvement, je ne sais pas...

**Mme Dominique Gillot, rapporteure.** – Il y a deux aspects distincts : assurer la mixité dans ce secteur de la recherche, mais aussi veiller à ce que les algorithmes ne comportent pas de biais liés au genre, à l'origine ethnique ou socioculturelle.

**M. Jean-Yves Le Déaut, président.** – Vous posez la question de la taxation des robots : mais où se situe la frontière entre ceux-ci et les machines ? La disparition du travail humain est une question pertinente, dont je ferais mention plutôt dans le corps du rapport, non dans les propositions, en demandant aux commissions compétentes de s'y pencher.

Quant au choix entre une taxation, des cotisations sociales, ou un autre moyen...

**Mme Dominique Gillot, rapporteure.** – Ce n'est pas une proposition mais une orientation pour la réflexion à venir. Si les tâches à faible valeur ajoutée doivent dans l'avenir être de plus en plus fréquemment assurées par des robots ou « agents autonomes », il y a lieu de s'interroger sur l'équilibre des comptes sociaux.

**Mme Catherine Procaccia.** – Mais ne croyez-vous pas que le système de protection sociale va évoluer ? La question doit être posée dans le cadre d'une réflexion globale sur les comptes sociaux.

**Mme Dominique Gillot, rapporteure.** – Le rapport de Mady Delvaux a suscité un débat autour d'une taxe robots, dans les médias et les milieux politiques.

**Mme Catherine Procaccia.** – M. Hamon lui-même...

**Mme Dominique Gillot, rapporteure.** – Oui, mon candidat la propose, mais je ne pense pas que ce soit une bonne solution : ce serait un signal négatif pour la recherche en IA. On en parle néanmoins, et France Inter y a consacré une émission entière, dans *On n'arrête pas l'éco*.

**Mme Fabienne Keller, sénatrice.** – Je salue votre énorme travail. La Chine est un continent, le nombre de chercheurs en IA y est considérable. Or vos propositions concernent surtout la France. Toutes devraient être, à tout le moins, à périmètre européen. Ma chère professeure de physique disait : « *Le nombre de femmes dans les sciences est le produit de facteurs cumulatifs, tous inférieurs à un* ». Effectivement, il y a moins de filles que de garçons au baccalauréat scientifique, encore moins dans les écoles supérieures, et finalement, elles ne sont que 24 % dans la recherche sur la robotique... On l'observe, plus qu'on ne l'explique. Une discrimination positive serait sans doute utile pour cesser de nous priver de l'intelligence féminine... qui n'a rien d'artificiel !

Pourriez-vous revenir sur la différence entre les robots et l'IA ? Celle-ci est essentiellement multidisciplinaire. L'Institut de recherche contre les cancers de l'appareil digestif (IRCAD) à Strasbourg travaille sur des



---

robots chirurgicaux capables de reproduire les meilleures opérations effectuées dans le monde par les praticiens, en traitant les données numérisées des corps mous. On parle à présent d'imprimer en 3D la prothèse nécessaire, calquée sur l'organe du patient, produite en temps réel.

Les applications touchent des domaines très divers, et pour les transports, dans le cadre de l'intermodalité, elles sont très intéressantes : des propositions de transport pourraient être faites sur le téléphone portable de l'usager, calculées en fonction des heures, des habitudes, des besoins, notamment pour les personnes handicapées.

Comment avez-vous pu traiter de la pluridisciplinarité de l'intelligence artificielle, quand les déclinaisons sont si vastes ?

**Mme Delphine Bataille, sénatrice.** – Au regard des investissements publics dans les pays dont vous avez pu étudier l'effort en IA, quelle est la place de la France, notamment pour les infrastructures de calcul intensif ? De nombreux acteurs industriels sont-ils équipés ? Les réglementations sont-elles adaptées ? Les contraintes juridiques ne sont-elles pas excessives ? Existe-t-il parmi les pays que vous avez étudiés un modèle d'organisation ? Au Japon, trois ministères, sciences et technologies, économie et industrie, intérieur et communications, ont chacun un centre de recherche, mais les trois se coordonnent. Avec quel pays la France entretient-elle la collaboration scientifique la plus aboutie ? Le plan national pour l'intelligence artificielle s'inspire-t-il d'exemples à l'étranger ?

**M. Jean-Yves Le Déaut, président.** – Nos rapporteurs disent que ces sujets sont internationaux, mais leurs propositions sont plutôt nationales alors qu'elles gagneraient à être davantage européennes. Il existe une convention internationale en biologie, dite d'Oviedo, dont l'article 13 interdit par exemple la modification du génome se répercutant aux descendants. Le moment n'est-il pas venu d'élargir ce type d'accord, en désignant une organisation pour le proposer ? Ne faudrait-il pas demander à l'Unesco de se charger d'une supervision internationale ? Nous avons par ailleurs un organisme européen avec lequel l'Office travaille, le *European parliamentary technology assessment* (EPTA), qui inclut le Conseil de l'Europe, et, parmi les pays observateurs, la Russie, les États-Unis le Japon : il faudrait sans doute organiser des auditions publiques et contradictoires des experts de tous les pays participants, ce serait une manière de commencer ce travail international.

La proposition n°8, qui encourage la recherche au service de la société, me paraît très bonne, elle s'inscrit bien dans la démarche de la stratégie nationale de recherche (SNR). Il faudrait faire un lien avec la SNR que nous venons d'évaluer et préciser que la question du handicap et de l'atténuation de la frontière entre l'homme et la machine pourrait être traitée...

Face à la thèse du robot autonome, je défends la thèse que toute machine issue de l'intelligence artificielle, tout robot, doit rester sous le contrôle de l'homme.

---

**Mme Dominique Gillot, rapporteure.** – Oui ! C'est le « bouton rouge ». Cela figure dans notre rapport. Et c'est l'esprit de la proposition n°2, même si nous n'entrons pas dans le détail.

**M. Jean-Yves Le Déaut, président.** – Il serait bon de le redire dans la conclusion.

**Mme Dominique Gillot, rapporteure.** – D'accord.

Le niveau pertinent est effectivement celui de l'Europe. Mais la France a des atouts majeurs à mettre en avant, son école mathématique, l'écosystème du développement et de la science, la prise de conscience des organismes de recherche... En trois ans, la prise de conscience s'est diffusée. La pluridisciplinarité, plus que la transdisciplinarité, se développe.

L'intelligence artificielle fait intervenir une juxtaposition de sciences, qui se conjuguent, s'hybrident et se renforcent mutuellement, grâce à la puissance de calcul, les données de masse, la reconnaissance d'image de plus en plus rapide. Le robot - d'observation, ou d'accompagnement médical, etc. – est une machine physique qui comprend une intelligence artificielle embarquée.

**Mme Fabienne Keller.** – Il peut y avoir de l'intelligence artificielle sans robot.

**Mme Dominique Gillot, rapporteure.** – Dans un téléphone portable, par exemple. Quant aux robots, ils vont des plus simples jusqu'à l'humanoïde.

Le président Jean-Yves Le Déaut suggère que l'Union européenne prenne en charge des missions en matière d'intelligence artificielle : mais si les services à Bruxelles sont performants sur les chartes éthiques, ils le sont beaucoup moins sur les aspects scientifiques. Lorsque nous avons rencontré les équipes de la Commission européenne, nous avons été fort déçus : nous étions mieux informés qu'elles ! Le CESE, qui prépare un rapport sur l'IA, fait le même constat. Les experts de la Commission européenne ne se sont pas emparés du sujet. Ils sont en retard par rapport aux scientifiques et aux politiques de notre pays. L'actualité est pourtant intense dans ce domaine !

**M. Claude de Ganay, rapporteur.** – Pour répondre au président, je dirai que les propositions n°1, 2 et 7 ont bien la vocation européenne qu'il réclame. Ce rapport était très attendu par la communauté scientifique : il dresse en effet un état des lieux sur le plan scientifique. Mady Delvaux a rédigé un rapport de compromis, non de propositions, tenant compte du cadre européen dans lequel elle intervenait. L'OPECST a plus de liberté, il peut être plus objectif. Je confirme ce qu'a dit Mme Gillot, nos interlocuteurs à Bruxelles ignoraient jusqu'à l'existence des rapports de la Maison blanche et de la Chambre des communes britannique. Ce travail a été passionnant, nous l'avons abordé sans *a priori*.

**M. Jean-Yves Le Déaut, président.** – Je mets aux voix le rapport sous réserve des corrections rédactionnelles que nous avons mentionnées.

*Le rapport est adopté à l'unanimité.*

---

## ANNEXE 3 : LISTE DES PERSONNES RENCONTRÉES

### I. PERSONNES RENCONTRÉES PAR LES RAPPORTEURS EN VUE DE L'ÉTUDE DE FAISABILITÉ

M. Raja Chatila, Directeur de recherche au Centre national de la recherche scientifique (CNRS), directeur de l'Institut des systèmes intelligents et de robotique (ISIR)

M. Max Dauchet, Professeur émérite à l'Université de Lille, Président de la Commission de Réflexion sur l'Éthique de la Recherche en sciences et technologies du Numérique (CERNA) d'Allistene<sup>1</sup>

Mme Laurence Devillers, Professeure à l'Université Paris IV Sorbonne et directrice de recherche au Laboratoire d'informatique pour la mécanique et les sciences de l'ingénieur (Limsi de Saclay)

M. Alain Fuchs, Président du Centre national de la recherche scientifique (CNRS), Professeur de chimie

M. Jean-Gabriel Ganascia, Professeur à l'Université Paris VI Pierre et Marie Curie (UPMC), directeur de l'équipe ACASA (Agents Cognitifs et Apprentissage Symbolique Automatique) au laboratoire d'informatique de Paris VI (LIP6), membre du COMETS (Comité d'éthique du CNRS)

M. Michael Matlosz, Président-directeur général de l'Agence nationale de la recherche (ANR), Professeur de génie des procédés

M. Antoine Petit, Président-directeur général de l'Institut national de recherche en informatique et en automatique (Inria), Professeur à l'Ecole normale supérieure de Cachan

M. François Taddei, Directeur du Centre de recherches interdisciplinaires (CRI) de l'Université Paris V René Descartes), directeur de recherche à l'Inserm, ingénieur en chef des Ponts, des Eaux et des Forêts

---

<sup>1</sup> Alliance des sciences et technologies du numérique, cette association regroupe les organismes de recherche concernés par les différents domaines du numérique, notamment par l'intelligence artificielle, comme la CDEFI, le CEA, le CNRS, la CPU, l'INRIA et l'Institut Mines-Télécom. Ses membres associés sont l'INRA, l'INRETS et l'ONERA.



---

## II. PERSONNES RENCONTRÉES PAR LES RAPPORTEURS EN VUE DE L'ÉTABLISSEMENT DU RAPPORT

### A. EN FRANCE

#### 1. À Paris (lors de différentes auditions de rapporteurs)

##### Mardi 25 octobre 2016

- M. Patrick Albert, entrepreneur (créateur et ancien directeur de ILOG), chercheur et pionnier dans le domaine de l'intelligence artificielle ;
- M. Stéphane Mallat professeur à l'École normale supérieure (ENS), chercheur en mathématiques appliquées.

##### Mardi 8 novembre 2016

- M. Marc Mézard, directeur de l'École normale supérieure (ENS) ;
- M. Marc Pouzet directeur des études du département d'informatique de l'École normale supérieure (ENS) ;
- M. Yves Caseau, directeur du numérique pour le groupe AXA, membre de l'Académie des technologies et responsable de sa commission TIC ;
- M. Pierre-Yves Oudeyer, directeur de recherche à l'INRIA, directeur du laboratoire *Flowers*, président du comité technique des systèmes cognitifs et développementaux de l'IEEE (*Institute of Electrical and Electronics Engineers*, en français Institut des ingénieurs électriciens et électroniciens) ;
- M. Grégory Bonnet, GREYC - Normandie Université) ;
- M. Alain Berger, société Ardans ;
- M. Olivier Boissier, Institut Henry Fayol - ARMINES ;
- M. Pierre-Antoine Chardel, Institut Mines Télécom ;
- M. Jean-Gabriel Ganascia, LIP6 - Université Paris 6 ;
- Mme Catherine Tessier, ONERA ;
- M. Nicolas Cointe et Mme Fiona Berreby, chercheurs en thèse de doctorat sur l'éthique de l'intelligence artificielle ;

- 
- M. Raja Chatila, directeur de recherche au Centre national de la recherche scientifique (CNRS), directeur de l'Institut des systèmes intelligents et de robotique (ISIR) ;
  - M. Benoît Girard, directeur de recherche CNRS à l'ISIR ;
  - M. Nathanaël Jarrassé, chercheur CNRS à l'ISIR.

### **Mercredi 9 novembre 2016**

- M. Pierre-Yves Oudeyer, directeur de recherche à l'INRIA, directeur du laboratoire *Flowers*, président du comité technique des systèmes cognitifs et développementaux de l'IEEE (*Institute of Electrical and Electronics Engineers*, en français Institut des ingénieurs électriciens et électroniciens) ;
- M<sup>e</sup> Alain Bensoussan, avocat, président de l'association pour les droits des robots et M<sup>e</sup> Marie Soulez, avocate spécialisée sur les TIC dans son cabinet ;
- M. Laurent Alexandre, président de DNA Vision, fondateur de Doctissimo, auteur, chirurgien-urologue et ancien élève de l'ENA ;
- M. Henri Verdier, entrepreneur et spécialiste du numérique français, directeur interministériel du numérique et du système d'information, adjoint à la secrétaire générale pour la modernisation de l'action publique, administrateur général des données (AGD), membre du conseil scientifique de l'Institut Mines-Télécom; du comité de prospective de l'ARCEP, du comité de prospective de la CNIL et de la Commission innovation 2030, ex-président de Cap Digital, pôle de compétitivité et de transformation numérique.

### **Jeudi 24 novembre 2016**

- M. Max Dauchet, Professeur émérite à l'Université de Lille, Président de la Commission de Réflexion sur l'Éthique de la Recherche en sciences et technologies du Numérique (CERNA) d'Allistene ;
- M. Cédric Sauviat, ingénieur, président de l'association française contre l'intelligence artificielle (AFCIA) et Marie David, ingénieur, éditrice, membre du bureau de l'association ;
- Mme Flora Fischer, chargée de programme de recherche au Cigref, réseau de grandes entreprises, pilote du groupe de travail sur l'intelligence artificielle en entreprise.

### **Lundi 28 novembre 2016**

- M. Nicolas Cointe et Mme Fiona Berreby, chercheurs en thèse de doctorat sur l'éthique de l'intelligence artificielle ;
- M. Claude Berrou professeur à Télécom Bretagne (Institut MINES-TELECOM), chercheur en électronique et informatique, membre de l'Académie des sciences, surtout connu pour ses travaux sur les turbo-codes, très utilisés en téléphonie mobile<sup>1</sup> ;
- Mme Laurence Devillers<sup>2</sup> professeure d'informatique à l'université Paris-Sorbonne et directrice de recherche du CNRS au Laboratoire d'informatique pour la mécanique et les sciences de l'ingénieur (Linsi de Saclay).

### Mercredi 30 novembre 2016

- M. François Taddéi directeur du Centre de recherches interdisciplinaires (CRI). (Inserm, université Paris Descartes), biologiste ;
- M. Igor Carron, entrepreneur, organisateur du principal « *meet-up* » en intelligence artificielle en France intitulé « *Paris Machine Learning* » ;
- M. Jill-Jen Vie, chercheur en thèse de doctorat à l'École normale supérieure Paris-Saclay sur le *deep learning*<sup>3</sup> ;
- M. Dominique Sciamma directeur de l'école de Design « *Strate* » à Sèvres ;

<sup>1</sup> Au début des années 1990, il met au point avec Alain Glavieux la première classe quasi-optimale de codes correcteurs utilisant des codes convolutifs, les turbocodes, aujourd'hui très utilisés pour la téléphonie mobile 3G et 4G (plusieurs brevets sur ces systèmes, dont le premier -brevet européen- remonte à 1991 et est détenu en copropriété par France Telecom, TDF et l'Institut Télécom). Les deux chercheurs, avec d'autres collègues de Télécom Bretagne, étendent également le principe turbo à des fonctions autres que le codage correcteur d'erreurs, en particulier l'égalisation. Claude Berrou fait partie des dix scientifiques français les plus cités dans les sciences de l'information. En 2003, il reçoit le Grand Prix France Télécom de l'Académie des sciences et la médaille Hamming, puis en 2005 le prix Marconi.

<sup>2</sup> Experte en traitement du langage et du signal et en apprentissage machine, ses recherches portent principalement sur l'« *affective computing* », le traitement automatique de la langue parlée, la détection des émotions « *real-life* », l'interaction homme-machine et la robotique affective et interactive. Laurence Devillers anime l'équipe de recherche *Dimensions affectives et sociales* dans les interactions parlées. Elle a participé/participe à de nombreux projets notamment sur les interactions affectives et sociales humain-robot (ANR Tecsan Armen, FUI Romeo, BPI Romeo2, Rex EU Humaine, EU Chistera Joker). Elle travaille sur les émotions mais aussi sur l'humour et l'empathie dans les systèmes de dialogue homme-machine. Ses travaux peuvent être utilisés pour des applications avec des robots, des objets connectés, des jeux sérieux, des centres d'appels pour différents domaines comme par exemple la santé, le bien-être et la sécurité. Elle anime le pôle co-évolution humain-machine de l'Institut de la société numérique (ISN - Paris Saclay), où elle mène des travaux en collaboration avec des chercheurs en droit et en sociologie sur mémoire du robot et responsabilité, et sur la réflexivité langagière. Elle est également membre du conseil d'administration de AAAC (*emotion-research.net*), membre de IEEE, ACL, ISCA and AFCP. Elle est aussi impliquée dans l'association Eurobotics dans les groupes de travail sur « *Natural Interaction with Social Robot* » et « *Socially intelligent robots* ». Elle a participé à la rédaction du premier rapport de la Cerna sur l'éthique du chercheur en robotique dont elle est membre.

<sup>3</sup> Thèse d'informatique à l'ENS et à l'université Paris-Saclay en 2016. En 2017, il est en postdoctorat au laboratoire RIKEN à Tokyo, sous la direction de Hisashi Kashima. Consultant du ministère de l'Éducation nationale sur le projet PIX de certification des compétences numériques, coproducteur de l'émission « *La Faute à l'algo* » pour la chaîne Nolife avec Michel Blockelet, sur le rôle grandissant de l'algorithmie dans nos vies.



- 
- David Sadek directeur de la recherche de l'Institut Mines-Télécom, spécialiste en intelligence artificielle ;
  - Verity Harding, directrice des affaires publiques de *Deep Mind* (Londres, Royaume-Uni) ;
  - Paul Strachman représentant d'ISAI aux États-Unis<sup>1</sup>, diplômé de l'École Nationale des Ponts & Chaussées d'un Master de la *London School of Economics and Political Science* (LSE) et d'un MBA de Stanford University.

## 2. À Paris (lors de l'audition publique du 19 janvier 2017)

- Mme Axelle Lemaire, secrétaire d'État chargée du numérique et de l'innovation ;
- M. Jean-Gabriel Ganascia, professeur à l'Université Pierre-et-Marie-Curie Paris VI ;
- M. Gérard Sabah, directeur de recherche honoraire au CNRS ;
- M. Yves Demazeau, président de l'Association française pour l'intelligence artificielle (AFIA) ;
- M. Bertrand Braunschweig, directeur du Centre de l'INRIA de Saclay ;
- M. David Sadek, directeur de la recherche de l'Institut Mines-Télécom ;
- M. Jean-Daniel Kant, maître de conférences à l'Université Pierre-et-Marie-Curie-Paris VI ;
- M. Benoît Le Blanc, directeur-adjoint de l'École nationale supérieure de cognitive ;
- M. Jean-Marc Merriaux, directeur général de Canopé ;
- M. François Taddéi, directeur du Centre de recherche interdisciplinaire ;
- M. Olivier Esper, responsable des affaires publiques de Google France ;
- Mme Delphine Reyre, directrice Europe des affaires publiques de Facebook ;
- M. Laurent Massoulié, directeur du Centre de recherche commun INRIA-Microsoft ;
- M. Dominique Cardon, professeur de sociologie à l'Institut d'études politiques de Paris/Medialab ;

---

<sup>1</sup> Basé à New-York, il est chargé d'aider les sociétés du portefeuille d'ISAI à s'établir et se financer aux États-Unis et est également en charge des investissements d'ISAI outre-Atlantique. Après un passage chez PAI Partners puis chez Bain & Company, il passe 5 ans chez Equinox (centres de fitness haut de gamme, leader aux États-Unis) en tant que Directeur de la Stratégie. Il combine des activités de « business angel », de conseil et d'interim-management au sein de l'écosystème américain des *start-ups*.

- 
- M. Gilles Babinet, entrepreneur, digital champion auprès de la Commission européenne ;
  - M. Henri Verdier, directeur interministériel du numérique ;
  - Mme Marie-Claire Carrère-Gée, présidente du Conseil d'orientation pour l'emploi ;
  - M. Laurent Alexandre, entrepreneur (DNA vision) ;
  - M. Jean-Christophe Baillie, entrepreneur (Novaquark) ;
  - M. Jean-Claude Heudin, directeur de l'Institut de l'Internet et du multimédia ;
  - M. Gilles Dowek, directeur de recherche à l'INRIA et professeur attaché à l'ENS de Paris-Saclay ;
  - Mme Laurence Devillers, professeur à l'Université Paris-Sorbonne/LIMSI-CNRS ;
  - M. Serge Abiteboul, directeur de recherche à l'INRIA ;
  - M. Jean Ponce, professeur à l'École normale supérieure (ENS) ;
  - M. Serge Tisseron, psychiatre, chercheur associé à l'Université Paris Diderot-Paris VII ;
  - Mme Isabelle Falque-Pierrotin, présidente de la CNIL ;
  - M. Rand Hindi, membre du Conseil national du numérique, pilote du groupe de travail sur l'intelligence artificielle ;
  - M. Olivier Guilhem, directeur juridique chez SoftBank Robotics (ex-Aldebaran) ;
  - Me Alain Bensoussan, avocat, président de l'Association du droit des robots.

### **3. À Arcachon (lors d'un colloque du 26 au 30 septembre 2016, organisé par le CNRS)**

- M. Max Dauchet, Professeur émérite, Université de Lille, Président de la CERNA ;
- Mme Danièle Bourcier, Directeur de recherche émérite, CNRS, CERSA ;
- Mme Claire Lobet-Maris, Professeur, Université de Namur ;
- M. Alexei Grinbaum, Chercheur, CEA-Saclay, SPEC/LARSIM ;
- M. Raja Chatila, directeur de recherche CNRS, ISIR ;
- Mme Nozha Boujemaa, directeur de recherche Inria, conseillère du président pour le *big data* ;
- M. Jean-Gabriel Ganascia, Professeur UPMC et Institut de France, président du COMETS ;

- 
- M. Guillaume Piolle, enseignant chercheur à Supelec Rennes ;
  - M. François Pellegrini, Professeur et vice-président délégué au numérique de l'Université de Bordeaux, membre de la CNIL ;
  - Mme Nicole Dewandre, conseillère pour les questions sociétales, directeur général de CONNECT ;
  - Claude Kirchner, directeur de recherche Inria, président du COERLE ;
  - Sophie Vulliet-Tavernier, directrice des relations avec les publics et la recherche, CNIL ;
  - M. Nathanaël Jarrassé, chercheur CNRS à l'ISIR ;
  - Mme Laurence Devillers, professeur à l'Université de Paris IV, LIMSI ;
  - Mme Christine Balagué, titulaire de la Chaire Réseaux Sociaux de Télécom École de Management.

#### **4. À Paris lors d'un colloque sur l'intelligence artificielle à l'Assemblée nationale le 14 février 2017 organisé par la Commission supérieure du numérique et des postes (CSNP)**

- M. Jean-Yves Le Drian ministre de la Défense ;
- M. Jean Launay, député du Lot et président de la Commission supérieure du numérique et des postes (CSNP) ;
- M. Jean-Yves Le Déaut, député de Meurthe-et-Moselle, président de l'Office parlementaire d'évaluation des choix scientifiques et technologiques (OPECST) ;
- Lionel Tardy, député de la Haute-Savoie
- M. Laurent CELERIER, capitaine de vaisseau, commandement opérationnel de cyberdéfense (COMCYBER) ;
- Mme Nozha BOUJEMAA, directrice de recherche à l'INRIA ;
- M. Jean PONCE, professeur et directeur du département d'informatique de l'ENS ;
- M. Benjamin WERNER, directeur du département d'informatique de l'École Polytechnique
- M. Pierre-Jean BENGHOZI, professeur à l'École polytechnique ;
- M. Yves-Alexandre de MONTJOYE, professeur à l'Imperial College London
- M. Nikos PARAGIOS, professeur à CentraleSupélec ;
- M. Yann BONNET, Secrétaire Général - Conseil National du Numérique ;
- Mme Flora FISCHER, chargée de recherche au CIGREF ;
- M. Alain CLOT, président de FinTech France ;

- 
- M. Eric LEANDRI, président de QWANT ;
  - M. David BESSIS, président de Tinyclues ;
  - M. Konstantinos VOYIATZIS, vice-président de CIO – Edenred ;
  - Mme Vanessa HEYDORFF, directrice régionale France, Espagne et Portugal de Booking.com ;
  - M. Olivier CUNY, secrétaire général du groupe ATOS ;
  - M. Jean-Philippe DESBIOLLES, vice-président de Watson, IBM France ;
  - M. Hervé JUVIN, président de l'observatoire d'Eurogroup Consulting ;
  - M. Antoine BORDES, chercheur chez Facebook ;
  - M. Jean ROGNETTA, directeur de la rédaction de Forbes France ;
  - M. Eric SADIN, Ecrivain et Philosophe.

## B. À L'ÉTRANGER

### 1. À Genève, Suisse, les 21 et 22 septembre 2016

- Professeur Daniel Schneider, Faculté de psychologie et des sciences de l'éducation de l'Université de Genève, spécialiste de l'e-education ;
- Professeur Alexandros Kalousis, professeur à la Haute école spécialisée de Suisse occidentale et Université de Genève, spécialiste de *machine learning* ;
- Professeur Dominique Boullier, directeur du *SocialMedia Lab* de l'École polytechnique fédérale de Lausanne (EPFL), spécialiste de sociologie du numérique ;
- Professeur Didier Grandjean, enseigne la psychologie cognitive à l'Université de Genève, spécialiste des relations affectives hommes-machines ;
- M. Hervé Bourlard, directeur de l'Idiap (fondation affiliée à l'École polytechnique fédérale de Lausanne) ;
- M. Andrei Popescu-Belis, chercheur senior à l'Idiap et enseignant à l'École polytechnique fédérale de Lausanne ;
- Professeur Felix Schürmann, directeur adjoint du *Blue Brain project*. Professeur adjoint à l'École polytechnique fédérale de Lausanne (EPFL) ;
- Professeur Sean Hill, co-directeur du *Blue Brain project* (BBP) et co-directeur neuro informatique du *Humain brain project* (HBP) ;
- Professeur John Richard Walker, économiste, chercheur sénior sur le *Blue Brain Project* ;

- 
- Johann Christoph Ebell, directeur exécutif du projet HBP ;
  - Jean-Pierre Changeux, chercheur au HBP, École normale supérieure, Collège de France, Institut Pasteur ;
  - Bernd Stahl, directeur de l'éthique du HBP, directeur du centre de calcul et responsabilités sociales. Professeur à l'Université De Montfort (Royaume-Uni) ;
  - Christine Aircardi, chercheur associé au *Human Brain Project*, ingénieur de l'école nationale des ponts et chaussées.

## 2. À Londres, Royaume-Uni, les 14, 15 et 16 décembre 2016

- Madame l'Ambassadeur de France au Royaume-Uni, Son Excellence Sylvie Bermann ;
- Mme Natasha McCarthy, responsable de la stratégie, à la *Royal Society* sur le "*Machine Learning Project*" ;
- Mme Jessica Montgomery, conseillère stratégie (senior policy adviser), à la *Royal Society* sur le "*Machine Learning Project*" ;
- Mme Laura Wilton, conseillère stratégie (senior policy adviser), à la *Royal Society* sur le "*Machine Learning Project*" ;
- M. Paul Mason, directeur des technologies numériques et habilitantes (*Head of Digital and Enabling Technologies*), *Innovate UK* ;
- M. Phil Williams, responsable du groupe d'intérêt sur la robotique et les systèmes autonomes (*Robotics and Autonomous Systems Special Interest Group*), *Knowledge Transfer Network* ;
- M. Noel Sharkey, Université de Sheffield, Professeur de robotique et systèmes autonomes, co-fondateur de la *Foundation for Responsible Robotics* (Fondation pour une robotique responsable) ;
- M. Stephen Metcalfe, membre du Parlement, Président de la Commission pour la Science et la Technologie de la Chambre des Communes ;
- Dr Nico Guernion, responsable des partenariats (Français), *Alan Turing Institute* ;
- Dr Sandra Wachter, chercheuse en régulation et éthique des données, *Alan Turing Institute* ;
- Professeur Rose Luckin, éducation et nouvelles technologies, spécialiste en Intelligence Artificielle, *UCL Knowledge Lab, University College London* ;
- Dr Kedar Pandya, Directeur associé, *Engineering and Physical Sciences Research Council (EPSRC)* ;

- 
- Professeur Alan Winfield, Professeur de robotique, spécialiste des questions éthiques et sociétales, *University of the West of England Bristol* ;
  - Professeur Murray Shanahan, Professeur de robotique cognitive, *Imperial College London* ;
  - M. Arnaud Schenk, entrepreneur avec *First*, incubateur en *deep learning* ;
  - M. Bruno Marnette, entrepreneur, président de *prodo.ai* ;
  - M. Jacqueline I. Forien, fondatrice du salon du *Machine Learning*, professeur de mathématiques, master de *Machine Learning* ;
  - M. Alexandre Flamant, entrepreneur *Venture Capital* ;
  - M. Stéphane Chretien, chercheur au *National Physical Laboratory* ;
  - M. Gautier Marti, doctorant X-ENS, en contrat chez *Hellebore Capital Management*.

### 3. À Oxford, Royaume-Uni, le 15 décembre 2016

- Professeur Michael Wooldridge, Directeur du département de sciences informatiques, université d'Oxford ;
- Professeur Doyne Farmer, Co-Directeur, économies complexes, The Institute for New Economic Thinking (INET), Oxford Martin School ;
- Dr François Lafond, Postdoctorant ;
- Dr Rupert Way, Postdoctorant, INET ;
- Dr Anders Sandberg, chercheur, *Oxford Martin School Senior Fellow* et *Future of Humanity Institute (FHI)* ;
- Dr Stuart Armstrong, chercheur en intelligence artificielle et *machine learning*, *Future of Humanity Institute (FHI)* ;
- M. Miles Brundage, chercheur, *Strategic Artificial Intelligence Centre, Future of Humanity Institute (FHI)* ;

### 4. À Cambridge, Royaume-Uni, le 16 décembre 2016

- Dr Stephen Cave, Directeur exécutif du *Leverhulme Centre for the Future of Intelligence* ;
- Dr Huw Price, Directeur académique du *Centre for the Study of Existential Risk (CSER)*, et du *Leverhulme Centre for the Future of Intelligence* ;

- 
- M. Jens Steffersen, Administrateur, *Centre for the Study of Existential Risks (CSER)* ;
  - Susan Gowans, Administrateur, *Leverhulme Centre for the Future of Intelligence* ;
  - Pr Steve Young, Professeur en ingénierie de l'information (information engineering), Université de Cambridge, et membre de l'équipe de recherche d'Apple sur le logiciel *Siri* ; président de la *start-up* de reconnaissance vocale *VocalIQ* ;
  - Dr Sean Holden, chercheur, *Artificial Intelligence Group*, université de Cambridge ; représentant du *machine learning group*.

## 5. À Washington, D.C., États-Unis, le 23 janvier 2017

- Sénateur Ted Cruz, président de la sous-commission du Sénat américain sur l'espace, la science et la compétitivité (« *Chairman of the Senate Commerce Committee's Subcommittee on Space, Science and Competitiveness* ») ;
- Représentant John Delaney (D-Maryland), président du comité sur l'IA du Congrès des États-Unis (« *Chairman of the House AI Caucus* ») ;
- M. Jason Matheny, directeur de l'agence, Intelligence Advanced Research Projects Activity (IARPA) ;
- M. Andrew Borene, Intelligence Advanced Research Projects Activity (IARPA) Partner Engagement Lead ;
- Dr. Brian Pierce, directeur adjoint de l'innovation, Defense Advanced Research Projects Agency (DARPA) ;
- Mme Amanda Lloyd, chargée de l'international, Defense Advanced Research Projects Agency (DARPA)
- Mme Rebecca Keiser, directrice, Office of International Science & Engineering (OISE), National Science Foundation (NSF) ;
- M. Jim Kurose, directeur adjoint, Computer and Information Science and Engineering (CISE), National Science Foundation (NSF) ;
- M. Erwin Gianchandani, adjoint du Directeur de l'Informatique et de l'Ingénierie du Computer and Information Science and Engineering (CISE), National Science Foundation (NSF) ;
- Mme Lynne Parker, Directrice de la Division Information and Intelligent Systems (IIS), Computer and Information Science and Engineering (CISE), National Science Foundation (NSF).

Personnel de l'Ambassade de France aux États-Unis :



- 
- M. Gérard Araud, Ambassadeur de France aux États-Unis ;
  - Mme Minh-Hà Pham, conseillère pour la science et la technologie à l'Ambassade de France ;
  - M. Renaud Lassus, Ministre Conseiller pour les affaires économiques à l'Ambassade de France ;
  - M. Bernhard Hechenberger, Conseiller commercial, Service économique régional ;
  - M. Cameron Griffith, chargé de mission à l'Ambassade de France (Congressional Affairs Liaison) ;
  - M. Hervé Martin, attaché pour la science et la technologie à l'Ambassade de France ;
  - M. Xavier Morise, Directeur du bureau du CNRS à l'Ambassade de France ;
  - Mme Mireille Guyader, Directrice du bureau Inserm à l'Ambassade de France ;
  - M. Norbert Paluch, Conseiller Spatial et Représentant du CNES à l'Ambassade de France ;
  - M. Christophe Barre, Attaché pour la Technologie et l'Innovation au Service économique régional à l'Ambassade de France.

## **6. À Boston, États-Unis, les 24 et 25 janvier 2017**

- Pr. Alex "Sandy" Pentland, professeur d'informatique, chaire Marvin Minsky, MIT Media Lab ;
- Pr. David H. Autor, co-director of the School Effectiveness and Inequality Initiative economist and professor of economics, MIT ;
- Pr. Iyad Rahwan, professeur d'informatique, MIT ;
- M. Richard Stallman, Free Software Foundation, MIT ;
- Mme Latanya Arvette Sweeney, Professor of Government and Technology Harvard Data Privacy Lab, , Formerly Chief Technology officer at SEC ;
- M. David Parkes, professeur d'informatique et responsable du département d'informatique, Harvard University, School of Engineering and Applied Science ;
- M. Nicolas Mialhes co-fondateur du projet "The Future Society", Harvard Kennedy School, fondateur de la "AI Initiative" ;
- M. Richard Mallah, director of AI projects, Future of Life Institute (FLI), Cambridge Innovation Center, MIT ;



- M. Joseph Aoun, Président de North-Eastern University (Boston) ;
- Pr. Thomas A. Kochan, professor of industrial relations, work and employment, MIT Sloan School of Management ;
- Mme Chloe Hecketsweiler, journaliste au journal Le Monde, Fellow au MIT ;
- M. Jean-François Guillous, chercheur chez IBM pour le projet Watson ;
  
- M. Valery Freland, Consul Général de France, et son conjoint, M. Laurent Colomines ;
- Jean-Jacques Yarmoff, attaché pour la Science et la Technologie au Consulat de Boston.

## 7. À San Francisco, États-Unis, du 25 au 28 janvier 2017

- M. Philippe Perez, attaché pour la Science et la Technologie au consulat de France à San Francisco.

### a) À San Francisco

- M. Alex Dayon, directeur des produits Salesforce ;
- M. John Ball, directeur du département Einstein de Salesforce ;
- M. Thierry Donneau-Golencer, directeur de la recherche en intelligence artificielle chez Salesforce (projet Einstein) ;
- M. Gregory Renard, entrepreneur (XBrain), ambassadeur de la Fondation Sigfox et expert pour la NASA ;
- M. Akli Adjaoute, entrepreneur (Brighterion) ;
- M. Michel Morvan, entrepreneur (Cosmo) ;
- M. Ben Levy, fondateur de BootstrapLabs ;
- M. Nicolas Pinto (ex-startup Perceptio), chercheur en intelligence artificielle chez Apple ;
- M. François Chollet, chercheur chez Google et créateur de Keras et de « AI-ON » (Artificial Intelligence Open Network) ;
- M. Fabien Beckers, entrepreneur (Arterys) ;
- M. Johan Mathé (ex-Google X), chercheur chez Bay Labs ;
- M. Nicolas Poilvert (ex-startup Perceptio), chercheur chez Bay Labs ;
- M. Reza Malekzadeh (Partech Ventures).

---

*b) À l'Université de Berkeley*

- M. Michael Jordan, professeur d'informatique à l'université de Berkeley ;
- M. Stuart Russell, professeur d'informatique à l'université de Berkeley ;
- M. Russ Salakhutdinov, professeur d'informatique à l'Université Carnegie Mellon, directeur de la recherche en intelligence artificielle chez Apple ;
- Mme Marion Fourcade, professeur de sociologie à l'université de Berkeley, chercheuse en sociologie de l'innovation et en économie numérique ;
- Mme Valérie Issarny, directrice de recherches à l'INRIA, basée à l'Université de Berkeley et chargée des relations INRIA - États-Unis ;
- M. Ahmed El Alaoui, chercheur en doctorat sous la direction de Michael Jordan, professeur d'informatique à l'université de Berkeley.

*c) À l'Université de Stanford et dans la Silicon Valley*

- M. Oussama Khatib, directeur du centre de robotique, Université de Stanford ;
- Mme Fei-Fei Li, directrice du centre d'intelligence artificielle, Université de Stanford ;
- M. Greg Corrado, directeur de la recherche en intelligence artificielle chez Google, co-fondateur de l'équipe Google Brain ;
- M. Alex Acero, directeur du projet Siri chez Apple ;
- Mme Catherine H. Foster, directrice des affaires publiques chez Apple ;
- M. Larry Zitnick, directeur de la recherche en intelligence artificielle au centre Menlo Park de Facebook ;
- M. Michael Kirkland, directeur de la communication technologique de Facebook ;
- Mme Molly Jackman, directrice des affaires publiques chez Facebook ;
- Mme Brenda Tierney, responsable des relations avec les pouvoirs publics chez Facebook ;
- M. Jack Clark, responsable de la stratégie et de la communication, fondation « Open AI » créée par Elon Musk ;
- Mme Cathy Olsson, directrice de recherche à la fondation « Open AI ».

## 8. À Bruxelles, Belgique, 8 et 9 février 2017

- Mme Juha Heikkila chef de l'unité robotique et intelligence artificielle, Commission européenne, et Mme Cécile Huet, son adjointe ;
- M. Jean Arthuis, eurodéputé et président de la commission des Budgets du Parlement européen ;
- Mme Mady Delvaux, eurodéputée, rapporteure sur l'intelligence artificielle et la robotique ;
- Professeur Hugues Bersini, Université libre de Bruxelles-ULB ;
- M. Erastos Filos, de l'unité RTD.D2 « Systèmes de production avancés et Biotechnologies » ;
- M. Pascal Rogard, conseiller technologie à la RP ;
- Mme Catelijne Muller, rapporteure du Conseil économique et social européen (CESE) sur l'intelligence artificielle ;
- M. Pegado Liz, président du groupe d'étude du CESE sur l'intelligence artificielle ;
- Mme Marie-Laurence Drillon, administratrice à la section "Marché intérieur, production et consommation" du CESE.

## **ANNEXE 4 : PROGRAMMES DES DÉPLACEMENTS (ÉTATS- UNIS, ROYAUME-UNI, BELGIQUE, SUISSE, ARCACHON POUR UN SÉMINAIRE SUR L'ÉTHIQUE DE L'INTELLIGENCE ARTIFICIELLE)**

Chronologie des déplacements :

- Du 21 au 22 septembre 2016 : déplacement en Suisse à Genève ;
  
- Du 26 au 30 septembre 2016 : déplacement à Arcachon pour un séminaire sur l'éthique de l'intelligence artificielle organisé par la CERNA ;
  
- Du 13 au 16 décembre 2016 : déplacement au Royaume-Uni : Londres, Cambridge et Oxford ;
  
- Du 22 au 29 janvier 2017 : déplacement aux États-Unis : Washington DC, Boston et San Francisco ;
  
- Du 8 au 9 février 2017 : déplacement à Bruxelles.

## I. DÉPLACEMENT AUX ÉTATS-UNIS

---

## II. DÉPLACEMENT AU ROYAUME-UNI



---

### III. DÉPLACEMENT EN BELGIQUE



## IV. DÉPLACEMENT EN SUISSE

## V. DÉPLACEMENT À ARCACHON POUR UN SÉMINAIRE SUR L'ÉTHIQUE DE L'INTELLIGENCE ARTIFICIELLE

## ANNEXE 5 : PANORAMAS DE L'INTELLIGENCE ARTIFICIELLE

### I. AUX ÉTATS-UNIS

*Note réalisée avec le concours du service scientifique de l'Ambassade de France aux États-Unis et en particulier du Consulat de San Francisco*

S'il fallait un indice de la prise de conscience collective aux États-Unis de l'importance critique du développement de l'intelligence artificielle dans ses perspectives scientifiques et technologiques mais également et peut-être surtout économiques et sociétales, le fait que le Président Barack Obama ait choisi de consacrer la dernière de ses grandes apparitions publiques consacrées à la science et à la technologie à un entretien extrêmement médiatisé avec Joi Ito, le directeur du prestigieux MIT Media Lab, publié dans le plus grand magazine de vulgarisation technologique du monde, et entièrement dédié à une discussion autour de l'intelligence artificielle<sup>1</sup> constitue probablement un assez bon indicateur de ce que cette thématique technologique est le sujet brûlant du moment aux États-Unis, que ce soit dans l'expression de politiques publiques, la consolidation d'une puissance de recherche inégalée ou le développement rapide et sans précédent d'activités économiques liées.

#### **1. Les politiques publiques liées au développement de l'intelligence artificielle**

Consciente de la tension entre l'impact potentiellement bénéfique des progrès technologiques dans le champ de l'intelligence artificielle et de leur percolation dans un nombre croissant de champs d'activité économique d'une part, et de ses potentielles implications structurantes sur la société, relayées publiquement de surcroît par diverses personnalités en vue du monde scientifique et industriel, d'autre part, la Maison Blanche a mis en place en mai 2016 un sous-comité spécifique au sein du *National Science and Technology Council* (NSTC), chargé de suivre les évolutions du secteur et de coordonner les activités fédérales sur le sujet. En parallèle ont été tenues quatre sessions de travail publiques entre les mois de mai et juillet 2016, visant à engager la discussion avec le grand public et surtout à produire une large évaluation des opportunités, risques, et implications réglementaires et sociales de l'intelligence artificielle, de même qu'une série de recommandations dotées d'un plan stratégique afin de se donner les moyens

---

<sup>1</sup> "Barack Obama, neural nets, self-driving cars and the future of the world", *entretien du 24 août 2016 dans WIRED*. <https://www.wired.com/2016/10/president-obama-mit-joi-ito-interview/>

---

de les mettre en œuvre. Les conclusions de ces travaux ont donné lieu à pas moins de trois rapports produits par l'administration Obama entre octobre et décembre 2016.

### 1.1 *Preparing the Future of Artificial Intelligence* (NSTC, octobre 2016)

Ce premier rapport, assez peu détaillé, a pour objectif de produire un cadre général à la réflexion nationale, en amorçant un état des lieux de la recherche et des applications actuelles tout en posant des jalons initiaux quant à la possibilité et la nature d'une régulation.

En particulier, le rapport conclut, avec un optimisme affiché, que l'intelligence artificielle et l'apprentissage automatique font émerger le potentiel d'améliorer la vie des citoyens en permettant de résoudre certains des plus grands enjeux sociétaux. La santé, les transports, l'environnement, la justice ou encore l'efficacité du gouvernement lui-même sont autant de secteurs applicatifs mis en avant.

Ceci étant, les auteurs insistent sur l'importance de la réglementation dans l'accompagnement de ces avancées technologiques. Si un consensus se dégage pour conclure qu'une réglementation générale de la recherche en IA semble inapplicable à l'heure actuelle, et, partant, la réglementation actuelle est pour l'heure suffisante, l'approche explicite guidant les prochains arbitrages se devra de procéder selon le principe suivant : évaluer les risques que l'implémentation de l'IA pourrait réduire, de même que ceux qu'elle pourrait augmenter. L'analyse comparée des risques et des bénéfices permettra de justifier les futures décisions d'ordre réglementaire. En outre, le rapport insiste sur l'importance d'ajuster les réglementations afin de réduire, au maximum, les coûts et les barrières à l'innovation sans mettre en danger la sécurité du public ou la compétition équitable sur le marché.

Ce n'est pas pour autant diminuer le rôle pivot des autorités fédérales. De ce point de vue, le gouvernement a plusieurs rôles à jouer : un rôle d'organisateur du débat public et d'arbitre des mesures à mettre en place à l'échelle du pays ; un rôle de suivi attentif de la sécurité et de la neutralité des applications développées ; un rôle de producteur de réglementations encourageant l'innovation tout en protégeant le public ; un rôle d'accompagnateur de la diffusion de ces technologies tout en protégeant certains secteurs au besoin afin d'éviter des contrecoups économiques dévastateurs ; un rôle de soutien et de financeur de projets de recherche faisant avancer le domaine ; et enfin un rôle d'adoption en son sein même de ces avancées afin d'assurer un service public de meilleure qualité

En particulier, dans cette position, le gouvernement fédéral a un rôle clé à jouer dans l'avancement de l'IA en termes de R&D, en particulier à

---

travers la production d'une main d'œuvre en nombre suffisant, ainsi que d'un haut niveau de qualification et de diversité technique, non seulement d'un point de vue professionnel, mais également, et de manière plus large, du point de vue de la formation générale de la population.

En termes d'impact économique, le rapport conclut que le premier effet à court terme sera celui de l'automatisation d'un nombre grandissant de tâches. Si la productivité générale et la création de richesse ont toutes les chances d'être positivement impactées, l'impact sur l'emploi est lui plus difficile à évaluer. Ce dernier sera inégal en fonction des professions, et le sous-comité suggère que les professions à faible salaire seront les plus touchées, et que l'automatisation aggrave le fossé entre travailleurs hautement formés et travailleurs faiblement éduqués, s'accompagnant potentiellement d'une augmentation des inégalités économiques. Dans cette optique, le rôle du gouvernement serait ici d'assurer le maintien de certaines catégories de travailleurs pouvant être considérés comme complémentaires aux machines automatisées, plutôt que concurrents. En sus, les politiques publiques pourront s'efforcer d'assurer un partage général des bénéfices économiques engendrés.

### 1.2 The National Artificial Intelligence Research & Development Strategic Plan (NSTC, octobre 2016)

Ce plan stratégique, publié en octobre 2016, présente une série d'objectifs à poursuivre pour la recherche en IA financée par des fonds publics fédéraux, avec l'ambition de couvrir à la fois les efforts de recherche directement produits par des entités fédérales et ceux des organisations externes, au premier rang desquelles les universités, tout en essayant de mettre l'accent sur les domaines dans lesquelles l'industrie est moins susceptible d'intervenir. Sept priorités en constituent le cœur stratégique :

- Investissement fédéral soutenu dans la recherche à long terme afin de produire des percées scientifiques et technologiques dans les dix prochaines années, en particulier dans les nouvelles méthodologies nécessaires à la découverte de savoir à l'heure des grands volumes de données, l'amélioration des capacités de perception des systèmes d'intelligence artificielle, la compréhension profonde des capacités théoriques et des limites propres au développement de l'intelligence artificielle, la poursuite des efforts visant au développement d'une IA générique par opposition aux intelligences artificielles spécifiques, la nécessité de développer des systèmes d'intelligence artificielle extensible, l'accélération de la recherche sur une intelligence artificielle s'approchant des capacités humaines.
- Développement de méthodes de collaboration entre homme et intelligence artificielle : systèmes fonctionnant en parallèle, en

---

substitution au moment où l'individu atteint ses limites cognitives ou en substitution totale.

- Compréhension accrue des implications légales, sociétales et éthiques, en particulier développement de méthodes permettant de concevoir des systèmes d'intelligence artificielle conformes aux principes éthiques, légaux et sociétaux des États-Unis. Plus précisément, le rapport insiste sur l'importance d'assurer la justice, la transparence et la responsabilité des systèmes, dès la phase de conception.
- Sécurité et sûreté des systèmes, en particulier dans l'adaptation à des environnements complexes et/ou incertains.
- Développement de bases de données publiques partagées et d'environnements pour l'entraînement et le test de systèmes d'intelligence artificielle, à commencer par la mise à disposition de jeux de données fédéraux existants.
- Développement d'un spectre large de standards visant à assurer que les technologies émergentes répondent à des objectifs précis en termes de fonctionnalité et d'interopérabilité, ainsi qu'en termes de sécurité et de performance.
- Évaluation des besoins nationaux en termes de main-d'œuvre.

1.3 Artificial Intelligence, Automation and the Economy (Executive Office of the President, décembre 2016)

Dans la foulée des deux précédents, l'*Executive Office of the President* vient de produire un troisième rapport, cosigné par plusieurs branches, et centré sur les implications économiques, en particulier sur le marché du travail, du développement de l'intelligence artificielle.

S'il en souligne comme toujours les bénéfices économiques attendus, il soulève la question de la répartition de ces bénéfices, en évoquant en particulier le risque que cette répartition asymétrique contribue de manière décisive à creuser les inégalités entre les travailleurs hautement qualifiés et les autres. Bien qu'il soit hasardeux d'essayer de prédire exactement quels emplois seront le plus tôt impactés par l'automatisation conduite par l'intelligence artificielle (aussi bien en termes de destruction d'emplois que de création de nouveaux emplois, par exemple de supervision de l'intelligence artificielle), le rapport pointe que, d'après les experts, l'ampleur des volumes d'emplois directement menacés dans les décennies à venir serait comprise entre 9 et 47% du volume total d'emplois, le plus largement concentrés au sein des groupes de travailleurs les moins diplômés et les moins bien rémunérés, ce qui implique qu'un des effets directs de cette vague d'automatisation pour ce groupe résidera dans la pression vers le bas des salaires. Le rapport va même jusqu'à pointer très explicitement la menace selon laquelle l'automatisation en question pourrait bien ne

bénéficier qu'à quelques-uns, de par la nature hautement hégémonique (« *winner-takes-most* ») des marchés liés aux technologies de l'information.

Ceci étant, le rapport insiste par contraste sur la capacité des politiques publiques et des incitations institutionnelles à réguler les effets du changement technologique (« *Technology is not destiny* »). Il formule dès lors un certain nombre de propositions stratégiques dont il estime qu'il est encore plus important qu'elles soient mises en œuvre au regard du contexte d'émergence massive à moyen terme de l'intelligence artificielle et de l'automatisation :

- Investir dans et développer l'intelligence artificielle en reconnaissant les bénéfices, à la fois en termes de croissance de la productivité et de volonté de rester à la pointe de l'économie de l'innovation, en menant des politiques volontaristes d'accès à cette économie, en particulier du point de vue de la diversité.
- Éduquer et former les Américains aux emplois du futur, que ce soit de manière précoce, dans la capacité des politiques d'enseignement à rendre accessible financièrement l'enseignement postsecondaire, ou dans l'apprentissage tout au long de la carrière professionnelle.
- Soutenir les travailleurs en transition (en particulier en accroissant la disponibilité des dispositifs de sécurisation de l'emploi et de la transition vers l'emploi, mais également l'accroissement des salaires).

## **2. Quelques repères sur la recherche en intelligence artificielle aux États-Unis**

L'objectif de cette section est de proposer quelques repères sur la recherche en intelligence artificielle aux États-Unis, sans volonté d'exhaustivité dans la présentation de ce paysage scientifique.

### 2.1 Quelques grands centres

L'excellence de certaines universités comme le MIT, Stanford et Harvard n'est plus à démontrer et celles-ci abritent des équipes et des chercheurs qui jouent un rôle important dans la recherche en IA.

#### **MIT : Computer Science and Artificial Intelligence Laboratory (CSAIL)**

Le Massachusetts Institute of Technology (MIT) possède avec le CSAIL un centre dédié à l'IA qui est remarquable de par la forte diversité dans les thématiques traitées, le nombre et la qualité de ses équipes de recherche. L'ensemble des cinquante équipes regroupe mille deux cents personnes dont environ trois cents chercheurs, huit cents étudiants (six cents *graduate students*, deux-cents *undergraduate students*). Par rapport à la démographie usuelle en France, il est intéressant de noter la proportion

---

beaucoup plus forte d'étudiants (pour ces thématiques, la répartition usuelle en France est plutôt de 50 % de chercheurs et 50 % d'étudiants).

### **Carnegie Mellon University (CMU)**

Carnegie Mellon University (CMU) est une université moins connue mais pourtant essentielle dans le paysage de l'informatique avec ses douze récipiendaires du prix Turing (par comparaison, la France n'en compte qu'un seul) qui est l'équivalent du prix Nobel pour l'informatique. Avec son institut de Robotique, qui a été le premier créé aux États-Unis en 1970, CMU a joué et joue toujours un rôle central dans ce domaine. Cet institut abrite notamment le *Field Robotic Center* qui conçoit des robots capables de travailler dans des conditions extrêmes (explorations polaires et explorations spatiales par exemple) et le *NanoRobotics lab* dont l'objet d'étude est centré sur les nanorobots avec notamment des applications en médecine très prometteuses.

### **Stanford Artificial Intelligence Laboratory (SAIL)**

Pour illustrer la forte interaction entre le monde académique et le monde industriel, on peut noter qu'en lien avec le SAIL, l'université de Stanford a créé un laboratoire conjoint avec Toyota, le *Stanford-Toyota Research Center*, dont l'objet d'étude n'est pas exclusivement le véhicule autonome mais de manière plus large, le développement des interactions entre l'homme et les machines intelligentes. Ce laboratoire pluridisciplinaire regroupe vingt et un chefs de projets qui travaillent sur douze thématiques liées à l'intelligence artificielle<sup>1</sup>.

### **Berkeley Artificial Intelligence Research (BAIR)**

Beaucoup plus petit que le CSAIL du MIT (vingt-cinq professeurs pour une centaine d'étudiants diplômés), ce laboratoire couvre les principaux domaines de l'IA (vision par ordinateur, apprentissage automatique, traitement automatique des langues et robotique). Ce laboratoire se caractérise par l'excellence scientifique de certains de ces membres dont quelques professeurs emblématiques comme Michael Jordan et Stuart Russell, mais aussi le très prometteur Sergey Levine qui applique avec succès des techniques d'apprentissage automatique en robotique<sup>2</sup>. On notera que l'Université de Californie à Berkeley vient de lancer un Centre de Recherche dédié aux interactions entre l'homme et l'intelligence artificielle (*Centre for Human-Compatible Artificial Intelligence - CHCAI*) piloté par Stuart Russell.

## 2.2 Quelques acteurs importants aux États Unis

Les États-Unis ont une capacité d'attractivité exceptionnelle et ont réussi à attirer plusieurs grands scientifiques de la discipline.

---

<sup>1</sup> <http://aicenter.stanford.edu/research/#people>

<sup>2</sup> <https://www.technologyreview.com/lists/innovators-under-35/2016/pioneer/sergey-levine/>



---

### 2.2.1 Les pionniers

**Alan Turing (1912-1954)** : d'origine britannique, il est considéré comme le père fondateur de l'informatique avec ses contributions fondamentales sur la décidabilité. Il est resté célèbre dans le monde de l'intelligence artificielle par sa proposition du test de Turing qui vise à déterminer si une machine est dotée d'intelligence artificielle. Ce test qui consiste à tester la capacité de confondre un ordinateur avec un interlocuteur humain est aujourd'hui remis en cause pour prendre en compte la grande variété des critères et caractéristiques d'un être ou d'une machine intelligente.

**John McCarthy (1927-2011) et Marvin Lee Minsky (1927-2016)** : John McCarthy a été professeur à Stanford et au MIT. Ce mathématicien est à l'origine du terme « Intelligence Artificielle » et l'inventeur du langage de programmation LISP qui était très populaire du temps des systèmes experts. Ses travaux autour de la logique symbolique sont essentiels en théorie des jeux (il est à l'origine de l'élagage alpha-bêta permettant de réduire l'espace d'exploration dans un arbre de possibilités). Il est avec Marvin Minsky l'un des organisateurs de l'école d'été à Dartmouth où le concept d'IA a émergé. Marvin Minsky est un spécialiste des sciences cognitives, co-fondateur du laboratoire d'intelligence artificielle au MIT. Il fut notamment le créateur du premier calculateur basé sur les réseaux de neurones.

### 2.2.2 Les contemporains

**Les trois prix Turing américains en IA : Edward Feigenbaum (1994), Leslie Valiant (2010) et Judea Pearl (2011)** : Après des études à l'université Carnegie-Mellon, Edward Feigenbaum est devenu professeur d'informatique et codirecteur scientifique du *Knowledge Systems Laboratory* à l'université Stanford. Il a obtenu le Prix Turing pour ses travaux dans les domaines de l'étude et la construction de systèmes d'intelligence artificielle à grande échelle. Judea Pearl était professeur à New York University (NYU) et a reçu le prix Turing pour ses travaux sur les calculs de probabilités et les approches bayésiennes en Intelligence Artificielle. Leslie Valiant est un britannique, professeur à Harvard, et reconnu notamment pour son modèle PAC en théorie de l'apprentissage.

**Les récipiendaires de IJCAI/AAAI (principales conférences du domaine), Michael I. Jordan et Barbara J. Grosz** : Michael I. Jordan est professeur à Berkeley et lauréat du prestigieux prix IJCAI en 2016 pour ses travaux en apprentissage automatique. Michael I. Jordan est également membre de l'académie des sciences et a obtenu la chaire d'excellence de la Fondation des Sciences Mathématiques de Paris en 2012. Barbara J. Grosz est professeur à l'université d'Harvard et spécialiste du Traitement Automatique des langues et des systèmes multi-agents. Elle a reçu le prix IJCAI en 2015 et le prix de l'ACM/AAAI Allen Newell en 2009.

---

**Une responsable de la société savante, Manuela Veloso** : Manuela Veloso est professeur à CMU et préside la principale société savante en IA, *l'Association for the Advancement of Artificial Intelligence (AAAI)*. Cette scientifique qui travaille dans le domaine des multi-agents défend une approche baptisée « *Symbiotic Autonomy* » qui est basée sur la collaboration entre humains et robots en fonction du contexte.

**Des scientifiques impliqués dans les enjeux sociétaux liés à l'IA, de Gary Marcus à Stuart Russell, en passant par Elon Musk** : Gary Marcus est professeur de psychologie à New York University et est connu pour ses romans et aussi son leadership dans le mouvement pour remplacer le test de Turing par une série de tests baptisée « *Turing Olympics* »<sup>1</sup>. On ne présente plus Elon Musk : ce Sud-Africain d'origine est devenu créateur d'entreprises innovantes (Tesla Motors, SpaceX) et symbolise parfaitement la puissance d'innovation technologique américaine. Il est le cofondateur et le coprésident d'OpenAI<sup>2</sup> qui vise à promouvoir *l'open source* en Intelligence Artificielle. La promotion du logiciel libre est particulièrement importante pour la recherche mais aussi pour l'émergence de *start-ups* en particulier en France dans le domaine. Quant à Stuart Russell, professeur à Berkeley, est un scientifique de renom qui explore différentes facettes de l'intelligence dans ses travaux. Il est notamment connu pour ses contributions sur l'apprentissage par renforcement hiérarchique et ses réflexions sur les impacts sociétaux de l'IA, notamment le danger potentiel des armes autonomes. Stuart Russell a été titulaire d'une chaire Blaise Pascal de la Région Île-de-France de 2012 à 2014.

### 2.3 Quelques tendances dans différents domaines de l'Intelligence Artificielle

**Vision par ordinateur (*Computer Vision*)** : La principale tendance dans ce domaine de recherche est caractérisée par le passage de méthodes dites supervisées (nécessitant une intervention humaine) à des méthodes peu supervisées voire non supervisées. L'objectif est de pouvoir indexer des quantités de données de plus en plus grandes. L'indexation consiste à proposer des algorithmes pour extraire des éléments sémantiques caractéristiques liés à la perception telle que la reconnaissance d'objets, la détection de séquences dans des vidéos, la description de contenus (personnes, activités, mouvements, ...). Ce domaine est essentiel pour faire face aux défis scientifiques posés par le *big data*.

**Apprentissage automatique (*Machine Learning*)** : Ici encore, comme le souligne Yann LeCun, la principale tendance est le développement de l'apprentissage non supervisé. Une autre tendance forte est la recherche sur la compréhension par des non-experts des conclusions, actions et décisions retournées par les programmes d'IA. La DARPA vient de lancer un

---

<sup>1</sup> <http://vision.stanford.edu/Science-2015-Turing.pdf>

<sup>2</sup> <https://www.openai.com/about/>

---

programme spécifique sur cette thématique<sup>1</sup>. Ceci implique des défis scientifiques mais aussi des problèmes pour former les générations futures à ces techniques, pour les démythifier et les comprendre. Un autre défi important pour ce domaine est souligné dans le rapport du département de la défense US<sup>2</sup>. Il s'agit, notamment pour les technologies d'apprentissage profond (*deep learning*), de satisfaire aux propriétés souhaitables pour tout logiciel critique : preuve, maintenance, évolutivité, robustesse. Ces aspects posent des problèmes scientifiques complexes et nécessitent des collaborations entre des communautés scientifiques distinctes. L'importance de l'accès aux données et les enjeux économiques pour ces thématiques liées à l'apprentissage automatique fait que le secteur industriel économique et notamment les GAFA (Google, Apple, Facebook, Amazon) américains proposent une recherche de pointe dans ces domaines.

**Robotique** : la robotique est un secteur qui a également énormément évolué dans la dernière décennie, d'une part, en raison des progrès technologiques notamment sur les capteurs et, d'autre part, en raison des enjeux économiques et sociétaux qui ont suscité de nombreux investissements. Côté US, on peut citer les travaux du MIT sur le robot guépard<sup>3</sup> mais aussi le robot humanoïde proposé par Google / Alphabet et Boston Dynamics<sup>4</sup> capable de courir dans les bois. Concernant les perspectives, nul doute que les nanorobots seront au cœur de nombreux progrès scientifiques et technologiques dans les prochaines années.

#### 2.4 Les États-Unis, un acteur majeur de la recherche en IA

Les États-Unis sont ainsi, comme d'ailleurs dans la plupart des domaines scientifiques, un acteur majeur de la recherche en Intelligence Artificielle. Le principal atout du pays est sans doute sa capacité à :

- attirer les meilleurs scientifiques et les talents les plus prometteurs,
- proposer de grands programmes et laboratoires thématiques en lien avec ses agences (NSF<sup>5</sup>, NASA<sup>6</sup>, DARPA<sup>7</sup>, NIH, ...) et
- promouvoir l'innovation à travers un écosystème très efficace impliquant le monde académique et le monde industriel.

Il est également important de souligner la perméabilité des domaines scientifiques qui fait que l'innovation scientifique et technologique se retrouve souvent à la convergence de plusieurs domaines. L'IA a ainsi

---

<sup>1</sup> <http://www.darpa.mil/program/explainable-artificial-intelligence>

<sup>2</sup> Randolph Franklin

<sup>3</sup> [https://www.youtube.com/watch?v=\\_luhn7TLfWU](https://www.youtube.com/watch?v=_luhn7TLfWU)

<sup>4</sup> <https://www.theguardian.com/technology/video/2015/aug/18/humanoid-robot-atlas-boston-dynamics-run-video>

<sup>5</sup> <https://www.nsf.gov/funding/programs.jsp?org=IIS>

<sup>6</sup> <https://ti.arc.nasa.gov/tech/dash/physics/quail/>

<sup>7</sup> <http://www.darpa.mil/tag-list?tt=73&type=Programs>

---

bénéficié des progrès récents dans le domaine du *big data* et du calcul haute performance (HPC).

Nul doute que la convergence des mondes numérique et physique avec le développement de l'internet des objets et des véhicules autonomes sera la source de nombreux progrès où l'intelligence artificielle jouera un rôle majeur.

D'autres domaines comme la cybersécurité sont au cœur de la politique scientifique dans le numérique aux États-Unis avec des connexions fortes dans le monde de l'intelligence artificielle pour détecter et contrer les attaques.

Dans tous ces domaines les collaborations scientifiques entre la France et les États-Unis sont une réalité. INRIA possède par exemple quarante équipes associées entre des équipes françaises et américaines. Le CNRS possède un Groupe de recherche international (GDRI) impliquant les États-Unis dans le domaine de l'intelligence artificielle et un laboratoire commun (UMI) avec Georgia Tech.

### **3. La course industrielle à l'armement**

Naturellement, les promesses de croissance économique ouvertes par cette nouvelle rupture technologique ne pouvaient laisser indifférent l'écosystème des grands acteurs privés de l'innovation aux États-Unis.

#### 3.1 Les reconfigurations internes des géants industriels

Il est notable que 2016 aura été l'année durant laquelle les géants industriels technologiques, notamment de la côte ouest, ont opéré une série de réorganisations internes destinées à mettre en avant des services de *cloud computing* spécialement conçus pour les avancées en intelligence artificielle.

C'est ainsi que par exemple Google/Alphabet vient de recruter Fei-Fei Li, directrice des *Artificial Intelligence and Vision Labs* de Stanford pour diriger son groupe d'apprentissage automatique dans le *cloud*, lequel inclura rapidement des services en reconnaissance d'image, reconnaissance vocale, traduction automatique ou compréhension du langage naturel. Microsoft, Amazon ou IBM ont tous engagé des restructurations similaires.

Une des raisons pour focaliser ce type d'effort réside dans la prédiction selon laquelle, le *cloud* étant promis à constituer une part dominante de la génération de leur chiffre d'affaires, l'infusion de cette partie de leur activité par les technologies d'intelligence artificielle et d'apprentissage automatique constitue un avantage comparatif décisif.

---

Plus encore, il s'agit pour ces sociétés de faire percoler ces technologies dans l'ensemble des divisions et produits associés des susdites sociétés. Microsoft vient par exemple de créer une division entière au sein de laquelle son vice-président exécutif pour la recherche et la technologie, Harry Shum, conduit plus de cinq mille ingénieurs et informaticiens à travailler à la diffusion des technologies d'intelligence artificielle dans l'ensemble des produits de la société (le moteur de recherche Bing, l'assistant personnel Cortana ou les activités robotiques de la société en constituent les exemples les plus notables). Comme souvent, Google/Alphabet avait été un précurseur en la matière avec sa division Google Brain (portée par Andrew Ng, désormais à Baidu) qui travaille avec l'ensemble des autres divisions de Google/Alphabet. De la même manière, un ingénieur de Facebook sur cinq travaille désormais à partir des résultats en apprentissage automatique du pourtant récent *Applied Machine Learning Group*.

Il n'est d'ailleurs pas inintéressant de noter que la rareté relative des talents en intelligence artificielle (à l'inverse de ceux en programmation classique) conduit les grands industriels technologiques de la Silicon Valley à injecter des ressources significatives dans la formation (et la rétention) de ses employés : Google/Alphabet a monté un programme de formation interne en apprentissage automatique, et Facebook offre des formations similaires tout en construisant des parcours internes de transfert vers des missions à plein temps de recherche dans le domaine exclusif de l'intelligence artificielle.

### *3.2 Les stratégies d'acquisition*

Une autre voie d'immersion pour les entreprises industrielles réside naturellement dans le déploiement d'une stratégie d'acquisition de sociétés plus petites, d'autant qu'elle s'accompagne souvent d'une volonté de rétention des personnels clés de ces dernières (ce qui se désigne par le terme d'*acqui-hire*).

Les chiffres parlent d'eux-mêmes : on passe d'un volume global d'acquisition de *start-ups* spécialisées en intelligence artificielle de sept en 2012 pour tout le territoire américain, à trente-trois en 2014, puis trente-sept en 2015 et quarante-deux simplement pour les trois premiers trimestres de 2016.

Là encore, Google/Alphabet est à la fois un précurseur en la matière, puisqu'il a acquis dès 2013 la *start-up* d'apprentissage automatique issue de l'Université de Toronto DNNresearch, puis surtout en 2014 pour 600 M\$ la société britannique DeepMind Technologies, connue pour avoir enfanté un programme capable de battre le meilleur joueur mondial de go, et celui qui pratique cette stratégie avec le plus d'intensité et de constance (onze acquisitions au total, dont les plus récentes sont la *start-up* de recherche visuelle Moodstock et la plateforme de langage naturel Api.ai).

---

Mais ses concurrents directs ne sont pas en reste : Intel a acquis cinq *start-ups* (dont trois en 2016), de même qu'Apple (deux en 2016), Twitter quatre (dont la très prometteuse *start-up* britannique de traitement d'image Magic Pony pour 150M\$) et Salesforce, qui a démarré son effort en 2016, trois.

De manière intéressante, on voit désormais apparaître dans la course des entreprises industrielles non initialement liées dans leur cœur technologique à l'informatique et à l'Internet, le meilleur exemple en étant General Electric (deux acquisitions en novembre 2016)

### 3.3 A l'autre bout du spectre, l'expansion de l'écosystème de start-ups en intelligence artificielle

Dans un marché de l'investissement en capital-risque sinon atone, du moins plafonnant après les records de l'année précédente, les *start-ups* liées à l'intelligence artificielle et à l'apprentissage automatique connaissent *a contrario* une envolée spectaculaire de leur attractivité financière. Le troisième trimestre 2016 est à cet égard archétypal : une croissance impressionnante tant en volume global (705 M\$) qu'en nombre d'accords (71), soit une hausse respective de seize et 22 % par rapport au trimestre précédent, avec une tendance qui continue à accélérer.

L'élément le plus frappant réside cependant sans doute dans le degré inégalé de transversalité sur des verticales sectorielles par les *start-ups* liées à l'intelligence artificielle, lesquelles attaquent désormais tous les marchés : le langage, la vision, l'automobile, la cybersécurité, la robotique, le marketing, l'intelligence économique, la génération de texte, l'Internet des objets, le commerce, les technologies financières, l'assurance, l'agriculture ou l'imagerie satellitaire. Mais la diversité de champs applicatifs au sein de ces marchés est tout aussi spectaculaire : si on considère le secteur qui a généré ces deux dernières années le plus d'investissement en capital-risque pour des *start-ups* en intelligence artificielle, celui de la santé, on trouve aussi bien des sociétés qui travaillent sur la gestion du risque que sur la recherche génétique, la nutrition, l'imagerie médicale, le diagnostic, le monitoring individuel, la gestion des urgences hospitalières, la découverte de médicaments, l'oncologie ou la santé mentale. L'intelligence artificielle est donc bien devenue un marché total.

## II. EN CHINE

*Note réalisée avec le concours du service scientifique de l'Ambassade de France en Chine*

L'**intelligence artificielle** (IA) est une discipline scientifique recherchant des méthodes de résolution de problèmes à forte complexité logique ou algorithmique. Par extension elle désigne, dans le langage courant, les dispositifs imitant ou remplaçant l'humain dans certaines mises en œuvre de ses fonctions cognitives.

Historiquement, elle trouve son point de départ avec les travaux de Turing dans les années 1950 et a suivi l'évolution de l'informatique. Le développement des technologies informatiques et en particulier l'explosion récente des puissances de calculs et de la capacité de capter et manipuler des quantités importantes de données ont permis, à l'aide d'algorithmes théorisés dans les années 70, la réalisation de programmes informatiques surpassant l'homme dans certaines de ses capacités cognitives emblématiques.

Le sujet se trouve à l'interface de nombreuses disciplines. L'informatique est au centre mais en lien avec les modèles mathématiques sous-jacents – réseaux et statistiques pour l'analyse de données massives –, la biologie et en particulier les neurosciences, au moins comme source d'inspiration pour affiner les architectures, les sciences sociales comme les sciences cognitives ou la linguistique, mais aussi souvent les sciences liées aux applications concernées, de la ville intelligente à la médecine en passant par l'éducation ou l'automobile.

La Chine se trouve au cœur de l'évolution récente de ces techniques et déploie d'importants moyens pour devenir leader dans le domaine.

Parmi ses atouts, mentionnons l'importance de la part du PIB consacrée à la recherche et la mise en avant de la recherche autour de l'IA ; sa position dominante dans le classement du TOP 500 (deux supercalculateurs en tête de cette liste) ; la collusion État / géants du net et de l'informatique / instituts de recherche / universités / *start-ups* ; un marché très important et friand des avancées potentielles du secteur ; une opinion publique peu préoccupée par les questions philosophiques soulevées ni par les questions concernant la protection des données qui sont des enjeux importants pour l'évolution et les applications de l'IA.

Ces avantages et les succès chinois annoncés en la matière sont à nuancer par certains handicaps : politique de communication parfois un peu arrogante et certaines fois déconnectée de la réalité (que ce soit pour la hauteur effective des financements ou la qualité des résultats obtenus) ; choix parfois guidés par une volonté plus de coller aux « indicateurs » que de faire progresser la science ; la plupart des progrès visibles reposent encore sur des architectures conçues par des scientifiques occidentaux ; existence de

multiples réseaux parfois concurrents se livrant à une compétition pas toujours très constructive (cela oblige d'ailleurs pour se faire une idée à identifier les différents réseaux) ; la pluridisciplinarité qui semble requise pour des avancées sérieuses en IA n'est pas dans la tradition scientifique chinoise ; et puis, mais ce n'est pas propre à la Chine, il y a une nette tendance à qualifier d'IA tout algorithme un peu original et pas trop bête.

### Généralités

Le 13<sup>ème</sup> plan quinquennal comprend une liste de quinze « nouveaux grands projets - innovation 2030 » qui structurent les priorités scientifiques du pays et correspondent chacun à des investissements de plusieurs milliards d'euros. Parmi ces quinze projets, on trouve un projet de « Recherche sur le cerveau » et des projets d'ingénierie intitulés « Mega données », « Réseaux intelligents » et « Fabrication intelligente et robotique ». Si le terme « Intelligence Artificielle » n'apparaît pas explicitement, il est clairement sous-jacent dans ces quatre projets, ce qui signifie qu'un effort considérable va être développé dans cette direction.

À la manière locale, certains chercheurs présentent pour la communauté scientifique chinoise une feuille de route ambitieuse :

1) Actuellement, développement de machines ayant une intelligence par domaine ("*Artificial Narrow Intelligence*" - ANI) (reconnaissance d'objets, compréhension du langage naturel, assistant personnel, ...). Ces techniques pouvant être appliquées dans un grand nombre de domaines, l'objectif est que la Chine soit leader mondial (en 2020 ?).

2) Vers 2025-2030, développement de systèmes ayant une intelligence générale, comparable à celle du cerveau humain (« *Artificial General Intelligence* » - AGI). Ici aussi, l'ambition de la Chine est d'être la première à disposer d'un tel système.

3) Puis le développement d'intelligences suprahumaines. C'est d'après ces chercheurs la seule solution pour résoudre les problèmes de la planète - les intelligences humaines, trop limitées, ne peuvent appréhender ce type de problème dans leur ensemble, et donc sont incapables de proposer des solutions efficaces.

Les recherches envisagées ne sont pas seulement théoriques, les applications multiples sont un moteur important : détection des émotions, interface homme-machine, analyse d'images, contrôle de drones, de robots ou d'avatars, interface en langage naturel, automobiles autonomes, etc. Les systèmes permettront aussi de faire l'analyse des *big data*. L'exemple suivant ne semble pas gêner les collègues qui l'évoquent : « Piloté par le gouvernement et l'organisme central de planification, le dispositif de notation de la population devrait récupérer automatiquement les



---

informations sur les citoyens d'ici 2020. Il scrutera les activités en ligne, etc, pour générer un score individuel. Il semble que si un seuil est dépassé, l'individu concerné se verra privé d'un certain nombre de droits et de services ». Une expérimentation a lieu à Suining ([https://www.washingtonpost.com/world/asia\\_pacific/chinas-plan-to-organize-its-whole-society-around-big-data-a-rating-for-everyone/2016/10/20/1cd0dd9c-9516-11e6-ae9d-0030ac1899cd\\_story.html?utm\\_term=.71ca38d649e1](https://www.washingtonpost.com/world/asia_pacific/chinas-plan-to-organize-its-whole-society-around-big-data-a-rating-for-everyone/2016/10/20/1cd0dd9c-9516-11e6-ae9d-0030ac1899cd_story.html?utm_term=.71ca38d649e1) ). D'autres dispositifs d'évaluation sont étudiés, comme le Sesame Credit du distributeur en ligne Alibaba (<http://www.bbc.com/news/world-asia-china-34592186> ).

Les réseaux du Service pour la Science et la technologie de ce Poste diplomatique ont pu identifier quelques unités saillantes dans le paysage mouvant travaillant autour de ce sujet, en se reposant essentiellement sur les retours de quatre missions (une mission Découverte Chine sur l'informatique neuro-mimétique, une mission d'experts sur l'IoT, une mission d'experts sur le calcul haute performance et une mission d'experts sur le *big data*).

## Cerveau

Ainsi la recherche en intelligence artificielle est en partie portée par le projet « cerveau » (qui comporte bien sûr par ailleurs une importante composante plus strictement étiquetée « sciences de la vie »), à travers le pôle « *brain-like/ brain-inspired computing* » qui est en train de se constituer. Il implique en particulier trois acteurs majeurs de la recherche : la Chinese Academy of Sciences (CAS), l'Université Tsinghua et l'Université de Pékin (Beida) qui pour l'instant ont développé leurs propres réseaux. Le programme « *CASIA brain* » est une collaboration entre six instituts de l'Académie des sciences chinoise qui implique, entre autres, l'institut d'automatique (CASIA) et l'institut de technologie informatique (ICT-CAS). De son côté, l'université Tsinghua pilote le *Center for Brain-Inspired Computing Research* (CBICR) qui relie des partenaires dans toute la Chine et se présente comme résolument « transdisciplinaire » au moins au sein de l'informatique. Le groupe de l'Université de Pékin semble moins structuré nationalement mais entretient d'importantes coopérations avec les pays anglo-saxons. Dans tous ces cas, l'idée directrice est de travailler simultanément sur des aspects *hardware* et logiciel pour concevoir des circuits susceptibles de fonctionner « comme le cerveau » pour apprendre à résoudre les problèmes posés.

Dans la veine de l'informatique neuro-mimétique, on relève aussi des travaux intéressants à Hangzhou (Projet Darwin, *College of Computer Science*, Université de Zhejiang) et à Shanghai (*Center for Brain Like Computing and Machine Intelligence* de l'Université Shanghai Jiaotong).

## *Big data*

---

Parfois au sein des mêmes institutions, les laboratoires d'informatique (leur département « logiciel »), parfois au sein de pôles constitués autour des enjeux dits « *big data* » ou « IoT » développent des outils et intègrent des méthodes avec des objectifs de même nature – du moins pour les objectifs à court terme.

Outre les laboratoires des grandes institutions pékinoises (*CAS Key Lab of Network Data Science and Technology* ou *Institute for Data Science Tsinghua*), on relève le laboratoire *Pattern Recognition & Intelligent System Lab.* et le laboratoire du logiciel de télécommunication intelligent et multimédia de la *Beijing University of Post and Telecommunications (BUPT)*, la *School of Computer Software* à Tianjin, le *National Key Laboratory for Novel Software Technology* de l'Université de Nankin, le *Shenzhen Institute of Advanced Technology* de la CAS, ou encore le *Shanghai Advanced Research Institute* de la CAS (centré sur l'IoT), l'*Institute of Media Computing* de l'Université Fudan, également à Shanghai.

### Géants privés

Les enjeux sont nationaux et les financements suivent. Mais les entreprises privées sont aussi des moteurs puissants du secteur.

En tête, l'entreprise *Baidu* (qui a développé le moteur de recherche chinois, site le plus consulté en Chine et 5ème au niveau mondial, indexant près d'un milliard de pages, cent millions d'images et dix millions de fichiers multimédia) communique beaucoup sur le sujet, consacre une part conséquente de sa recherche au sujet (*Institute of deep learning, Big data lab, ...*) et, comme les géants américains, dispose d'un flux de données permettant d'envisager des applications dans de nombreux domaines.

L'entreprise considère, comme ses concurrents, que l'IA est son prochain challenge comme solution clé pour des applications en vision, parole, traitement du langage naturel et sa compréhension, génération de prédictions et de recommandations, publicité ciblée, planification et prise de décision en robotique, conduite autonome, pilotage de drones,... Elle travaille en étroite relation avec de nombreuses universités et *start-ups*.

Les résultats algorithmiques de l'*Institute of Deep Learning* de Baidu sont impressionnants et du meilleur niveau mondial, malgré son existence récente. Mentionnons pour donner une idée, la capacité de leurs algorithmes à retrouver une image dans une base de données de 10 milliards d'images en moins d'une seconde. Ils ont aussi de bonnes performances sur les *benchmarks* ICDAR (1<sup>er</sup> sur 5/8 évaluations parmi 4 tâches).

Sur FDDB (*Face Detection Data Set and Benchmark*, voir <http://www.cs.umass.edu/fddb/>) et sur la base de données de visages LFW, ils progressent vite (8 % d'erreurs en décembre 2015, 2,3 % en septembre 2016 et

---

bientôt 1 %). L'entreprise annonce aussi la meilleure précision sur la collection de benchmarks KITTI (<http://www.cvlibs.net/datasets/kitti/>) orientés pour la conduite de voitures autonomes. Baidu développe aussi des applications de reconnaissance d'image pour la plateforme de services *Baidu Nuomi* : une application permet par exemple de reconnaître le restaurant (et le plat) en prenant une photo de nourriture dans un restaurant.

Il n'est pas évident d'évaluer la maîtrise théorique des ingénieurs mais l'entreprise montre une incontestable efficacité pour implémenter rapidement les dernières innovations du secteur.

Les autres géants chinois du net, comme Alibaba (distribution) ou Tencent (réseaux sociaux), tirent eux aussi dans la même direction : développement et diffusion grand public d'applications plus ou moins convaincantes mais manifestement exploitant des techniques d'IA un peu évoluées, même s'ils semblent moins présents dans la recherche « fondamentale ». Huawei (telecom, téléphones), entreprise qui accorde une grande importance à la recherche fondamentale et dont la recherche s'internationalise rapidement (pôle mathématique implanté en France il y a deux ans) a mis en avant début janvier 2017 un concept de téléphone intelligent, dont on ne peut encore savoir s'il ira effectivement plus loin que ceux de ses concurrents.

Plus globalement concernant les données, les pôles universitaires peuvent aussi compter sur le support des industriels comme, par exemple, National Grid, China Mobile, China Unicom, China Mobile, Shanghai Meteorological Bureau ou Environmental Monitoring Center.

### **Innovation et *start-ups***

Le paysage chinois de l'innovation - incitation à très grande échelle pour soutenir l'innovation privée, en particulier par la multiplication de petites structures - conjugué à l'effet de mode « attractif » du secteur IA fait éclore dans tout le pays et en particulier aux abords des (voire dans les) universités, des entreprises cherchant à exploiter ce filon, souvent en proposant des services originaux ; sous l'œil (attentif et bienveillant) de l'État et des géants. Ces initiatives ne sont pas toutes scientifiques. Il s'agit souvent d'abord de marketing, mais pas toujours ou pas seulement.

On peut mentionner par exemple la société Cambricon, très proche de l'ICT de la CAS : le professeur Chen Yunji continue de développer des accélérateurs neuronaux (à l'ICT de la CAS) de la famille DianNao à partir de travaux réalisés en coopération avec des chercheurs du CEA (Marc Duranton) et de l'INRIA (Olivier Temam) ; son frère Chen Yianshi est maintenant le P-DG d'une *start-up* « [Cambricon Technologies Corporation Limited](#) » fondée en mars 2016, comptant 100 employés en septembre 2016 et ayant un objectif de 200 pour fin 2016, qui travaille sur des accélérateurs neuronaux inspirés fortement de la famille DianNao. Le projet de la société

---

est d'utiliser ces accélérateurs pour permettre le traitement intelligent des images et des signaux embarqués (au lieu d'architectures GPU / DSP actuelles) améliorant les performances énergétiques d'un facteur 4000. Le chercheur Marc Duranton qui a eu accès (limité) à la société estime que cette perspective est prometteuse.

Sur le marché des robots de service, intégrant de l'IA en particulier pour la partie reconnaissance vocale, mentionnons par exemple Yunji Technology (robots déployés dans des hôtels à Pékin et Shanghai), iFLYTEK (moteur de reconnaissance vocale) ou Ninebot Inc (Transport intelligent).

Certaines sociétés françaises du secteur, comme par exemple *Air Visual* (suivi et prédiction de la pollution de l'air) ou *WosomTech* (reconstruction 3D prédictive ; approchée par *Lenovo* récemment) ont une expérience en Chine et connaissent un peu le paysage de l'intérieur.

### **Personnes ressources à privilégier**

Mentionnons tout d'abord les trois principaux acteurs de ce qui devrait constituer le pôle « brain-like/inspired computing » du projet de recherche sur le cerveau Il s'agit de HUANG Tiejun, SHI Luping et XU Bo, travaillant respectivement à l'Université de Pékin, à l'Université Tsinghua et à l'Institut d'Automatique de la CAS.

Le NLPR (*National Laboratory of Pattern Recognition*) dirigé par TAO Jinhua est l'un des laboratoires de l'Institut d'Automatique de la CAS. Il est notable qu'il héberge le principal laboratoire franco-chinois du domaine, le LIAMA (Laboratoire d'Informatique, d'Automatique et de Mathématiques Appliquées) du domaine. L'Intelligence Artificielle, et en particulier ce qui tourne autour du « *brain-inspired computing* » est l'un des sujets mis en exergue lors de la dernière rencontre organisée par le LIAMA (Pékin, Octobre 2016).

Le Professeur TAO Jinhua a encadré au NPLR la thèse d'un étudiant qui a obtenu récemment un « one-million prize » décerné par l'entreprise *Baidu* - pour son travail sur la reconnaissance vocale et la production de parole. L'une de ses étudiantes actuelles travaille sur la reconnaissance des émotions dans la voix humaine. C'est nous semble-t-il un contact précieux dans ce secteur.

ZHOU Zhihua est directeur adjoint de l'équipe LAMDA (Learning And Mining from DATA), affiliée au National Key Laboratory for Novel Software Technology et au Department of Computer Science & Technology. C'est un chercheur qui a aujourd'hui une véritable envergure internationale et qui est une valeur montante de l'apprentissage statistique en Chine.

Les Hub French Tech récemment labellisés à Pékin, Shanghai et Shenzhen peuvent constituer d'excellents points d'entrée pour accéder aux *start-ups* françaises du secteur ayant des intérêts en Chine.

QIU Xipeng : vice dean du Big Data School de l'Université de Fudan.

---

### III. AU JAPON

*Note réalisée avec le concours du service scientifique de l'Ambassade de France au Japon*

L'intelligence artificielle est considérée comme l'élément clé de la révolution numérique au Japon. Le gouvernement japonais en a fait le socle de sa nouvelle stratégie en science, technologie et innovation, qui vise à mettre en place une « société 5.0 », société super-intelligente et fer-de-lance à l'échelle mondiale. Il a annoncé une vague d'investissements massifs dans le domaine, avec 27 milliards de yens pour la seule année 2016, à travers la création de centres de recherche et technologies dédiés à ce domaine. Les partenariats publics-privés entre ces nouveaux centres et les grands groupes japonais se multiplient, afin d'exploiter le potentiel de création de valeur que constitue l'intelligence artificielle sur des applications ciblées.

#### I. L'Intelligence artificielle, au coeur de la stratégie japonaise

##### **1-L'intelligence artificielle (AI) dans le « 5e Plan-cadre de la science et de la technologie » et dans le « Livre blanc 2016 sur la science et la technologie »**

Les plans-cadre pour la science et de la technologie, élaborés tous les cinq ans depuis 1996 par le Conseil pour la Science, Technologie et Innovation (CSTI, dépendant directement du Cabinet Office, i.e. Les services du Premier Ministre japonais), définissent les priorités de l'Etat dans les domaines de la science, de la technologie et de l'innovation.

Entré en vigueur le 1er avril 2016 (le début de l'année fiscale 2016), le 5e Plan-cadre<sup>1</sup> repose sur un nouveau concept de « Société 5.0 », société ultra-intelligente, qui tirerait le meilleur bénéfice des opportunités offertes par les technologies numériques afin de redynamiser le secteur industriel, mais également dans l'optique de réformer la structure même de la société et assurer une prospérité inclusive, dans laquelle les citoyens bénéficieraient de services de haute qualité et adaptés à leurs besoins selon leur âge, sexe, région ou langue. Pour mettre en oeuvre cette société fer-de-lance à l'échelle mondiale, le Japon compte tirer profit de l'expansion très rapide des technologies de l'information et de la communication, parfois considérée comme la « 4e révolution industrielle », et s'appuyer sur trois piliers : l'Internet des objets (IoT), le *big data* et l'intelligence artificielle (IA).

L'intelligence artificielle ouvre la voie à des développements très attendus dans le domaine des transports, de la communication, de la

---

<sup>1</sup> [http://www8.cao.go.jp/cstp/english/basic/5thbasicplan\\_outline.pdf](http://www8.cao.go.jp/cstp/english/basic/5thbasicplan_outline.pdf)

---

traduction automatique ou de la robotique, notamment à l'horizon des Jeux Olympiques et paralympiques de 2020, que le Japon envisage comme une vitrine technologique pour se présenter comme LE pays leader en termes d'innovation.

Suite à la publication du 5ème Plan-cadre pour la Science et la Technologie par le CSTI, le Ministère japonais de l'Éducation, de la Culture, des Sports, des Sciences et de la Technologie (MEXT) a rédigé son Livre blanc 2016 pour la science et la technologie, intitulé « Vers la société ultra-intelligente mise en œuvre par l'IoT, le *big data* et l'IA- pour que le Japon soit un précurseur mondial », adopté en Conseil des ministres le 20 mai 2016. Ce rapport signale l'insuffisance au Japon de la formation et de la recherche fondamentale en informatique, qui devraient venir soutenir le développement des technologies de l'information et de la communication.

Enfin, à l'occasion de l'organisation du G7 au Japon, le Japon a pris l'initiative d'organiser une réunion ministérielle G7 dédiée aux Sciences et technologies de l'information et de la communication (format qui n'avait pas été mis en œuvre depuis 20 ans). Suite à la tenue de cet événement en avril à Takamatsu, les ministres en charge des Sciences et Technologies de l'Information (ICT) se sont accordés dans leur déclaration conjointe à promouvoir la R&D pour les technologies ICT, notamment dans le domaine de l'Intelligence artificielle. Ils ont mis l'accent sur l'importance de mettre en place des politiques adaptées, permettant de prendre en compte l'impact sociétal et économique colossal de ces technologies.

## **2-Mise en place de deux comités pour le suivi de ces stratégies**

### *Création d'un Comité de délibération sur l'IA et la société humaine*

Le 12 avril 2016, à l'issue de la réunion régulière du Conseil des ministres, Mme Aiko Shimajiri, Ministre chargée de la politique de la science et de la technologie, a annoncé la mise en place d'un « **Comité de délibération sur l'AI et la société humaine** » au sein du Cabinet Office. Il s'agit de la première entité gouvernementale ayant pour mission d'étudier les enjeux liés à l'IA selon cinq points de vue, à savoir : l'aspect éthique, l'aspect légal, l'aspect économique, l'aspect social et l'aspect R&D. Ce Comité, présidé par Mme Yuko HARAYAMA, membre exécutif permanent du CSTI du Cabinet Office, est composé de 11 experts. La première réunion s'est tenue le 30 mai 2016 en présence de la ministre Shimajiri. Le Comité s'est réuni sur une base mensuelle jusqu'en septembre 2016 pour analyser les activités nationales et internationales et approfondir la problématique. Il a choisi de se baser sur des cas d'application précis, mettant en œuvre des technologies qui devraient voir le jour à court terme : le véhicule autonome, l'automatisation de l'appareil de production et la communication homme/machine. Il souhaite également engager un débat avec le grand

---

public (par le biais essentiellement de séminaires ouverts et de questionnaires en ligne).

Le Comité devra remettre avant la fin de l'année 2016 ses conclusions, qui seront prises en compte dans la nouvelle Stratégie globale pour la Science, Technologie et Innovation qui sera publiée en juin 2017. Les discussions au niveau international en la matière sont prévues à partir de 2017. Mme Harayama a présenté à Paris le 17 novembre 2016, dans le cadre du *Technology Foresight Forum 2016* de l'OCDE<sup>1</sup> dédié à l'intelligence artificielle, les premières réflexions du comité :

- Ethique : Le citoyen peut-il accepter d'être manipulé pour modifier ses sentiments, convictions ou comportements, et d'être catégorisé/évalué, sans en être informé? Quel impact aura le développement de l'IA sur notre sens de l'éthique et les relations entre les hommes et les machines ? Dans la mesure où elle étend notre temps, notre espace et nos sens, est-ce que l'IA viendra affecter notre conception de l'humanité, notamment notre conception des facultés et des émotions humaines? Comment évaluer les actions et la création à partir de l'IA ?

- Légal : comment trouver le juste équilibre entre les bénéfices du traitement des *big data* par l'IA, et la protection des informations personnelles ? Est-ce que les cadres légaux existants pourront s'appliquer aux nouvelles problématiques soulevées par l'IA ? Comment clarifier la responsabilité dans le cas d'incidents impliquant de l'intelligence artificielle (par exemple pour le véhicule autonome) ? Quels sont les risques d'utiliser l'IA ? De ne pas utiliser l'IA ?

- Économique : comment l'IA va-t-elle changer notre manière de travailler ? Quelle politique nationale mettre en place pour favoriser l'utilisation de l'IA ? Comment l'IA va-t-elle modifier le monde de l'emploi ?

- Sociétal : Comment réduire les divisions liées à l'IA et répartir de manière équitable le coût social de l'IA ? Y-a-t-il une pathologie ou des conflits de société que peut potentiellement engendrer l'IA ? Peut-on assurer la liberté d'utiliser ou non l'IA et assurer le droit à l'oubli ?

- Éducation : Quelle politique nationale mettre en place pour faire face aux inégalités que ne manqueront pas de causer l'utilisation l'IA dans le domaine de l'éducation ? Comment développer notre capacité à exploiter l'IA ?

- R&D : Quelle R&D développer pour l'IA en respect de : l'éthique, la sécurité, la protection de la vie privée, la transparence, la contrôlabilité, la visibilité, la responsabilité ? Comment rendre disponible l'information liée à l'IA de manière à ce qu'un utilisateur puisse prendre la décision d'utiliser ou non l'IA ?

---

<sup>1</sup> <http://www.oecd.org/sti/ieconomy/Yuko%20Harayama%20-%20AI%20Foresight%20Forum%20Nov%202016.pdf>



---

Le comité cherche notamment à approfondir trois voies pour définir des politiques adaptées : la co-évolution de la société et de la technologie ; la recherche d'un équilibre entre les bénéfices (services personnalisés à coût abordable) et les risques liés à l'IA (discrimination, perte de protection des données à caractère privé, perte d'anonymat) ; la définition des limites de la prise de décision automatisée.

### *Un Conseil dédié à la stratégie des technologies liées à l'IA*

Par ailleurs, dans le cadre du « Dialogue privé-public pour les investissements du futur », le Premier Ministre Abe a présenté son projet de définir d'ici la fin de l'année la feuille de route présentant les objectifs de la recherche sur l'AI et ses applications industrielles et de mettre en place un « **Conseil de la stratégie des technologies liées à l'AI** ».

Ce Conseil, dont la réunion inaugurale a eu lieu le 18 avril 2016, est présidé par M. Yuichiro Anzai, Président de la JSPS (agence de financement de la recherche du MEXT dédiée à la recherche fondamentale), accompagné d'un conseiller, M. Kazuo Kuyma, membre exécutif permanent du CSTI, et composé de deux représentants du Keidanren (syndicat patronal des entreprises japonaises), des présidents de deux universités (Université de Tokyo et Université d'Osaka) et des présidents de cinq grands instituts de recherche et agences de financement : le NICT (*National Institute for Information and Communication Technologies*, dépendant du Ministère des affaires intérieures et communication, MIC), le RIKEN (principal centre de recherche pluridisciplinaire du MEXT), l'AIIST (*Advanced Institute for Science and Technologies*, centre de recherche multidisciplinaire du Ministère de l'Economie et de l'Industrie, METI), la JST (l'une des deux agences de financement du MEXT, orientée vers les projets de recherche appliquée) et la NEDO (agence de financement dépendant du METI).

Ce Conseil servira de quartier général de la R&D des technologies de l'AI et de leurs applications industrielles en regroupant les trois ministères impliqués dans l'AI : le MIC, le MEXT et le METI.

Deux Comités ont été mis en place sous ce Conseil : le « Comité de collaboration de recherche » (conseil des présidents des instituts de recherche et des agences de financement) et le « Comité de collaboration industrielle » qui se réunissent chacun mensuellement.

## **II. Mise en oeuvre : les 3 ministères impliqués mettent en oeuvre leurs propres initiatives**

Dans ce contexte, les trois ministères, respectivement en charge de la recherche, de l'industrie et des communications (MEXT, METI et MIC), se sont mobilisés avec une vitesse étonnante pour développer des initiatives

dans ce domaine. Ils ont chacun annoncé la création d'un centre de recherche sur l'intelligence artificielle en 2016.

- Le **Ministère japonais de l'Éducation, de la Culture, des Sports, des Sciences et de la Technologie** (MEXT) a officiellement lancé son centre, le « AIP Center» (*Advanced Integrated Intelligence Platform Project Center*), le 1er avril 2016, hébergé par le RIKEN (principal centre de recherche pluridisciplinaire du MEXT) et dirigé par le Professeur Masashi Sugiyama. Doté d'un budget de 1,45 milliards de yens pour l'année fiscale 2016, ce centre vise à construire une approche intégrée des technologies d'intelligence artificielle/ *big data* /IoT et cybersécurité. L'acceptation du terme intelligence artificielle est prise au sens large, incluant la composante informatique (algorithme, réseaux, architecture matérielle), mais également mathématique (algèbre linéaire, probabilités et statistiques, optimisation), ingénierie (intégration dans des systèmes) et applications dans différents domaines scientifiques (biologie, physique, science des matériaux...). Les activités s'articulent autour de 5 objectifs :

- Développer des technologies pour l'intelligence artificielle « fondamentale » (basé essentiellement sur le *machine learning* et le *deep learning*, avec un effort portant sur un apprentissage robuste et en temps réel sur des données pouvant être bruitées, hétérogènes, incomplètes...) ;

- Contribuer à l'accélération de la recherche scientifique (analyse et synthèse automatique d'articles scientifiques, brevets, résultats d'expériences, design efficaces de molécules et matériaux...) ;

- Contribuer à des applications concrètes à fort impact sociétal : problématique des soins dans le contexte du vieillissement de la population, gestion des infrastructures (inspection des fissures dans les ouvrages de génie civil tel que les ponts, etc...) résilience aux catastrophes naturelles ;

- Prise en compte des aspects éthiques, légaux et sociaux ;

- Développement des ressources humaines.

Ce centre (pour lequel un nouveau bâtiment est en cours de construction, au centre de Tokyo), dirigé par le professeur Masahi Sugiyama de l'Université de Tokyo, est composé de 28 personnes, réparties entre deux groupes de recherche, l'un dédié aux technologies génériques, et l'autre aux technologies dédiées à certaines applications. Il cherche à accroître ses effectifs, notamment en accueillant des chercheurs étrangers pour des durées longues. Par ailleurs, une vingtaine d'entreprises, dont Toyota Motors, NEC, *Sony Computer Science Laboratories*, NTT ou *start-up* spécialisée dans le *deep learning Preferred Networks* ont annoncé la mobilité de plusieurs de leurs chercheurs dans ce centre. Le centre devrait rapidement monter en puissance, avec un budget de 10 millions de yens qui sera octroyé par le MEXT pour l'année 2017, et plusieurs dizaines de millions de yens attenus de la part des partenaires privés.

- Le MEXT a également mis en place un laboratoire virtuel réunissant les projets de recherche financés par la JST (l'une des deux

---

agences de financement du MEXT, orientée vers les projets de recherche appliquée) sur l'intelligence artificielle, dit « **Network Labo** », dirigé par le Professeur Setsuo Arikawa, ancien président de l'Université de Kyushu.

Cinq projets de recherche sur l'AI sont déjà financés par la JST, pour une somme totale de 2,85 milliards de yens pour l'année fiscale 2016. La JST mettra en place trois nouveaux projets additionnels dans ce domaine, dotés d'un montant de 1,15 milliards de yens en 2016. Ces trois nouveaux projets relèvent respectivement de la médecine (identification de symptômes par traitement de données massives d'imagerie médicale), de la conduite autonome (par traitement de mégadonnées issues de caméras et de radars à extrêmement haute fréquence installés dans les véhicules) et de la cybersécurité. Ces huit projets en intelligence artificielle seront regroupés et pilotés via le « Network Labo », ce qui permettra de mutualiser certaines des technologies de base développées dans le cadre de ces projets et de diversifier les applications des résultats de recherche.

Le « **Network Labo** » et l'**AIP Center** collaborent étroitement afin mutualiser leurs ressources matérielles et humaines.

- **Le Ministère de l'Economie et de l'Industrie (METI)** a lancé également son centre dédié à l'IA le 1er mai 2015 : l'**AIRC** (*Artificial Intelligence Research Center*), hébergé par l'**AIST** (*Advanced Institute for Science and Technologies*, centre de recherche multidisciplinaire du METI). Il regroupe 348 personnes en 2016, dont 269 chercheurs (parmi lesquels 95 chercheurs permanents, 16 détachés d'universités, 17 détachés d'entreprises et 41 professeurs étrangers invités) et repose sur 3 piliers : plateformes de données, compétences en informatique et « *business models* » clairs sur des applications ciblées (axées sur la production industrielle, notamment la robotique industrielle ; les grands secteurs à fort impact sociétal et économique tels que la santé, le tourisme.. ; les disciplines scientifiques utilisant de grandes masses de données).

Le centre est financé par METI et son agence de financement la NEDO pour un montant de 19,5 milliards de yens pour l'année 2016, ainsi que par des contrats directs avec des acteurs industriels sur des applications ciblées (l'AIRC a notamment mis en place un laboratoire commun avec l'entreprise NEC).

L'AIRC vise à intégrer l'approche orientée données (*machine learning, deep learning*, modèles statistiques) et l'approche orientée « connaissance », utilisant une ontologie du domaine d'application considéré. Il mène également des recherches sur le cerveau humain. Enfin, il porte un effort particulier sur les infrastructures de calcul (il va notamment faire l'acquisition d'un superordinateur d'une puissance de 130 PetaFlops : l'*AI Bridging Cloud Infrastructure* (ABCI) en 2017, pour un financement de 19.5 milliards de yens en provenance du METI).

- **Le Ministère des affaires intérieures et communication (MIC)** a créé un cluster de recherche dédié à l'intelligence artificielle, hébergé par son centre de recherche, le NICT (*National Institute for Information and*

*Communication Technologies*). Ce cluster de recherche<sup>1</sup> se focalise sur l'IA inspirée par le cerveau humain, la reconnaissance vocale, la traduction automatique multi-langue ou l'analyse de la connaissance sociale. Il est doté d'un budget de 2,2 milliards de yens pour l'année 2016.

Si les trois ministères ont mis en place leurs centres de recherche en propre, ils se sont engagés à coordonner leurs actions, via le Conseil dédié à la stratégie des technologies liées à l'IA, à organiser des symposiums communs sur une base annuelle, à mettre en place un site web commun, et à mutualiser leurs ressources informatiques.

### **III. Les acteurs industriels investissement massivement dans le domaine**

On estime que le marché de l'intelligence artificielle au Japon devrait passer de 3,7 milliards de yens en 2015 à 87 milliards en 2030, dont 30,5 milliards de yens dans le domaine du transport et 42 milliards en incluant la production industrielle pour le transport.

Plusieurs grands groupes japonais<sup>2</sup> ont lancé de grands programmes ou laboratoires dédiés à ce domaine, pour un montant d'investissement devant atteindre 300 milliards de yens sur les trois ans à venir.

De nombreux projets sont développés en perspective des Jeux Olympiques et Paralympiques de 2020, qui verront un afflux de visiteurs étrangers à Tokyo, motivant le développement de systèmes de traduction automatique, tels que ceux développés par Panasonic (qui se positionne également sur la robotique, la reconnaissance d'image et la reconnaissance spatiale, en collaboration avec le NICT et plusieurs universités japonaises, ainsi que dans ses centres de R&D situés à Singapour et dans la *Silicon Valley*) ou le *chatbot* multilingue de KNT-CT. Tokyo sera à cette occasion également la cible privilégiée d'attaques physiques ou cyber, ce qui incite Fujitsu Laboratories développer des systèmes intelligents de protection

<sup>1</sup> Voir les pp. 6 et 7 : [http://www.nict.go.jp/en/data/nict-news/NICT\\_NEWS\\_1606\\_E.pdf](http://www.nict.go.jp/en/data/nict-news/NICT_NEWS_1606_E.pdf)

<sup>2</sup> Les exemples donnés peuvent être approfondis sur les sites suivants :  
<http://asia.nikkei.com/Business/Trends/Japan-tech-giants-double-down-on-AI>  
<http://news.panasonic.com/global/stories/2016/45102.html> <http://asia.nikkei.com/Tech-Science/Tech/Multilingual-chatbot-to-link-Japan-hotels-foreign-tourists>  
<http://asia.nikkei.com/Tech-Science/Tech/Cybersecurity-companies-adopting-AI-but-so-are-hackers>  
<http://www.japantoday.com/category/technology/view/nec-launches-ai-software-that-searches-video-for-specific-individuals>  
<https://robottaxi.com/en/service/#servicevision>  
<http://www.japantimes.co.jp/news/2016/06/02/business/corporate-business/honda-set-ai-research-base-tokyo/>  
<http://asia.nikkei.com/Business/Trends/Japan-to-launch-initiative-for-AI-based-drug-discovery>  
<http://www.japantoday.com/category/technology/view/ai-based-smart-energy-system-developed>  
<http://www.roboticsbusinessreview.com/deep-learning-startup-pfn-partners-fanuc-save-japanese-manufacturing/>  
<http://www.japantoday.com/category/technology/view/hitachi-develops-basic-ai-technology-that-enables-logical-dialogue-in-japanese>  
<http://asia.nikkei.com/Business/Consumers/NEC-s-AI-system-knows-who-you-are-and-what-you-want>

---

contre les attaques cyber, tandis que NEC met au point un système de reconnaissance faciale, NeoFace, basé sur l'IA.

Le secteur automobile draine également des investissements considérables, avec le projet « Robot Taxi » de DeNA et ZMP, ou le partenariat entre Honda et Softbank sur un projet d'IA pour l'assistance à la conduite automobile. Honda a par ailleurs annoncé en juin 2016 la création d'un centre de R&D à Tokyo dédié à l'intelligence artificielle, Honda R&D Innovation Lab Tokyo, qui regroupera l'ensemble des activités de la société dans le domaine de l'IA. Toyota a de son côté investi 1 milliard de dollars sur l'IA dans son centre de la Silicon Valley aux Etats-Unis.

Dans le domaine de la santé, L'entreprise *Preferred Network* a lancé avec le Centre national du Cancer et l'AIRC un projet conjoint de recherche sur l'IA pour le diagnostic précoce et le traitement du cancer. Fujitsu et NEC ont décidé d'investir dans la découverte de nouveaux médicaments grâce à l'IA tandis que Sony Computer Science Laboratories développe avec l'AIP Center un système médical basé sur l'IA, permettant de recommander automatiquement des traitements adaptés à des symptômes donnés.

Dans le secteur de l'énergie, Hazawa Ando Corp un développe un système de gestion intelligent qui a pour but de réaliser un « Net Zero Energy Building » utilisant l'énergie solaire.

Enfin, dans le secteur de l'industrie de production, l'IA devrait redonner un second souffle aux usines japonaises. Canon, qui a annoncé le retour de sa production au Japon se base sur l'IA pour construire une usine entièrement automatisée à l'horizon 2018. FANUC a également investi massivement dans l'IA pour ses usines de production, via un investissement de 900 millions de yens en 2015 dans la *start-up* Preferred Networks.

De son côté, Hitachi s'est positionné sur le développement d'outils d'IA pour l'aide à la décision par les entreprises. NEC fournit également des solutions de suivi de clients pour le commerce de détail.

#### **IV. Développement de la collaboration scientifique franco-japonaise sur l'intelligence artificielle**

Pour approfondir les opportunités de collaboration entre le Japon et la France dans ce domaine, le Service pour la science et la technologie de l'Ambassade de France au Japon a organisé, en partenariat avec l'Université de Tokyo (Policy Alternatives Research Institute), Elsevier et CEA Tech un événement de 3 jours, les 11, 12 et 13 octobre sur l'apprentissage profond (*deep learning*) et l'intelligence artificielle.

Cet événement était composé d'un symposium et d'un programme de visites, dont celle des nouveaux centres d'intelligence artificielle du MEXT (mis en oeuvre par le RIKEN) et du METI (mis en oeuvre par l'AIST), celle du Policy Alternative Research Institute de l'Université de Tokyo

---

(PARI) ainsi que du Laboratoire franco-japonais d'informatique<sup>1</sup>. Lors de ces visites, les responsables de ces trois nouveaux centres ont montré une forte volonté de développer les collaborations internationales (échanges de chercheurs, partage de grandes bases de données..).

Le symposium a réuni le 12 octobre à l'Université de Tokyo près de 200 personnes, représentants du gouvernement et d'agences gouvernementales (CSTI, MIC, MEXT, JST), experts français et japonais, dans le domaine du *deep learning* et du *machine learning*, issus d'institutions de recherche publique ou d'entreprises.

Cet événement a mis en valeur la qualité de la collaboration existante entre la France et le Japon ainsi que l'existence d'un cadre favorable (programmes de financements, possibilités d'échanges via le JFLI), que les acteurs peuvent à mettre à profit pour développer plus encore cette coopération.

Les différentes présentations et discussions ont fait émerger plusieurs thèmes à fort potentiel pour cette coopération, tels que l'apprentissage par renforcement ou la question de l'accès et de l'utilisation.

Pour entretenir cet élan, un groupe de travail réunissant les acteurs clés du domaine en France et au Japon est en cours de création.

De premières discussions ont également été lancées avec le CSTI, qui a manifesté son intérêt pour échanger avec les experts français sur les questions intelligence artificielle & éthique, en vue de préparer les recommandations pour le gouvernement japonais dans ce domaine.

Enfin, l'ambassade de France au Japon organisera en 2017 un cycle de débats d'idées sur le sujet de l'intelligence artificielle, axés sur l'évolution des modes de travail, les smart cities, etc.

## **Conclusion**

Le Japon mise fortement sur l'intelligence artificielle pour redynamiser son économie, notamment pour conserver, voire rapatrier, son appareil de production, via l'automatisation complète des usines.

Il s'agit de tirer profit du potentiel de l'intelligence artificielle (essentiellement sur les technologies du *deep learning*), de manière intégrée aux domaines de l'IoT et du *big data*.

Le fait que de nombreux acteurs académiques et industriels possèdent des infrastructures de calcul intensif (super-ordinateurs) est également un atout dans ce domaine.

---

<sup>1</sup> Le JFLI, réunissant l'Université de Tokyo, l'Université Keio, le NII (National Institute of Informatics, dépendant du MEXT) côté japonais, le CNRS, Inria et l'Université Pierre et Marie Curie côté français.

Le gouvernement japonais accompagne cette transition de manière très étroite, avec des investissements publics massifs, mais également par la constitution de comités de réflexion sur l'impact de l'utilisation de l'intelligence artificielle sur la société et la nécessaire adaptation des réglementations.

---

## IV. AU ROYAUME-UNI

*Note réalisée avec le concours du service scientifique de l'Ambassade de France au Royaume-Uni*

### **Une recherche britannique en IA dynamique et soutenue par des infrastructures de mutualisation et de mise en réseau.**

Plusieurs équipes recherche et centres d'innovation se sont développés au Royaume-Uni au cours des dernières années dans les domaines connexes de la robotique et de l'Intelligence artificielle.

Afin de répondre aux besoins d'interdisciplinarité et de mutualisation de la recherche en intelligence artificielle, plusieurs organismes et réseaux se sont développés ces dernières années au Royaume-Uni.

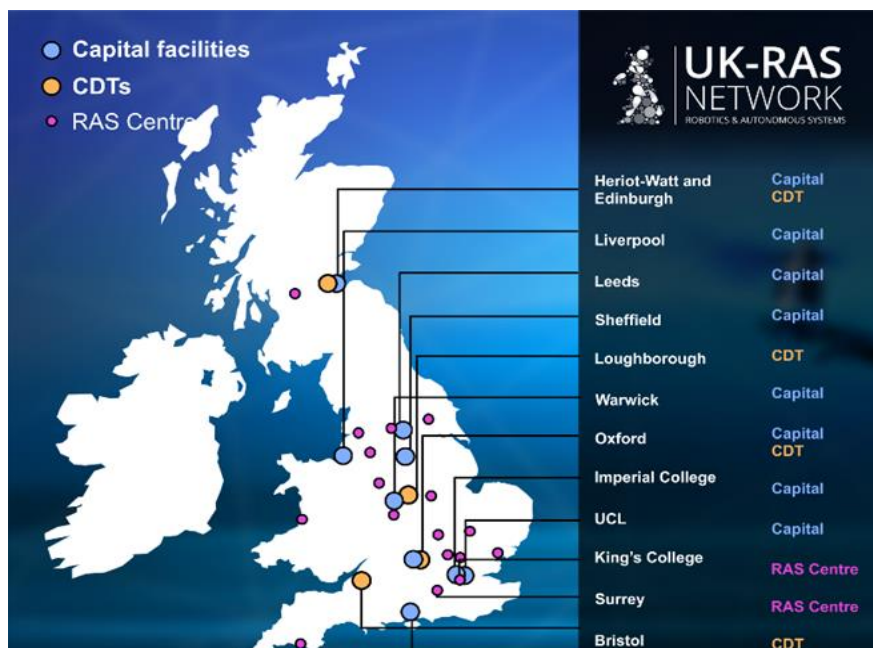
C'est le cas du réseau britannique de robotique et systèmes autonomes de l'Engineering and Physical Sciences Research Council (EPSRC), l'[EPSRC UK-RAS Network](#), créé en mars 2015.

L'EPSRC est une institution britannique qui fournit des financements pour la recherche et l'enseignement supérieur dans le domaine de l'ingénierie et des sciences physiques.

Le réseau EPSRC UK-RAS a pour mission d'assurer une collaboration entre acteurs académiques et industriels.

Il coordonne les activités de huit infrastructures (*Capital facilities*) et quatre centres de formation doctorale (*Centres for Doctoral Training - CDT*) sur la robotique et les systèmes autonomes, à travers le Royaume-Uni.





Crédits : EPSRC UK-RAS

Les universités participant au réseau EPSRC UK-RAS sont les suivantes :

Université d'Edimbourg et Université Heriot-Watt (**Edinburgh Centre for Robotics**), Imperial College London (**Hamlyn Centre**), Université de Leeds (**National Facility for Innovative Robotic**

**Systems**), Université de Sheffield (**Sheffield Robotics**), Université de Liverpool (**Centre for Autonomous Systems Technology**), Université de Southampton (**Autonomous Systems USRG**), Université de Warwick (**Warwick Manufacturing Group**), Université de Bristol (**Bristol Robotics Laboratory**), Loughborough University, University of Oxford (**Centre for Doctoral Training in Autonomous Intelligent Machines and Systems**)

En 2015, plusieurs universités (de Cambridge, Edinburgh, Oxford, University College London et Warwick) ont également créé avec l'EPSRC, l'[Alan Turing Institute](#), qui est le centre national pour la science des données, basé à la British Library. L'Institut a été nommé d'après le célèbre mathématicien et cryptologue britannique Alan Mathison Turing (1912-1954), qui a posé des fondements de l'informatique et contribué à la réflexion sur l'intelligence artificielle, en particulier avec le « test de Turing ».

Il existe également des structures spécifiques de partenariat public/privé dédiées à l'innovation, appelées [Catapult Centres](#), gérées par le Technology Strategy Board (renommé en 2014 « [Innovate UK](#) »<sup>1</sup>). Créés à partir de 2012, ces centres facilitent la R&D grâce à une collaboration entre chercheurs académiques et entreprises. Parmi ces centres, [Digital Catapult](#) effectue notamment des recherches en intelligence artificielle.

L'[Open Data Institute \(ODI\)](#) figure également parmi les acteurs majeurs de la R&D en intelligence artificielle. Il s'agit d'une société privée à

<sup>1</sup> Innovate UK est un organisme public britannique financé par le Department for Business, Energy & Industrial Strategy (BIS)

---

but non lucratif fondée en 2012 par Tim Berners-Lee (un des inventeurs du World Wide Web et actuel président du World Wide Web Consortium) et Nigel Shadbolt (Professeur en intelligence artificielle à l'University of Southampton). Le réseau de l'ODI comprend des entreprises, des établissements publics ainsi que des individus et vise à innover dans le domaine des données.

Ainsi, des liens entre les différents centres de recherche, qu'ils soient publics ou privés, ont été renforcés par une mise en réseau au niveau national.

Au titre de la structuration des activités dans le domaine, tant au niveau académique qu'industriel, on peut également citer The [Society for the Study of Artificial Intelligence and the Simulation of Behaviour](#) (AISB) qui est la société savante principale en intelligence artificielle au Royaume-Uni. Créée en 1964, elle réunit des membres au-delà du territoire britannique et bénéficie ainsi d'une portée internationale.

### **Le financement de la R&D en intelligence artificielle : vers un renforcement des capacités nationales suite au Brexit ?**

Les dépenses de R&D du Royaume-Uni (1,72 % du PIB en 2014) sont globalement inférieures à celles de la France (2,26 %) ou de l'Allemagne (2,84 %). La Confédération de l'Industrie Britannique (CBI) a ainsi recommandé une augmentation de la part des dépenses publiques de R&D afin d'atteindre un objectif de 3% du PIB en 2025. Cependant, les budgets alloués à la recherche au Royaume-Uni bénéficient d'un excellent retour sur investissement : chaque livre sterling dépensée en crédit d'impôt à la R&D entraînerait un investissement complémentaire évalué entre 1,53 et 2,35 livres sterling au Royaume-Uni<sup>1</sup>.

Une augmentation du financement de la R&D de 4,7 milliards de livres sur quatre ans, jusqu'en 2020-21, a été annoncée dans l'Autumn Statement 2016. Cet investissement ciblera notamment les domaines prioritaires que sont l'IA, la robotique, les technologies énergétiques intelligentes, les biotechnologies et la 5G. Ces fonds seront attribués de deux façons : à travers un nouveau fond industriel « Industrial Strategy Challenge Fund » (géré par Innovate UK et par les Research Councils), et via l'attribution de financements par UK Research & Innovation ([UKRI](#)) qui regroupera les deux entités précédentes. Actuellement, le gouvernement britannique investit environ 6,3 milliards de livres par an pour la R&D – en plus du financement des établissements d'enseignement supérieur – ainsi que de 2 milliards de livres en crédits d'impôt pour les entreprises qui effectuent des activités de R&D.

Au niveau fiscal, le ministre anglais des finances George Osborne a annoncé en juillet 2017 – suite au Brexit – que l'impôt sur les sociétés continuerait de baisser, pour atteindre 15% en 2020 contre 20% en 2016. Ce

---

<sup>1</sup> Source : <https://taj-strategie.fr/cout-chercheurs-monde/> 31 janvier 2017

---

changement impacterait indirectement le crédit d'impôt dédié à la recherche (*Research and Development Tax Credit*) en faisant augmenter sa valeur.

Cependant, des chercheurs universitaires s'inquiètent d'une éventuelle diminution des fonds de R&D suite au Brexit<sup>1</sup>. En effet, l'Union européenne est active sur le thème de l'intelligence artificielle de multiples façons, par exemple à travers le financement de projets par le programme Horizon 2020 ou encore avec sa contribution financière à des événements tels que l'« European Robotics Forum » qui aura lieu en mars 2017 à Edimbourg. Plusieurs organismes européens existent dans le domaine de l'intelligence artificielle, notamment l'association [EuRobotics](#) AISBL et le partenariat public-privé pour la robotique en Europe ([SPARC](#)).

### **L'affirmation d'une stratégie britannique en intelligence artificielle, notamment dans un cadre industriel**

Innovate UK a créé en 2013 un « Robotics and Autonomous Systems Special Interest Group » ([RAS-SIG](#)). Ce pôle d'intérêt commun est chargé d'analyser l'état de la recherche en robotique et systèmes autonomes ainsi que d'élaborer une stratégie nationale unissant des acteurs publics et privés. Le pôle RAS-SIG a publié en 2014 une stratégie pour la robotique, l'automatisation et les systèmes autonomes à l'horizon 2020<sup>2</sup>.

La recherche académique en IA entre dans le cadre d'une stratégie nationale de soutien à l'industrie et en particulier à l'« Industrie 4.0 ». Une nouvelle stratégie industrielle, appelée « [Building our Industrial Strategy: Green Paper](#) » a été lancée le 23 janvier 2017. Cette stratégie se concentre autour de dix piliers, dont le premier est « Investir dans la science, la recherche et l'innovation ». Dès 2012, la robotique et les systèmes autonomes (RAS) avaient été identifiés par le gouvernement britannique comme étant l'un des huit domaines technologiques à développer dans le cadre de l'« UK Industrial Strategy ».

La chambre des communes (House of Commons) a publié un rapport sur la robotique et l'intelligence artificielle en octobre 2016<sup>3</sup>. Dans ce rapport, les députés demandent la création d'une Commission permanente sur l'Intelligence Artificielle. Cette nouvelle Commission serait basée à l'Alan Turing Institute, qui est le centre national pour la science des données. Le gouvernement a émis une réponse<sup>4</sup> prudente, en janvier 2017, affirmant

---

<sup>1</sup> Nature, "[How scientists reacted to the Brexit](#)", 24 June 2016 ; Nature, "[Brexit watch: UK researchers scramble to save science](#)", 22 July 2016 ; The New York Times, "[Brexit May Hurt Britain Where It Thrives: Science and Research](#)", 17 October 2016

<sup>2</sup> RAS-SIG, [Robotics, Automation and Autonomous Systems \(RAAS\) 2020 Strategy](#), July 2014

<sup>3</sup> House of Commons, Science and Technology Committee, [Robotics and artificial Intelligence, Fifth Report of Session 2016-17](#), October 2016

<sup>4</sup> House of Commons, Science and Technology Committee, [Robotics and artificial Intelligence, Fifth Report of Session 2016 - Appendix: Government response](#), 11 January 2017.

---

qu'il continuerait à financer les Catapult Centres et le nouveau Industrial Strategy Challenge Fund (ISCF).

En novembre 2016, le Government Office for Science a publié un rapport<sup>1</sup> examinant les opportunités et implications de l'IA pour la société et pour le gouvernement. Ce rapport aborde les questions de la gestion des risques éthiques et juridiques ainsi que des avantages que pourrait apporter l'IA à la société et au gouvernement.

### **Les nécessaires adaptations du marché du travail et des formations universitaires**

Les nouvelles technologies de robotique et d'IA seraient susceptibles à la fois de créer de nouveaux métiers et de faire disparaître des emplois. Le rapport « Technology at work: V2.0 »<sup>2</sup> publié en janvier 2016 par l'entreprise financière Citi, en collaboration avec l'Université d'Oxford, prévoit que 35% des emplois au Royaume-Uni pourraient être automatisés d'ici 2020. Cela place néanmoins le pays dans la fourchette basse en termes de risques (57% en moyenne dans l'OCDE). Ce rapport souligne également qu'une automatisation accrue pourrait conduire à de plus grandes inégalités économiques. Les emplois vulnérables sont principalement ceux peu qualifiés, tels que les travailleurs en centres d'appels, les vendeurs et les ouvriers des industries manufacturières. Inversement les emplois à faible risque d'automatisation sont ceux aux activités variées requérant de l'intelligence sociale et de la créativité. Les secteurs moins impactés seraient l'éducation, la santé et le social.

Des risques d'automatisation pèsent également sur des emplois qualifiés. Par exemple en finance, les algorithmes « robo-advice » permettent de recommander des produits d'épargne et de placement à un client de manière similaire à un conseiller financier<sup>3</sup>. Dans le secteur juridique, l'entreprise ROSS Intelligence a créé un service d'analyse de la jurisprudence à destination des cabinets d'avocats. La diffusion de ce type de « robots » pourrait conduire à des bouleversements sectoriels. Le cabinet de conseil Deloitte a publié en janvier 2016 des estimations sur l'impact de l'automatisation sur le marché du travail. Ainsi, ces estimations prédisent que 114 000 emplois du secteur juridique au Royaume-Uni pourraient être automatisés au cours des 20 prochaines années<sup>4</sup>.

---

<sup>1</sup> Government Office for Science, [Artificial intelligence: opportunities and implications for the future of decision making](#), 9 November 2016

<sup>2</sup> Oxford Martin School and Citi, [Technology at Work v2.0: The Future Is Not What It Used to Be](#), January 2016

<sup>3</sup> John Gapper, ["Robots are better investors than people"](#), *Financial Times*, March 16, 2016

<sup>4</sup> Deloitte, ["Automation transforming UK industries"](#), Press Releases, 22 January 2016

Ces transformations à venir du marché du travail interrogent sur la nécessité d'adapter les formations universitaires<sup>1</sup>. Selon le site Whatuni, le Royaume-Uni compte au moins 36 universités qui offrent une formation en IA, représentant un total de plus de 117 diplômes universitaires<sup>2</sup>. L'intelligence artificielle pourrait également impacter la formation avec l'arrivée de nouveaux services numériques. Le professeur Rose Lucking de l'University College London (*Knowledge Lab, Institute of Education*) effectue des recherches sur les liens entre numérique et éducation, en particulier en ce qui concerne les applications l'IA pour l'éducation ainsi que l'apprentissage analytique et le *big data*.

Au niveau des débouchés, de nombreuses entreprises d'intelligence artificielle sont basées au Royaume-Uni<sup>3</sup>.

### Quelques exemples d'innovations et d'applications marquantes en intelligence artificielle

- Pour la première fois quand l'histoire britannique un véhicule autonome a circulé dans les rues publiques, en octobre 2016. Ce véhicule autonome a été développé au Royaume-Uni par [Transport Systems Catapult](#) dans le cadre du projet *LUTZ Pathfinder*, initié en 2014.



Self-Driving Pods. Credits: Transport Systems Catapult

- Après avoir obtenu des autorités britanniques l'autorisation de lever les restrictions de vol, c'est depuis son *Prime Air Lab*, installé dans le *Cambridgeshire*, qu'*Amazon* a effectué sa première livraison par drone d'un colis à un client en décembre 2016.
- La société *DeepMind* spécialisée dans l'apprentissage profond (*deep learning*), créée par le Dr. Demis Hassabis, titulaire d'un doctorat de l'*University College London*, a été rachetée par *Google* en 2014 pour plus de 628 millions de dollars américains.
- La *start-up* de reconnaissance vocale *VocalIQ* développée par le Professeur Steve Young de l'*Information Engineering Division* à

<sup>1</sup> David Matthews, "[The robots are coming for the professionals](#) - Do universities need to rethink what they do and how they do it now that artificial intelligence is beginning to take over graduate-level roles?", *Times Higher Education*, July 28, 2016

<sup>2</sup> Ces Bachelors' et Master's Degree comportent des intitulés avec les termes : "Artificial Intelligence", "Robotics", "Computer Science", "Intelligent Systems", "Automation" ou encore "Cybernetics". <https://www.whatuni.com/degree-courses/search?subject=artificial-intelligence>

<sup>3</sup> Par exemple : *Business Insider UK*, "[10 British AI companies to look out for in 2016](#)", Jan. 5, 2016.

---

l'université de Cambridge a été rachetée par *Apple* en 2015 et contribue à l'évolution du logiciel *Siri*.

- Le Dr. Zoubin Ghahramani, professeur en ingénierie d'information à l'Université de Cambridge et membre de l'Alan Turing Institute, est co-fondateur de la *start-up Geometric Intelligence* qui est basée à New-York et qui développe des technologies de *machine learning*. L'entreprise de quinze personnes a été rachetée fin 2016 par Uber et deviendra *l'Uber AI Labs* à San Francisco.



## V. EN SUISSE

*Note réalisée avec le concours de l'Attaché de coopération scientifique et universitaire de l'Ambassade de France en Suisse*

### 1. - Une contribution scientifique française essentielle

Marvin Lee Minsky de l'Université de Carnegie-Mellon, définit à la fin des années cinquante l'intelligence artificielle comme « la construction de programmes informatiques qui s'adonnent à des tâches qui sont, pour l'instant, accomplies de façon plus satisfaisante par des êtres humains car elles demandent des processus mentaux de haut niveau tels que : l'apprentissage perceptuel, l'organisation de la mémoire et le raisonnement critique ». Ces dix dernières années, le domaine de l'intelligence artificielle a connu des progrès très importants notamment avec l'invention d'algorithmes d'apprentissage profond (*deep learning* algorithms, parfois aussi appelés réseaux de neurones) dus notamment au Français Yann Le Cun<sup>1</sup>. La stratégie mise en place dans ces algorithmes, mime le fonctionnement des neurones du cerveau où des connections se créent, disparaissent ou se renforcent en fonction des stimuli. Notons que dans le cadre biologique il n'y a pas d'**intelligence** sans **apprentissage**. Il s'agit à partir d'une phase d'apprentissage sur des objets connus, comme par exemple des images, des sons, des caractères, de pouvoir déterminer des paramètres d'un algorithme afin que celui-ci puisse donner des réponses pour des objets totalement nouveaux avec une erreur faible. Ici l'erreur est définie comme la différence entre la réponse de l'algorithme et la réponse exacte obtenue pour des données test. De manière plus précise, en déterminant un **nombre fini de paramètres** de l'algorithme avec un ensemble **d'apprentissage fini fixé**, on veut que l'erreur entre la réponse exacte et la réponse de l'algorithme soit minimale pour un ensemble **test fini fixé**. La détermination des paramètres se réalise avec un procédé d'optimisation. Un algorithme d'apprentissage profond comporte quatre éléments :

- des **données de spécification** (données d'entraînement et données de test) ;
- une **fonction coût** à minimiser pour déterminer les paramètres ;
- une **procédure d'optimisation** ;
- un **modèle** permettant de relier les variables d'entrée aux variables de sortie.

---

<sup>1</sup> <https://www.college-de-france.fr/site/yann-lecun/>

Donnons tout de suite quelques exemples que ces algorithmes permettent de traiter : la reconnaissance de visages, la traduction automatique, la reconnaissance d'objet dans une scène vidéo ou la classification (partition d'un ensemble en classes d'éléments de propriétés similaires). Pour tous ces exemples, la dimension des « exemples d'apprentissage » doit être de l'ordre de 10 millions de cas et la dimension de l'ensemble des paramètres est facilement de 1000 ou plus. Cette difficulté des dimensions est appelée la « malédiction de la grande dimension » (the curse dimensionality). Ainsi, pour être efficace, les algorithmes d'apprentissage profond doivent utiliser des **données massives** et disposer d'une bonne puissance de calcul. De plus, plusieurs couches d'algorithme d'apprentissage profond associé à de la classification sont nécessaires pour traiter les exemples complexes évoqués plus haut. Cela correspond à une décomposition en tâches plus simples pour traiter un problème compliqué. Dans le cas de la reconnaissance d'un visage, on ne cherchera pas à reconnaître la totalité du visage en une seule fois, mais certains de ses éléments séquentiellement. Ce point de vue de l'apprentissage conjugué au traitement séquentiel de tâches plus simples et l'augmentation de la puissance des calculateurs ont permis à l'Intelligence artificielle de sortir de l'impasse du traitement global de tâches complexes.

*La « pensée magique » est inappropriée*

L'approche consistant à créer d'énormes ensembles de données, puis à appliquer les algorithmes d'apprentissage profond dont on dispose aujourd'hui est inappropriée pour 'créer de la connaissance' même si elle est souvent proposée. En effet, il n'existe pas d'algorithme universel (pouvant traiter toutes sortes de données), cela peut se démontrer mathématiquement. De même, suivant les problèmes que l'on veut traiter, il faut introduire de l'**a priori** sur la 'connaissance' recherchée (ce qui en termes techniques de la discipline s'appelle les hyperparamètres et les termes de pénalisation).

Il faut quand même citer la réussite de la stratégie d'apprentissage combinée à une mesure de similarité (principes fondamentaux des algorithmes d'apprentissage profond) mise en œuvre pour le traducteur automatique « Google translate ». A partir de bases de données de textes traduits, Google translate propose une traduction par similarité à l'existant déjà traduit. Cet algorithme d'apprentissage profond donne de bons résultats là où des approches analytiques plus linguistiques avaient échoué.

### **Des questions théoriques à approfondir**

*Il n'y a pas d'algorithme universel possible*

Du point de vue de la logique, il n'est pas possible d'avoir un processus d'inférence général (c'est-à-dire valable pour n'importe quel ensemble) à partir d'un nombre limité d'exemples. Cette contradiction logique est levée pour les algorithmes d'apprentissage car ils ne fournissent qu'une réponse probabiliste. Néanmoins même dans le cadre probabiliste, si on recherche une grande généralité pour le domaine d'application, c'est-à-dire qu'un algorithme puisse s'appliquer pour toutes les données la réponse



---

reste négative. Prenons l'exemple des algorithmes de classification par apprentissage, il ne sera pas possible d'en trouver un qui est meilleur que les autres (ce résultat est connu mathématiquement sous le nom du « no free lunch theorem ») si il doit s'appliquer à toutes les données possibles. Dans la pratique pour avoir un taux d'erreur faible on introduit de la connaissance **a priori** sur les données que nous traitons.

### *La malédiction de la grande dimension<sup>1</sup>*

Outre la question de la puissance de calcul que le traitement de grands ensembles de données pose, il y a d'autres aspects théoriques limitant.

La minimisation de fonctions coût dépendant d'un très grand nombre de variables et non convexes est mathématiquement difficile car ces fonctions peuvent avoir des comportements beaucoup plus complexes qu'en petites dimensions. On peut s'en faire une représentation en pensant à un paysage de montagne vu de haut et très accidenté qui présente beaucoup de sommets et de creux, sachant que l'on recherche le creux le plus profond, celui-ci ne sera pas facile à déterminer.

Ce résultat est pénalisant car le calcul des paramètres nécessite toujours la minimisation d'une fonction coût.

Dans les ensembles de grande dimension les **éléments ne sont pas uniformément répartis**, de la même manière que dans le ciel que l'on observe, les étoiles de la Voie lactée, ne sont pas distribués de manière homogène. Cela veut dire que l'on doit prendre en compte cette difficulté lorsque l'on recherchera des « structures » (ou de la connaissance) dans ces grands ensembles de données.

## **2 - Des réussites et des potentialités**

L'exemple de « Google translate » a déjà été cité, mais on pourrait aussi mentionner la publicité ciblée sur le web, l'analyse de données issues du génome, la reconnaissance de la parole comme domaine d'activité où les algorithmes d'apprentissage profond ont permis une réelle avancée.

Parmi les potentialités des algorithmes d'apprentissage profond, on peut noter :

- la médecine personnalisée ;
- les humanités digitales ;
- le déplacement autonome des robots ou des véhicules ;
- les outils pédagogiques pour l'apprentissage des sciences.

## **3 - L'Intelligence artificielle en Suisse**

---

<sup>1</sup> <https://www.college-de-france.fr/site/yann-lecun/seminar-2016-02-19-15h30.htm>

---

Il faut tout d'abord mentionner le centre Google Research Europe localisé à Zürich qui sera dirigé par le Français Emmanuel Mogenet. Le groupe se concentrera sur trois axes principaux de recherche: outre l'[intelligence artificielle](#), il se consacrera aussi à la [compréhension et au traitement automatique du langage naturel](#) utilisée notamment pour les applications vocales, ainsi qu'à la [perception artificielle](#).

Le SGAICO<sup>1</sup> - Swiss Group for Artificial Intelligence and Cognitive Science. Le SGAICO rassemble des chercheurs, des praticiens, ainsi que d'autres personnes intéressés au domaine de l'intelligence artificielle et des sciences cognitives (AI/CO), sous la houlette de la société suisse d'informatique. Le SGAICO vise la promotion des technologies intelligentes pour l'innovation dans la société. Il apporte une plateforme d'échanges en AI/CO où l'industrie et les universités peuvent se rencontrer. Jana Koehler<sup>2</sup> est présidente du SGAICO.

Il y a aussi le laboratoire d'intelligence artificielle et d'apprentissage automatique<sup>3</sup> dirigé par le professeur Boi Faltings à l'École polytechnique fédérale de Lausanne, ainsi que l'Automatic Control Laboratory<sup>4</sup> à l'École polytechnique fédérale de Zürich.

L'institut de recherche Idiap<sup>5</sup> (anciennement Institut d'intelligence artificielle perceptive), situé à Martigny (État du Valais, Suisse), est une fondation de recherche autonome, indépendante et à but non lucratif spécialisée dans la gestion d'informations multimédia et dans les interactions homme-machine multimodales. Fondé en 1991 par la municipalité de Martigny, l'École polytechnique fédérale de Lausanne, l'Université de Genève et Swisscom, il est reconnu internationalement pour ses travaux en traitement et reconnaissance de la parole, apprentissage artificiel, vision par ordinateur, authentification biométrique et interface homme-machine. Il est associé à l'EPFL par un plan de développement commun. Son budget, qui s'élève à plus de 10 millions de francs suisses, est financé à 50 % par des projets de recherche récompensés selon des processus concurrentiels et à 50 % par des fonds publics. Alors qu'il employait une trentaine de personnes en 2001, l'Idiap a en 2016 une centaine d'employés, dont 80 chercheurs.

L'Idiap, alors dirigé par le professeur Hervé Boulard<sup>6</sup>, a été l'institution hôte du Pôle de recherche national IM2 (2001-2013) dédié à la Gestion interactive et multimodale de systèmes d'information.

### *La propriété des données en Suisse*

---

<sup>1</sup> <http://www.s-i.ch/fr/sgaico/>

<sup>2</sup> <https://user.enterpriselab.ch/~takoehle/>

<sup>3</sup> <http://ic.epfl.ch/intelligence-artificielle-et-apprentissage-automatique>

<sup>4</sup> <https://control.ee.ethz.ch/overview.en.html>

<sup>5</sup> <https://www.idiap.ch/linstitut>

<sup>6</sup> <http://people.idiap.ch/boulard> `boulard [at] idiap.ch`

---

La loi suisse sur la protection des données (LPD) est une des plus restrictives d'Europe, mais son domaine d'application est restreint. A ce propos, notons qu'un certain nombre de sociétés pensent installer en Suisse leur centre de données pour cette raison.

Pour la gestion des données de médecine personnelle la LPD n'est pas adaptée. En effet, à partir d'applications dites de « bien-être », comme les montres connectées ou des objets connectés, les entreprises qui développent les applications informatiques enregistrent des données physiologiques personnelles (rythme cardiaque, pression artérielle...) sur leurs serveurs. Ces données personnelles sont par la suite utilisées par ces entreprises comme elles le souhaitent. Les milieux scientifiques suisses sont conscients des dérives que cela pourrait entraîner, et développent des solutions innovantes où l'individu décide de l'usage de ses données de santé. Citons par exemple la plateforme midata<sup>1</sup> développée par la Haute école spécialisée de Berne et qui accueille déjà deux projets, dont un sur la sclérose en plaque avec des patients de l'hôpital universitaire de Zürich.

Du point de vue légal en Suisse, les juristes considèrent que le Droit ne se préoccupe que des états et non des processus. Ainsi des principes généraux suffisent et il n'y a pas besoin de lois spécifiques pour encadrer ces technologies.

#### **4 Un regard critique sur l'Intelligence artificielle**

Dans son ouvrage *La société automatique* le philosophe Bernard Stiegler voit dans le développement de l'Intelligence artificielle et des automatismes un risque de disparition de l'abstraction et de la pensée. L'argument repose sur l'antagonisme entre l'automatisme et le processus cognitif reposant sur l'observation, la construction d'un modèle abstrait permettant d'interpréter l'observation et l'analyse de l'adéquation des deux. L'intelligence artificielle a permis de développer des procédures de « saisie prédictive » de texte qui utilisent les textes que l'on a déjà écrits et la fréquence d'apparition d'une séquence. Ce qui est le plus utilisé se substitue à ce que l'on pourrait créer ou penser.

---

<sup>1</sup> <https://www.midata.coop/fr.html>

---

## ANNEXE 6 : SUR LE DROIT COMPARÉ DE LA ROBOTIQUE

Les éléments suivants sont issus des deux ouvrages, « *Droit des robots* » de MM. Alain et Jérémie Bensoussan et « *Comparative handbook : robotics technologies law* », ainsi que du livre blanc du Symop « *Droit de la robotique* » publié en 2016. Ils font souvent l'objet d'un parti-pris favorable à la création d'une personnalité juridique pour les robots.

En 2015, le marché de la robotique représentait plus de 26 milliards de dollars (contre 15 milliards de dollars en 2010). La robotique et l'intelligence artificielle pourraient représenter une révolution industrielle d'ampleur au moins comparable à celle d'Internet. **Le développement de technologies d'intelligence artificielle peut potentiellement concerner quatre secteurs d'investissement clés : la défense, l'industrie, l'entreprise et le secteur personnel.** La Commission européenne a évalué, en 2012, que la seule robotique de service pourrait constituer un marché de plus de 100 milliards d'euros à l'horizon 2020. La robotique pose des questions juridiques essentielles. Le soutien des pouvoirs publics est important, au travers de partenariats public-privé, à l'instar du plan France Robots Initiatives (100 millions d'euros), et de la mise en place d'une fiscalité attractive afin de favoriser les investissements dans le domaine de la robotique<sup>1</sup>.

Il est proposé, ci-dessous, d'effectuer des **comparaisons juridiques sur les points suivants : le statut légal du robot, les régimes de responsabilité appliqués aux robots, la protection des données collectées et stockées par le robot, la propriété intellectuelle de la création robotique, la reconnaissance des contrats concernant ou conclus par des robots, les robots dans le secteur de la santé et le cas des voitures autonomes.**

### I - Le statut légal du robot

La rupture technologique provoquée par les progrès conséquents de la robotique et de l'intelligence artificielle pourrait poser la question de la création d'une personnalité juridique pour les robots. La création d'une telle personnalité juridique permettrait d'offrir à l'entité robot dotée d'intelligence et de liberté décisionnelle une catégorisation juridique alignée sur ses capacités, ainsi qu'une insertion apaisée et sécurisée dans le tissu social. L'élément clé dans la conception d'une personnalité robot est la liberté décisionnelle que celui-ci aura acquis grâce à son intelligence et à sa capacité d'apprentissage. Bien que le robot voie ses capacités limitées par l'humain qui les lui aura attribuées, il demeure qu'il jouira d'une certaine

---

<sup>1</sup> Cf. par exemple le dispositif d'amortissement accéléré prévu par la loi de finances pour 2014.

liberté, même résiduelle, son comportement ne pouvant être complètement prédit.

**Tableau comparatif sur le statut légal du robot**

<b>Afrique du Sud</b>	S'il n'existe pas de statut légal ni de législation propre aux robots, certaines dispositions peuvent leur être appliquées, comme le <i>Protection of Personal Information Act 4</i> de 2013, ou le <i>Medecines and Related Substances Amendment Act 14</i> de 2015. De même, le <i>South African Bureau of Standards (SABS)</i> est un organe autonome établissant les normes de standardisation qui, de fait, concernent les robots, notamment en appliquant les standards ISO 10218-1 :2011 et ISO 8373 :2012.
<b>Allemagne</b>	Il n'existe pas de statut légal du robot ; cependant, de nombreuses dispositions contenues notamment dans le BGB (Code civil), dans le StGB (Code pénal), dans le TMG (loi sur les télémedias), ou encore le BDSG (loi fédérale sur la protection des données).
<b>Belgique</b>	Il n'existe pas de statut légal du robot ; le robot est reconnu comme un objet juridique mais est défini d'un point de vue légal comme une machine industrielle.
<b>Brésil</b>	Il n'existe pas de statut légal du robot ou de la machine intelligente, à l'exception des drones qui sont reconnus comme des objets juridiques. Les robots demeurent des objets juridiques non-identifiés d'un point de vue légal.
<b>Chine</b>	Les robots n'ont pas de statut légal reconnu ; cependant, ils sont identifiés comme des « produits », et sont donc considérés comme des objets conformément à au droit des biens chinois ( <i>PRC Property law</i> ).
<b>Costa Rica</b>	Les robots sont classés dans la catégorie des biens, et sont considérés comme des objets juridiques.
<b>Espagne</b>	Il n'existe aucun droit des robots ni statut légal des robots ; les robots sont considérés comme des objets, et n'ont donc pas de personnalité juridique. Aucun projet de loi concernant l'octroi d'un statut légal du robot n'est pour l'instant à l'étude.
<b>États-Unis</b>	Bien que les technologies robotiques représentent un marché émergent considérable, la législation tant fédérale que locale n'accorde pas de statut légal aux robots.
<b>France</b>	Les robots n'ont pas de statut légal, et sont considérés comme des biens. Cependant, la personnalité juridique étant une construction juridique qui n'est pas exclusive aux êtres humains, les robots pourraient bénéficier d'une adaptation de la législation en se voyant attribués une personnalité juridique.
<b>Grèce</b>	Il n'existe pas de législation propre aux robots ; cependant, certains textes visent la régulation de certains types de technologies robotiques, comme la législation sur les drones ou la législation sur les voitures autonomes.

<b>Israël</b>	Il n'existe pas de législation spécifique aux robots ; cependant, de nombreuses lois peuvent s'appliquer aux robots, comme sur la propriété, la propriété intellectuelle ou la responsabilité relative aux dommages causés par une opération robotique.
<b>Italie</b>	Si la directive 2006/42/CE (directive Machines) a été transposée dans l'ordre juridique interne, il n'existe pas de statut légal du robot. Certaines dispositions, notamment en matière de responsabilité, existent, mais elles ont été pensées pour des machines industrielles, et non pour des robots.
<b>Japon</b>	La législation nationale apporte plusieurs définitions contenant le terme robot - comme celle introduite par le cadre juridique de la radio qui définit le robot comme une machine ou un équipement sans-fil qui transmet automatiquement des données. Il n'existe cependant pas de législation pertinente pour les robots, du fait de la difficulté d'appliquer une régulation spécifique pour chaque type de robot.
<b>Liban</b>	Il n'existe aucun statut légal du robot ; une réflexion peut cependant être menée sur l'octroi d'une personnalité juridique aux robots.
<b>Portugal</b>	Les robots n'ont pas de statut légal ; ils sont considérés par la loi comme des « machines » et sont en ce sens des objets soumis aux lois et régulations en application, notamment en matière de commerce, de transports, d'investissement ou de production de machines et d'équipements.
<b>Royaume-Uni</b>	Il n'existe pas de cadre juridique applicable aux robots. Cependant, la réflexion autour de l'élaboration d'une personnalité juridique robot pourrait notamment s'appuyer sur les sept « principes de la robotiques » proposés par l'EPSRC et l'AHRC ou par la stratégie « RAS 2020 » proposée par le groupe d'intérêt RAS-SIG à destination du Gouvernement.
<b>Suisse</b>	Il n'existe aucun statut légal du robot, et seule la loi sur la transparence peut pour le moment s'appliquer aux robots.

## II - Régimes de responsabilité

<b>Afrique du Sud</b>	Un régime de responsabilité spécial est appliqué aux robots : celui de la responsabilité du fait des produits défectueux en vertu de la loi CPA ( <i>Consumer Protection Act 68 of 2008</i> ). Ainsi, la section 61 de la loi CPA impose une responsabilité objective aux fabricants, importateurs, distributeurs et vendeurs de biens, notamment de robots.
<b>Allemagne</b>	Le régime de responsabilité allemand distingue principalement : le régime de responsabilité contractuelle du régime de responsabilité civile, et la responsabilité objective de la responsabilité subjective. Aucun de ces régimes ne semble convenir pour un régime de responsabilité du robot, étant donné qu'il peut être très complexe - voire impossible - de retracer la chaîne de causalité et l'intentionnalité ou la négligence dans un cas de dommage causé par un robot autonome à une personne

	juridiquement responsable. Cependant, dans l'état actuel de la législation, le régime de responsabilité objective semble le plus adéquat à appliquer aux robots ; et un régime de responsabilité subjectif s'applique au fabricant en cas de comportement intentionnel ou négligeant entraînant une défaillance du robot.
<b>Belgique</b>	Le régime de responsabilité est principalement subjectif ; mais les mécanismes de responsabilité actuels ne sont pas nécessairement adaptés à la coopération homme-robot : étant donné que les robots sont des ensembles complexes mêlant matériel informatique et logiciels, parfois intégrant de l'intelligence artificielle, il peut être impossible, ou très long et très coûteux, de prouver une faute qui aurait causé des dommages. Il en serait de même dans la perspective de l'application d'un régime de responsabilité objective.
<b>Brésil</b>	Il n'existe pas de dispositions spécifiques concernant le régime de responsabilité appliqué aux robots ; de manière générale, le régime de responsabilité objectif est appliqué dans les relations civiles ou commerciales, étant donné que les robots sont considérés comme des objets.
<b>Chine</b>	Le régime de responsabilité par défaut est le régime objectif, et le régime de responsabilité subjectif ne s'applique que dans certains cas avec des circonstances spécifiques. Les robots étant considérés comme des produits, leur régime de responsabilité est défini par la loi sur la qualité des produits ( <i>RPC Product Quality Law</i> ). Ainsi, le fabricant et le vendeur assument une responsabilité objective concernant la qualité du produit délivré au client. De même, en cas de dommages physiques ou à la propriété du client, le vendeur assume une responsabilité civile fondée sur la faute et doit donc, selon l'article 42 de la loi sur la qualité des produits, assurer la compensation du dommage causé.
<b>Costa Rica</b>	Les régimes de responsabilité objective et de responsabilité subjective sont tous deux appliqués en fonction des cas ; de manière générale, quiconque cause un dommage à autrui a une obligation de le réparer. En fonction de la tâche exécutée par le robot et du risque qui y est associé, l'un de ces deux régimes s'applique. Cependant, ces deux mécanismes ne sont pas adaptés à la coopération humain-robot.
<b>Espagne</b>	Un régime de responsabilité objective s'applique de manière générale s'applique aux produits ; cependant, dans le cas très techniques, la Cour peut inverser la charge de la preuve et imposer au fabricant de démontrer que le produit n'est pas défectueux. Un robot étant un produit très complexe dont il peut être difficile de prouver la défaillance, ces précédents d'inversement de la charge de la preuve dans des cas très techniques peuvent s'appliquer pour des dossiers liés à des dommages causés par des robots.
<b>États-Unis</b>	Il est probable que de nombreux types d'équipements robotiques ou quasi-robotiques seront soumis à des standards de responsabilité stricte du fait des produits défectueux, notamment dans le cas des robots médicaux. Il serait en effet difficile de prouver la négligence du fabricant dans le cas des robots. En outre, le droit des contrats fait émerger d'autres formes de responsabilités ; ainsi, les fabricants et les distributeurs doivent veiller à inclure des renoncements de garanties appropriés et des limitations de responsabilité dans leur documentation sur les produits, les divulgations et les contrats pour la vente de ces produits.

<p><b>France</b></p>	<p>Le robot est considéré comme une chose dans le droit civil français. La responsabilité du fait des choses, en tant que régime de responsabilité objectif prévu par l'article 1384 du Code civil, signifie que pour qu'elle soit appliquée, la chose doit être impliquée dans le dommage et qu'elle joue un rôle actif (comme le fait d'être en mouvement ou de toucher la victime) dans l'occurrence dudit dommage. C'est l'individu considéré comme « gardien » de la chose qui est responsable de la réparation du dommage cause. Cependant, si le dommage est causé par une faille de sécurité du robot, le régime de responsabilité pour le dommage causé par la chose s'applique au fabricant du robot.</p>
<p><b>Grèce</b></p>	<p>Le fabricant assume une responsabilité du fait des produits défectueux : en vertu de l'article 6 de la loi 2251/1994, si un produit défectueux cause un quelconque dommage à un consommateur ou à leur propriété, le fabricant doit compenser ce dommage indépendamment du fait qu'il y ait eu négligence ou faute de sa part. Dans le cas des robots, identifier la partie responsable en cas de dommage peut être un exercice complexe, d'autant plus qu'une technologie robotique peut être composée d'un logiciel en <i>Open source</i> qui, par nature, vise à être modifié. Selon le professeur David Vladeck (université de Georgetown), une distinction d'application de régime de responsabilité selon le caractère semi-autonome ou pleinement autonome d'un robot. Dans le cas d'un robot semi-autonome, donc servant d'outils à un humain ou à une entité juridique, il est possible d'appliquer la responsabilité du fait des produits défectueux, assumée par le fabricant. Cependant, dans le cas de robots pleinement autonomes, il est plus adéquat de recourir à un régime de responsabilité objective quand il est presque impossible d'identifier la partie fautive, ou si c'est la machine elle-même qui est fautive, étant donné que celle-ci ne dispose pas de capacité juridique.</p>
<p><b>Israël</b></p>	<p>En l'absence de régime de responsabilité spécifique aux robots, différents régimes de responsabilité pourraient s'appliquer selon les cas : la loi sur la responsabilité civile des produits défectueux de 1980 (<i>Defective products Law</i>), qui impose un régime de responsabilité objective sur les fabricants et importateurs de produits défectueux lorsqu'une partie ayant subi le dommage soit physiquement blessée. L'ordonnance de responsabilité civile (<i>Torts Ordinance</i>) indique que, pour attester d'une négligence, le plaignant doit prouver l'existence d'une obligation de diligence, le manquement à ce devoir par le défendeur, et une connexion causale entre le manquement à ce devoir et le dommage causé au plaignant. Les termes d'un contrat entre un fabricant et un utilisateur peuvent soulever une cause d'action fondée sur la garantie fabricant/importateur du robot.</p>
<p><b>Italie</b></p>	<p>En vertu de la directive 59/92/EC, le produit est supposé comme sûr par définition. Ainsi, en cas de dommage causé par l'utilisation du produit, il est de la responsabilité du fabricant que de vendre uniquement des produits sûrs.</p> <p>Trois parties prenantes peuvent être responsables d'un dommage commis par un robot : le fabricant, ce qui implique que sont engagées la responsabilité du créateur et celle du programmeur ; le distributeur, si le bien a été produit et importé depuis l'extérieur de l'Union européenne ; et l'utilisateur, qui peut être responsable de dommages causés à l'encontre de tierces parties par un objet sous son contrôle et durant son utilisation.</p>
<p><b>Japon</b></p>	<p>La loi du 1<sup>er</sup> juillet 1995 sur la responsabilité civile des produits défectueux dispose que le fabricant est considéré comme responsable et doit compenser le dommage subi ; cependant, le fabricant ne saurait être responsables des dommages résultant uniquement de la qualité du produit en lui-même.</p>
<p><b>Liban</b></p>	<p>La loi de 2000 sur la protection des consommateurs tient une personne pour</p>



	responsable si cette personne a introduit un produit défectueux sur le marché. Le cas du régime de responsabilité des robots en cas de dommage causé à une personne tierce soulève de nombreuses questions : est-ce qu'un robot est responsable dans ses actions et ses révisions ? Qui doit être tenu responsable ?
<b>Portugal</b>	<p>Concernant le régime de responsabilité contractuelle, il n'existe aucune règle applicable aux accords dans le champ de la robotique ou impliquant des robots. Concernant le régime de responsabilité non contractuelle, il est nécessaire de distinguer trois cas :</p> <ol style="list-style-type: none"> <li>1. Le cas où la partie subissant le dommage doit prouver la faute, ce qui correspond à la règle générale du droit Portugal pour le régime de responsabilité non-contractuelle.</li> <li>2. Le cas où la faute est présumée en raison du non-respect du devoir de chacun de surveiller tout objet déplaçable en leur possession et le devoir de chacun à prévenir le danger qui résulte d'une dangereuse activité déclenchée ou bénéficiant à cette personne ou légale entité.</li> <li>3. Le cas où le régime de responsabilité objective s'applique, lorsque le fabricant met le produit sur le marché ou quiconque ayant le contrôle sur un véhicule terrestre.</li> </ol> <p>En vertu du décret-loi 383/89, le régime de responsabilité d'un fabricant dont le produit a causé un dommage à cause d'un défaut de fabrication. Dans ce cas-ci, les robots peuvent à la fois qualifiés de « fabricants » (robots concepteurs) et de « produits ».</p>
<b>Royaume-Uni</b>	<p>Les robots étant considérés par le droit britannique comme des produits, ils se voient appliqués le régime de responsabilité du fait des produits défectueux (<i>product liability</i>). Cependant, le cadre juridique actuel n'est pas adapté pour le cas des robots, car il ne peut pas apporter toutes les solutions aux problèmes de responsabilité qui émergent et émergeront de l'utilisation des robots. La principale difficulté réside dans la complexité du robot en tant que produit composé de éléments matériels et de logiciels et dont la création et réalisation d'opérations est difficilement traçable.</p> <p>Les robots n'ayant pas de personnalité juridique, le régime de responsabilité du fait des produits défectueux ou des actes qui causent des dommages ne pourra reposer sur le robot lui-même. Cependant, désigner la personne responsable pour les dommages causés par un robot sera une tâche difficile, du fait du nombre de sujets impliqués par la création, la commercialisation et le fonctionnement des robots.</p>

### III - Protection des données collectées et stockées par le robot

<b>Afrique du Sud</b>	La personne morale ou physique qui détermine pourquoi et comment le robot collecte les informations personnelles est responsable de la protection de ces informations.
<b>Allemagne</b>	<p>Le propriétaire du robot est responsable du traitement des données et de son respect aux lois de protection des données personnelles en vigueur. Le règlement 2016/679 sur la protection des données personnelles sera, en outre, applicable dès 2018.</p> <p>Lorsque des données personnelles sont stockées ou traitées par le robot, le propriétaire ou l'opérateur traitant les données doivent en notifier l'autorité fédérale</p>

	<p>de protection des données personnelles. En outre, chaque Lander dispose, en plus de l'autorité fédérale, d'une autorité de protection des données indépendante.</p> <p>Les responsables privés de robots doivent enregistrer ces procédures auprès des autorités de supervision compétentes.</p> <p>L'enregistrement obligatoire ne s'applique pas si le responsable du traitement a désigné un responsable de la protection des données ou si le responsable du traitement recueille, traite ou utilise des données à caractère personnel. Ainsi, la plupart des entreprises allemandes ont nommé un responsable de la protection des données.</p> <p>Si une partie tierce procède au traitement des données personnelles sous l'autorité de l'entité responsable, la loi considère que cette partie tierce fait partie de l'entité responsable et le transfert ou l'accès aux données personnelles ne requiert pas d'autorisation légale.</p> <p>La section 9 de loi fédérale sur la protection des données (BDSG) pose des exigences fondamentales en matière de mesures organisationnelles et techniques de sécurité des données.</p>
<b>Belgique</b>	<p>Le Code pénal belge permet à un propriétaire de poursuivre un individu ayant piraté ou saboté les données contenues dans un robot, qui est considéré comme un système informatique.</p> <p>Les données collectées et stockées par un robot sont protégées par la loi du 8 décembre 1992 sur la protection de la vie privée, cette loi transposant la directive 9546/EC du 24 octobre 1995.</p> <p>Tout propriétaire de robot doit notifier en amont toute opération de traitement des données collectées et stockées par le robot à la commission vie privée</p> <p>Le règlement général sur la protection des données adopté en 2016 par le Parlement européen impose une protection des données par défaut pour tous les produits, services et systèmes utilisant des données personnelles.</p> <p>Le propriétaire du robot doit assurer la sécurité et la confidentialité de ses données personnelles collectées et stockées par le robot.</p>
<b>Brésil</b>	<p>Il n'existe pas de régime juridique de protection des données personnelles spécifique aux robots ; de fait, les dispositions générales s'appliquent, notamment l'article 5 de la Constitution et les dispositions que l'on peut trouver dans le Code civil, dans le Code pénal, dans le Code de la consommation ou encore dans la loi civile de l'Internet du 23 avril 2014.</p>
<b>Chine</b>	<p>Si la loi sur la protection des informations personnelles (<i>Personal Information Protection Act</i>) adoptée en décembre 2015, elle ne répond pas aux « canons » européens de la protection des données : cette loi définit (et protège) les « informations » personnelles comme celles qui permettent d'identifier un utilisateur (nom, âge...) et les informations qui permettent la localisation et l'heure de consommation d'un service par un utilisateur.</p> <p>Cependant, il existe quelques textes, comme la décision du comité permanent de l'Assemblée nationale populaire sur le renforcement de la protection de l'information sur les réseaux, qui peuvent s'appliquer aux robots collectant et stockant des données à caractère personnel.</p>
<b>Costa Rica</b>	<p>Le propriétaire d'un robot gère la collecte et le stockage de données personnelles ; et les propriétaires ainsi que les vendeurs doivent respecter les dispositions de la loi n° 8968 sur la protection des individus concernant le traitement des données personnelles.</p>

<p><b>Espagne</b></p>	<p>La protection des données personnelles collectées par un robot ou un système informatique est assurée par la loi organique 15/1999 et le décret royal 1720/2007. Le propriétaire ou l'utilisateur d'un robot (désigné comme responsable du robot) doit respecter la législation de protection des données concernant le traitement de données personnelles au travers d'un robot. De même, le responsable gérant le système de traitement des données doit respecter certaines obligations posées par l'Agence espagnole de protection des données (AEPD), notamment sur la notification du type de données détenues, l'identité du responsable, ou encore le type de traitement de données et son but. Le responsable du robot est également tenu d'adopter les mesures techniques nécessaires pour assurer la sécurité des données personnelles afin d'empêcher leur altération, leur perte, ou tout traitement ou accès non-autorisé.</p>
<p><b>États-Unis</b></p>	<p>Il est possible d'attribuer à certaines données une paternité, ce qui les place sous la protection du droit d'auteur. C'est notamment le cas des lignes de code qui composent l'intelligence du robot. Les nombreux composants constituant un équipement robotique peuvent être également protégés par les lois sur le secret industriel. La question de la commercialisation, de l'utilisation d'un robot et des opérations et traitements effectués celui-ci pose la question du consentement de personnes tierces dont les données personnelles peuvent être collectées (par exemple, une personne photographiée par un drone).</p>
<p><b>France</b></p>	<p>Les activités robotiques sont soumises au respect de la loi Informatique et libertés de 1978. Ainsi, les propriétaires de robots gérant le système de traitement des données doivent respecter les obligations posées par la CNIL, c'est-à-dire la notification standard sur les utilisations du robot, sur le type de logiciel utilisé, sur les systèmes de sécurité installés pour protéger les données d'intrusions tierces non-autorisées, sur les données personnelles stockées dans le robot, et sur les informations fournies aux utilisateurs concernant le traitement de leurs données personnelles. Tout traitement de données personnelles doit être signalé en amont à la CNIL, et l'utilisation de données « sensibles », comme les données médicales, doit être autorisée par la CNIL. En outre, les propriétaires de systèmes de traitement de données personnelles sont soumis à une obligation de sécurité et de confidentialité des données.</p>
<p><b>Grèce</b></p>	<p>La collecte et le traitement de données à caractère personnel par un robot est régi par la loi 2472/1997 (DPL) et la loi 3471/2006 (PECL) qui transposent respectivement les directives européennes 95/46/EC et 2002/58/EC. En vertu de la loi DPL, le responsable des données (<i>data controller</i>) doit notifier par écrit à l'Autorité hellénique de protection des données (DPA) l'établissement d'une base de données ou le début d'un traitement de données. En vertu de la loi DPL, le traitement de données personnelles doit être confidentiel. Il doit être effectué exclusivement par les personnes agissant sous l'autorité et les instructions du responsable des données. Les informations stockées par un robot mais également les informations concernant la structure, les caractéristiques et les opérations effectuées par un robot peuvent être, dans certaines circonstances, être qualifiées d'informations confidentielles. Cependant, la législation ne fait pas mention spécifiquement des informations confidentielles stockées dans un robot ; néanmoins, le cadre juridique actuel peut s'appliquer aux robots.</p>
<p><b>Israël</b></p>	<p>Les utilisateurs de robots sont soumis aux mêmes lois de la protection des données qu'une personne collectant des données d'une différente manière.</p> <p>La loi 5741-1981 sur la protection de la privée indique que les détenteurs de bases de données contenant des informations appartenant à plus de 10 000 personnes ont l'obligation de s'enregistrer et d'identifier l'objectif guidant l'entretien d'une telle</p>

	base de données, les moyens de collecte des données contenues dans cette base, et les mesures de sécurité utilisées pour protéger les données.
<b>Italie</b>	<p>Le Code italien de protection des données régit toutes les activités de collecte de données à caractère personnel. Ainsi, toute personne dont les données personnelles sont connectées doit être informé de l'identité de celui la personne physique ou morale qui les recueille et dans quel but. Le fait que les robots ne bénéficient pas d'une personnalité juridique a un impact significatif sur ce le cas des opérations de collecte de données effectuées par un robot : qui est la personne qui doit informer la personne dont les données sont collectées ?</p> <p>Le droit italien a posé des exigences minimales de sécurité, la section 32 du Code de protection des données indiquant notamment que les données doivent être protégées par de mesures qui seraient adéquates et cohérentes avec l'état de l'art de la technologie. Ainsi, si les mesures minimums de sécurité ne peuvent être évitées, la loi introduit un niveau de standard bien supérieur en matière de sécurité des données, car ces standards doivent être cohérent avec l'état de la technologie mais également du niveau de risque et du type de données traitées.</p>
<b>Japon</b>	<p>Étant donné qu'un robot est en capacité de stocker et de traiter des données personnelles, l'utilisateur d'un tel robot peut être considéré comme un opérateur commercial détenant des informations personnelles, qui sont protégée par la loi de protection des informations personnelles, texte amendé en 2015 dont l'application est prévue en 2017.</p> <p>Si un robot collecte plus de 5 000 dossiers d'informations personnelles en répondant à des visées commerciales, l'utilisateur du robot est un opérateur commercial qui doit répondre à de nombreuses exigences, telles que la spécification obligatoire de l'utilisation des informations personnelles ou la mise en place des mesures nécessaires permettant d'assurer une gestion sécurisée des informations personnelles. A noter qu'avec l'amendement du texte, à partir de 2017, tout robot stockant et traitant des données sera considéré comme un opérateur commercial, même si ce robot a stocké et traité moins de 5 000 dossiers d'informations personnelles.</p>
<b>Liban</b>	Le Liban ne dispose pas d'autorité nationale de protection des données ni de loi de protection des données.
<b>Portugal</b>	<p>Si le robot est capable de collecter et de traiter des données personnelles, le responsable des données sur le territoire portugais, ou utilisant un équipement situé sur le territoire portugais, doit se conformer aux règles applicables comme énoncées dans la loi n°67/98 sur la protection des données personnelles. Une notification de traitement des données personnelles par le responsable des données doit être envoyée en amont à l'Autorité portugaise de protection des données.</p> <p>Dans le cas du traitement des données « sensibles », leur traitement est interdit sauf si une autorisation spéciale est accordée par l'autorité ou si une des quelques exceptions juridiques liées aux raisons de l'intérêt public s'appliquent. Le responsable des données est soumis à un devoir de confidentialité et doit installer des mesures techniques et organisationnelles adéquates afin de protéger les données personnelles contre les accidents, dommages ou actes malveillants.</p>
<b>Royaume-Uni</b>	<p>Le traitement de données personnelles est encadré par la loi sur la protection des données de 1998, qui dispose que le respect de la législation sur le traitement des données personnelles incombe au responsable de la base de données. Le responsable doit notifier en amont l'<i>Information Commissioner's Office</i> (ICO) de tout traitement de données personnelles en vertu de la section 18 de la loi sur la protection des données.</p> <p>Le responsable de la base de données doit également respecter de nombreuses</p>

	obligations imposées par la loi, notamment le traitement équitable et licite des données uniquement aux fins légales spécifiées pour lesquelles les données ont été obtenues et la mise en place de mesures de sécurité afin de prévenir tout traitement non-autorisé ou illicite, ainsi la perte, la destruction ou l'endommagement des données.
<b>Suisse</b>	Les provisions juridiques standards contenues dans la loi fédérale sur la protection des données et les différentes lois cantonales de protection des données sont appliquées aux questions de collecte et de traitement de données à caractère personnel par des robots.

#### IV - Propriété intellectuelle de la création robotique

La production et l'utilisation de robots et de technologies similaires dans le domaine public, industriel ou privé, peut être limitée par divers mécanismes de protection de la propriété intellectuelle, notamment avec l'existence de brevets qui protègent l'innovation technique, le droit d'auteur ou le savoir-faire, non breveté mais dont le détournement peut engager la responsabilité de l'auteur dudit détournement.

<b>Afrique du Sud</b>	<p><u>Droit d'auteur</u> : le droit d'auteur est actuellement défini par les dispositions de la loi sur le droit d'auteur de 1978 (<i>Copyright Act 98 of 1978</i>). Le droit d'auteur n'a pas besoin d'être enregistré ; il s'applique automatiquement. Cependant, la création robotique n'est pas prévue par ce texte. Néanmoins, un <i>Copyright Amendment Bill</i> devrait être promulgué au cours de l'année 2017 qui pourrait affecter le droit d'auteur pour les créations robotiques.</p> <p><u>Dépôt de marque</u> : si un robot peut constituer une marque déposée, il n'est pas prévu qu'une création robotique puisse déposer une marque.</p> <p><u>Brevet</u> : si un modèle de robot peut être breveté, il n'est pas prévu qu'une création robotique puisse être brevetée.</p>
<b>Allemagne</b>	Dans l'état actuel de la législation, il n'est pas possible pour un robot de détenir une quelconque propriété intellectuelle ou industrielle, car ils sont considérés au regard de la loi comme des objets, et ne sont donc pas porteurs de droits. Seuls les humains peuvent exercer un droit de propriété sur une création.
<b>Belgique</b>	Ni le droit national ni le droit communautaire ne prévoient la création par un robot ou un ordinateur. Aucun texte ne considère les robots comme des fournisseurs de service ; les robots sont considérés comme des objets, des outils, et de fait, cela exclut leurs créations du champ de la protection de la propriété intellectuelle et du droit d'auteur tel que défini par le droit communautaire.
<b>Brésil</b>	Les brevets, les marques déposées et le régime de droits d'auteur peuvent offrir aux robots une protection intellectuelle. Cependant, la loi sur la propriété industrielle et la loi du 19 février 1998 sur le droit d'auteur ne permettent pas aux robots de breveter leurs créations ni d'être considérés comme des auteurs. Ces lois considèrent que la

	création ne peut être qu'une humaine, et non issue de machines.
<b>Chine</b>	<p><u>Brevet</u> : les robots étant considérés au regard de la loi comme des produits, il est difficile qu'ils puissent déposer un brevet. Cependant, si une création humaine résultant de l'usage d'un robot ou une création proprement robotique remplissent les critères de brevetage, il est possible d'obtenir une protection de la propriété intellectuelle en vertu du droit des brevets.</p> <p><u>Droit d'auteur</u> : en vertu de la loi sur le droit d'auteur (<i>PRC Copyright Law</i>), une création souhaitant bénéficier du droit d'auteur doit remplir trois critères : l'originalité, la duplication et la légalité. Or, le critère d'originalité est ici conditionné au fait qu'une création doit être accomplie par la pensée indépendante d'un auteur et doit refléter la philosophie et l'intelligence de jugement de l'auteur, plutôt qu'une simple copie ou compilation de matériaux. De fait, les robots ne peuvent satisfaire le critère d'originalité, et leurs créations ne peuvent donc pas être protégés en vertu du droit d'auteur.</p> <p><u>Dépôt de marque</u> : comme de nombreux pays, la Chine a ratifié la Classification de Nice sur les biens et les services, dont la classe 7 catégorise les robots comme des biens et dont les actions en tant que fournisseur de service ne sont pas explicitement pris en compte. Cependant, à l'exception des robots répondant à la définition de la classe 7, les robots peuvent être classés selon leur fonction et leur tâche. Ainsi, les robots médicaux, d'entraînement et d'éducation peuvent être considérés comme de classe 10, 41, 44, etc.</p> <p><u>Savoir-faire</u> : le droit de la concurrence (<i>PRC Anti-Unfair Competition Law</i>) protège le secret d'affaires ; à ce titre, le propriétaire légal du robot peut protéger les techniques du robot en tant que secret d'affaires.</p>
<b>Costa Rica</b>	Dans l'état actuel de la législation, un robot ne peut détenir un droit d'auteur ou des droits intellectuels ni déposer une marque sur une création qu'il aurait développée.
<b>Espagne</b>	Si les robots, en tant que produits, peuvent être brevetés et protégés par le droit d'auteur et le dépôt de marque, la création robotique ne peut en bénéficier.
<b>États-Unis</b>	<p>De nombreuses classifications de brevet de conception et de brevet d'utilité peuvent s'appliquer aux robots.</p> <p>Les marques déposées pour la robotique sont protégées de la même manière que les autres outils électroniques ou techniques.</p> <p>La législation ne fait aucune mention de la protection de la propriété intellectuelle ou industrielle de la création robotique.</p>
<b>France</b>	<p><u>Brevet</u> : il est possible d'accorder un brevet à une création résultant d'un processus industriel impliquant un ordinateur ou un robot.</p> <p><u>Droit d'auteur</u> : le régime juridique actuel réserve la paternité d'une création à un individu, excluant de fait une paternité exclusivement robotique.</p> <p><u>Savoir-faire</u> : si les créations produites par des composants robotiques ne sont pas éligible à la protection assurée par le Code de la propriété intellectuelle, le savoir-faire, défini comme « un ensemble d'informations pratiques non-brevetées résultant de l'expérience et de l'expérimentation », peut être utilisé comme un outil pour protéger la création robotique.</p> <p><u>Dépôt de marque</u> : de même que dans les pays ayant ratifié la Classification de Nice, les robots sont considérés comme des biens et leurs actions en tant que fournisseur de services ne sont, de fait, pas prises en compte.</p>

<p><b>Grèce</b></p>	<p><u>Droit d’auteur</u> : toute création effectuée au recours d’un robot en tant qu’outils est protégé par la loi 2121/1993 sur le droit d’auteur. Cependant, il n’est pas possible de protéger le droit d’auteur d’une personne non-humaine, autant une entité juridique qu’un robot.</p> <p><u>Dépôt de marque</u> : la Grèce a ratifié la Classification de Nice, et les robots sont à ce titre couverts par la classe 7 ; mais au vu du nombre de services que peut offrir un robot, il est également possible de les inclure dans d’autres classes, comme la classe 12, 28, 41, 44, etc.</p> <p><u>Brevet</u> : un robot peut, en tant que tel ou en tant que nouvelle méthode, faire l’objet d’un brevet. Cependant, la législation ne prévoit pas de dispositions permettant le brevetage de la création robotique.</p>
<p><b>Israël</b></p>	<p>Les robots peuvent être, en tant que le résultat d’efforts collaboratifs, être protégés par les droits de la propriété intellectuelle ; cependant, le droit d’auteur n’a nul besoin d’enregistrement et s’applique automatiquement.</p>
<p><b>Italie</b></p>	<p>Comme dans le reste de l’espace communautaire, la création de robots peut bénéficier d’une protection de la propriété intellectuelle, et les robots peuvent être brevetés. Le droit d’auteur (<i>copyright</i>) n’existe pas en tant que tel en Italie ; le régime de protection de la propriété intellectuelle protège les inventions et les droits des auteurs, ces droits étant une conséquence naturelle de la création d’un bien protégé. Cette protection est accordée dès lors que le produit est marqué d’un caractère de nouveauté d’originalité. Le savoir-faire est quant à lui protégé par le secret industriel.</p>
<p><b>Japon</b></p>	<p><u>Brevet</u> : la loi sur le brevet indique qu’une invention industriellement applicable peut être autorisée à obtenir un brevet. Les robots peuvent être protégés en tant qu’inventions. Une invention est brevetée lorsqu’elle exprime une nouveauté, une démarche inventive et a une application industrielle. Un inventeur ne peut être qu’un être humain, les robots n’étant pas des sujets titulaires de droits et d’obligations, et ne peuvent de fait pas exprimer leurs idées.</p> <p><u>Droit d’auteur</u> : si le robot peut assister l’humain dans le travail de création, un travail est considéré comme une production dans laquelle les idées et pensées sont exprimées de manière créative et qui relève de n’importe quel champ artistique. De fait, un robot ne peut être considéré comme un auteur et en recevoir la protection.</p>
<p><b>Liban</b></p>	<p>Il existe de nombreux systèmes de protection de la propriété intellectuelle, notamment le brevet, le droit d’auteur ou le dépôt de marque et de concept, qui peuvent être utilisés et combinés pour offrir une protection totale à toute invention, et notamment aux robots.</p>
<p><b>Portugal</b></p>	<p><u>Brevet</u> : toute nouvelle invention dans le domaine de la robotique qui implique une activité inventive et appropriée à une utilisation industrielle peut être brevetée.</p> <p><u>Droit d’auteur</u> : il existe une protection juridique du droit d’auteur pour les programmes informatiques en tant qu’ouvrage littéraire ; il n’existe aucune raison pour laquelle cela ne pourrait être appliqué aux logiciels robots.</p> <p><u>Dépôt de marque</u> : le dépôt de marque peut être utilisé dans le domaine de la robotique afin d’identifier le résultat de développement du robot (le produit final, le logo et le nom du robot peuvent être protégés).</p>
<p><b>Royaume-Uni</b></p>	<p><u>Brevet</u> : en vertu de la loi sur les brevets de 1977, une invention est brevetable si elle est nouvelle, si elle implique une démarche inventive, si elle offre des applications industrielles et n’est pas interdit par la loi, ce qui permet aux robots satisfaisant ces critères d’être brevetés.</p>

	<p><u>Conception</u>: la conception d'un modèle de robot peut remplir les conditions pour être protégée en vertu de la loi sur les modèles enregistrés de 1949 et la loi sur le droit d'auteur, les modèles et le brevet de 1988.</p> <p><u>Droit d'auteur</u>: en vertu de la loi sur le droit d'auteur, les modèles et le brevet de 1988, un logiciel peut être protégé par le droit d'auteur en tant qu'ouvrage littéraire, et tout comme une compilation de données si cela représente un travail original et non-copié qui a nécessité de la part de l'auteur l'application de compétences, du discernement et du travail. Si une compilation de données peut correspondre à la définition de la base de données telle qu'énoncée dans la section 3A (1) de cette loi, alors les bases de données contenues dans les robots peuvent être protégés.</p> <p><u>Dépôt de marque</u>: les robots étant considérés aux yeux de la loi comme des produits, les fabricants peuvent avoir recours au dépôt de marque. Ainsi, les robots et les services délivrés par celui-ci peuvent être protégés par le dépôt de marque.</p> <p><u>Création robotique</u>: il n'existe aucune loi ni aucune jurisprudence traitant de la question de l'application des droits de propriété intellectuelle aux créations robotiques.</p>
--	---

## V - Contrats concernant ou conclus par des robots

<b>Afrique du Sud</b>	Il est possible de conclure un contrat de manière électronique, comme dans de nombreux pays mais la question des contrats intelligents (passés par et entre un ou plusieurs robots) n'est pas abordée. Cependant, un contrat peut techniquement être conclu par une personne utilisant un robot, il est de fait nécessaire de développer concernant la capacité de contractualisation d'un robot.
<b>Allemagne</b>	Les logiciels ne peuvent exprimer de déclaration d'intention. Cependant, des contrats conclus par des robots peuvent être possible au regard du droit allemand si et seulement si ces déclarations sont associées à un utilisateur et sont considérés comme appartenant à l'utilisateur. Un contrat conclu par des robots peut également être valide si les parties contractantes ont décidé, du fait de leur liberté contractuelle, d'une procédure où ce sont les robots qui établissent les clauses du contrat. Néanmoins, ces solutions ne permettent pas de résoudre tous les problèmes liés à la question des contrats conclus par des robots.
<b>Belgique</b>	Les contrats conclus automatiquement par des robots ne sont pas abordés par le droit national. Cependant, les deux parties d'un contrat doivent stipuler dans celui-ci qu'un système autonome pourra prendre un certain nombre de décisions qui les contraindront.
<b>Brésil</b>	Les contrats conclus par des robots ou impliquant une décision effectuée par un robot ne sont pas abordés ou régulés par un cadre juridique spécifique ; cependant, les décisions automatiques relèvent de la responsabilité de ceux qui utilisent ces systèmes autonomes ou qui décident d'incorporer ce système dans le processus contractuel.
<b>Chine</b>	Un robot, considéré comme un produit, peut également être l'objet d'un contrat de vente entre deux parties. De fait, le robot étant un produit, il ne peut être reconnu comme l'une des parties concluant un contrat mais peut en être l'objet.
<b>Costa Rica</b>	Il n'existe aucun cas enregistré de contrat passé automatiquement par un robot.



<b>Espagne</b>	Les contrats passés par des robots ne sont pas abordés par la législation espagnole. De fait, des contrats civils ou commerciaux passés par l'intermédiaire d'interactions robotiques peuvent se voir appliquées les règles contractuelles classiques sur les ventes et les droits et obligations généraux prévus par le Code civil et le Code de commerce.
<b>États-Unis</b>	De nombreuses activités, surtout financières, ont recours, grâce aux <i>Fintech</i> , à des contrats passés automatiquement par des robots. Ainsi, il est admis que 70% des opérations à Wall Street sont effectuées par des robots.
<b>France</b>	Le robot peut être l'objet d'un contrat ; cependant, les contrats « intelligents », passés par des robots notamment via une chaîne de blocs utilisant une monnaie virtuelle comme le <i>Bitcoin</i> (qui n'est pas reconnue comme monnaie dans le droit interne) ne sont pas reconnus par le Code civil.
<b>Grèce</b>	Les systèmes autonomes, comme les robots, sont de plus en plus utilisés dans les transactions, et notamment dans le cadre de transactions boursières. À ce stade d'installation et d'utilisation de contrats intelligents, il est possible d'affirmer que les concepts et principes traditionnels de droit civil, en identifiant la plateforme opératrice derrière le robot.
<b>Italie</b>	Bien que de nombreuses entreprises y aient recours, les contrats concernant la vente ou la location de robots ne disposent pas encore de modèles ou de standards comme c'est les cas des contrats concernant les ordinateurs. Le système juridique italien est un système de responsabilité limité : les dommages doivent être prouvés, et la charge de la preuve est du ressort du plaignant. Les dommages doivent être la conséquence directe et immédiate de actes ou faits préjudiciables, et les dommages doivent être évités de façon prévisible. La jurisprudence montre que, de manière répétée, lorsque les produits « high-tech » sont concernés, il existe un écart de connaissance entre le fabricant ou le vendeur et l'acheteur, ce qui signifie que le fabricant a une meilleure connaissance des implications que peut avoir l'utilisation de certaines technologies, notamment les robots. Le fabricant/vendeur doit respecter un certain niveau de diligence. Ainsi, il est essentiel que l'acheteur soit dûment informé sur de nombreux aspects du robot.
<b>Japon</b>	Il n'existe pas de standards ou modèles pour les contrats concernant les robots. Du fait de la complexité des structures et des mécanismes composant les robots, il est difficile pour les parties d'un contrat impliquant un robot de déterminer si un robot rempli les exigences des utilisateurs et acheteurs. De nombreux différends concernant les contrats sur les robots naissent du fait des interprétations des différentes parties du contrat. Les contrats conclus par des robots, ou automatiquement, sont d'usage courant au Japon dans le secteur services financiers à destination des particuliers. Cependant, les robots contractants participent à la procédure de formation du contrat, mais la décision finale demeure prise par un humain.
<b>Liban</b>	La vente et l'achat de robots peut être régie par deux textes : le Code des obligations et des contrats, et le Code de la consommation. L'article 401 du Code des obligations et des contrats impose notamment deux obligations aux vendeurs : une obligation de délivrer l'objet de la vente, et une obligation de garantir une utilisation paisible de l'objet par l'utilisateur et de le garantir contre les défauts. La vente de robot devrait être conditionnée au respect de mesures de sécurité et d'exigences techniques ; cependant, cela n'est pas spécifié pour le moment dans la législation. Il n'existe pas non plus d'encadrement des contrats robots par le Code des obligations et des contrats, ce texte ne limitant la liberté de contrat qu'au respect de l'ordre

	public, des règles obligatoires et des valeurs éthiques et morales.
<b>Portugal</b>	Il n'existe pas de cadre légal spécifique aux contrats sur les robots ; le principe de liberté contractuelle des parties s'applique. Il n'existe aucune législation spécifique concernant les contrats intelligents ; cependant, les contrats passés via des ordinateurs sont des extensions de la volonté humaine, en ce sens que la machine agit en fonction des instructions et scénarios calculés et saisis par le programmeur.
<b>Royaume-Uni</b>	Un contrat juridiquement exécutoire exige, entre autres, l'intention de créer des relations juridiques et, si le contrat n'est pas conclu directement par une partie, le pouvoir d'agir en son nom. Lorsque des moyens électroniques et des machines automatisées sont utilisés comme des outils qui facilitent simplement la conclusion d'un contrat conformément à des exigences strictement prédéterminées établies par les parties, alors un accord contraignant peut être formé. Cependant, si des robots peuvent, grâce à un haut niveau d'automatisation, conclure des contrats, il demeure un flou important. Les robots ne sont dès lors plus considérés comme des outils permettant la communication des intentions de chaque partie mais un système autonome capable de décisions immédiate et en interaction avec son environnement, et de fait les parties peuvent ne pas être informées sur le contenu ou l'existence même du contrat et donc de la création de relations juridiques. Il existe un flou juridique sur la question des contrats passés par, tout particulièrement lorsque ces contrats contiennent des clauses contraignantes.

## VI - Les robots dans le secteur de la santé

Il existe une certaine diversité des robots médicaux, notamment les robots d'aide au diagnostic, les robots chirurgicaux, et les robots d'aide aux soins ou de maintien de l'autonomie des personnes âgées ou handicapées.

Les robots médicaux sont considérés par les agences nationales et internationales de produits de santé comme des "dispositifs médicaux" ; l'autorité nationale compétente en France est l'Agence nationale de sécurité du médicament et des produits de santé (ANSM).

La mise en circulation sur le marché des robots médicaux est encadrée par des exigences de conformité issues de la réglementation européenne transposée dans le Code de la santé publique, et sont classés en quatre catégories, chacune d'elle étant associée à des règles d'évaluation et de contrôle spécifiques.

Le code de la santé publique impose aux professionnels de santé d'engager leur responsabilité en cas d'incident liée à l'utilisation d'un robot médical défaillant ; en outre, il existe trois normes ISO encadrant la sécurité des robots médicaux : la norme ISO 13482 :2014 révisée par la norme ISO/DIS 13485.2 qui énonce les exigences de sécurité pour les robots de soins personnel, et la norme ISO 13485 :2003 qui fixe des exigences relatives à la qualité des dispositifs et services médicaux fournis.

La protection des données médicales en France est assurée par plusieurs textes. D'une part, la notion de données « relatives à la santé » sont définies par l'article 1111-7 et 1111-8 du code de santé publique, les délibérations de la CNIL et la décision du 6 novembre 2003 de la Cour de

justice de l'Union européenne. D'autre part, la collecte et le traitement des données dites « sensibles », qui comprennent les données de santé telles que, est strictement encadré par la loi Informatique et libertés qui impose une obligation de déclaration des traitements de ces données personnelles à la CNIL.

### Tableau comparatif

<b>Afrique du Sud</b>	<p>Les robots médicaux sont sujets de la loi de 1965 sur les médicaments et substances relatives, texte qui fut amendé par la loi de 2015 sur les médicaments et les substances relatives.</p> <p>L'Autorité de régulation des produits de santé d'Afrique du Sud (SAH-PRA) régule l'utilisation des robots dans le secteur médical.</p>
<b>Allemagne</b>	<p>Les robots médicaux, désignés « dispositifs médicaux », sont notamment soumis aux dispositions de la loi sur les dispositifs médicaux (MGP) qui transpose les directives 93/42/CEE, 90/385/CEE et 98/79/CE. La régulation sur les dispositifs médicaux est, de fait, presque harmonisée au sein de l'espace communautaire. La commercialisation des dispositifs médicaux est conditionnée à l'obtention d'un marquage de conformité CE. La loi MGP ne contient aucune disposition concernant la responsabilité des dispositifs médicaux en cas de dommages causés. De fait, le régime général de responsabilité s'applique (objectif/subjectif ; contractuel/civil). La loi BDSG (section 3) définit les données médicales comme des données « sensibles », et leur traitement et utilisation est ainsi soumise à des exigences élevées. Le patient concerné doit en outre avoir déclaré expressément son consentement.</p>
<b>Belgique</b>	<p>La commercialisation de dispositifs médicaux est soumise aux dispositions de la directive 93/42 qui a été transposée dans l'arrêté royal du 15 juillet 1997 relatif aux dispositifs médicaux implantables actifs et dans l'arrêté royal du 18 mars 1999 relatif aux dispositifs médicaux. Ainsi, la commercialisation d'un dispositif médical est soumise à l'obtention préalable du marquage CE qui indique sa conformité avec les exigences de sécurité et de santé.</p> <p>L'arrêté royal du 18 mars 1999 oblige les professionnels de santé à avertir l'Agence fédérale des médicaments et des produits de santé de tout incident impliquant l'utilisation d'un dispositif médical ; ils peuvent être tenus responsables en cas de non-communication de tout incident.</p> <p>Plusieurs standards ISO sont appliqués aux dispositifs médicaux : ISO 13482 :2014, ISO 13485 :2003 et ISO/DIS 13485.2.</p>
<b>Brésil</b>	<p>L'Agence nationale de vigilance sanitaire (ANVISA) a édicté des règles spécifiques concernant l'usage de technologies dans le domaine médical, notamment sur les équipements et les produits, les logiciels, les exigences de sécurité, les données des patients, les exigences de gestion des documents électroniques, les technologies de vidéo et de surveillance et les applications de diagnostic et les produits liés.</p> <p>Bien qu'il n'existe pas de droit de la protection des données, le secteur médical est soumis à de nombreuses dispositions mentionnant le caractère sensible des données et enregistrements médicaux qui sont</p>

	donc soumis à la confidentialité.
<b>Chine</b>	<p>Le droit chinois tient une posture prudente vis-à-vis de l'usage de robots médicaux, tout en encourageant le développement. Le droit chinois prévoit ainsi une prise en charge des événements indésirables impliquant que le fabricant ou l'utilisateur d'un robot médical doit en faire rapport à la structure compétente. De même, sont prévus un système de rappel de produits défectueux et de sanctions à l'égard des entités ne respectant les dispositions précédentes.</p>
<b>Costa Rica</b>	<p>Les robots médicaux, dénommés « dispositifs médicaux », sont soumis aux dispositions du règlement n°34482-S pour l'enregistrement, la classification, l'importation et le contrôle d'équipement et de matériel biomédical.</p> <p>Les dispositifs médicaux ne peuvent être utilisés sans respecter les exigences de sécurité et de santé des patients, des utilisateurs et des parties tierces. Ces exigences incluent notamment d'avoir un mandataire en justice au Costa Rica, un certificat de conformité, une description du matériel utilisé, et l'intégralité des documents fournis en langue étrangère traduits en espagnol et dûment formalisés par le consulat du Costa Rica.</p>
<b>Espagne</b>	<p>Les robots médicaux sont désignés comme des dispositifs médicaux. Le décret royal 1591/2009 vise à garantir la qualité des dispositifs médicaux en prévoyant des inspections et contrôles périodiques au moyen de différentes procédures. Ce même décret permet à l'Agence espagnole des médicaments et produits sanitaires d'autoriser de manière individuelle et spécifique selon des considérations relatives aux services de santé la commercialisation et le déploiement d'un dispositif médical qui n'a pas satisfait l'ensemble des procédures de validation indiquées à l'article 13 du décret, comme la vérification de certification CE, la déclaration de conformité CE ou l'obtention du label CE. Les fabricants de dispositifs médicaux sont soumis au régime de responsabilité objectif, et les professionnels de santé utilisant des dispositifs médicaux sont également soumis au régime de responsabilité civile.</p>
<b>France</b>	<p>L'Agence nationale de sécurité du médicament et des produits de santé considère les robots médicaux comme des dispositifs médicaux ; l'ANSM est en charge de la régulation du marché des systèmes médicaux.</p> <p>La commercialisation de dispositifs médicaux est soumise aux obligations de conformité indiquées dans le Code de la santé publique et est conditionnée à l'obtention d'un marquage CE reflétant sa conformité aux exigences de sécurité et de santé précisées par la législation européenne. Les dispositifs médicaux sont classés en quatre catégories selon leur niveau de risque potentiel pour la santé.</p> <p>Les professionnels de santé sont dans l'obligation de reporter auprès de l'ANSM toute défaillance, tout incident ou tout risque d'incident causé par un dispositif médical. Si un incident impliquant l'usage d'un dispositif médical défectueux, leur responsabilité est mise en cause.</p> <p>Plusieurs standards ISO sont appliqués aux dispositifs médicaux : ISO 13482 :2014, ISO 13485 :2003 et ISO/DIS 13485.2.</p>

<p><b>Grèce</b></p>	<p>La Grèce a transposé les directives 98/79/EC et 93/42/CEE sur les dispositifs médicaux, impliquant que, comme tout État membre, la Grèce doit prendre les mesures nécessaires pour assurer la mise sur le marché et la mise en service des dispositifs médicaux uniquement si ceux-ci ne compromettent pas la sécurité et la santé des patients, utilisateurs ou d'autres personnes.</p> <p>La collecte et le traitement de données « sensibles », telles que les données médicales, est interdit par la loi DPL. Cependant, la collection et le traitement de données sensibles peut être permis par la DPA selon des conditions spécifiques, comme le consentement écrit de la personne à qui appartiennent ces données.</p>
<p><b>Israël</b></p>	<p>La Knesset a promulgué la loi 5772-2012 sur les dispositifs médicaux puis les règlements 5773-2013 sur l'enregistrement des dispositifs médicaux. L'ambition de ces textes est de réguler le champ des dispositifs médicaux en Israël, et notamment leur conception, invention, expérimentation, production, utilisation, distribution, exportation et importation afin de maximiser la protection des utilisateurs.</p>
<p><b>Italie</b></p>	<p>La directive 2007/47/EC a été transposée en Italie avec le décret législative<sup>1</sup> n°60 du 20 janvier 2010, qui a amendé le décret législative N°47 du 24 février 1997. Selon le décret législatif n°60, un dispositif médical doit être développé et produit de manière à ce qu'il ne puisse, dans son usage ordinaire et prévu, mettre en danger la santé ou la sécurité du patient. Si un dispositif est installé et utilisé correctement mais représente néanmoins un risque pour le patient, le ministère de la santé peut ordonner son retrait et en interdire la vente et l'utilisation en informant la Commission européenne. De plus, si un utilisateur constate qu'un dispositif a causé un incident ou a un problème technique, il doit en informer le ministère afin que celui-ci en informe le fabricant, et évalue l'incident et les mesures à prendre.</p> <p>Les sections 75 à 94 du Code en matière de la protection des données personnelles traitent de la question des robots médicaux. La section 78 dispose notamment que le médecin doit expliquer clairement au patient la manière dont ses données personnelles seront traitées pour la poursuite de son traitement. Le patient est amené à indiquer son consentement par écrit. En outre, les données personnelles de santé doivent être dûment protégées, et doivent notamment être cryptées avant toute communication par voie électronique.</p>
<p><b>Japon</b></p>	<p>En vertu de la loi sur les médicaments et les dispositifs médicaux de 2014, la production et l'utilisation de robots médicaux doivent respecter certaines exigences :</p> <ul style="list-style-type: none"> <li>- L'enregistrement des entreprises de fabrication (et de leurs usines) auprès du ministère de la santé, du travail et des affaires sociales.</li> <li>- L'obtention d'une autorisation de fabrication et commercialisation du ministère</li> <li>- Pour être commercialisé au Japon, tout dispositif médical doit être approuvé en amont par le ministère.</li> </ul> <p>Concernant la question de la protection des données médicales, il n'existe pour le moment aucune loi spécifique. En principe, la protection</p>

<sup>1</sup> Un « decreto legislativo » correspond, dans l'ordre juridique italien, à une ordonnance.

	des données médicales est assurée par les dispositions de la loi sur la protection des informations personnelles.
<b>Portugal</b>	<p>En vertu de la section 7 de la loi sur la protection des données personnelles, toute donnée médicale collectée, stockée ou traitée par un robot est considérée comme une donnée personnelle sensible.</p> <p>Il n'existe pas de dispositions spécifiques au régime de responsabilité pour les dommages causés par un dispositif médical. Dans ces cas, les règles générales sur le régime de responsabilité non-contractuel et le régime de responsabilité du fait des produits défectueux.</p>
<b>Royaume-Uni</b>	L'utilisation de robots dans le secteur de la santé au Royaume-Uni est conditionnée par le règlement sur les dispositifs médicaux de 2002, qui transpose les directives 93/42/CEE, 98/78/CE et 90/385/CEE. Les fabricants de dispositifs médicaux doivent engager une procédure d'évaluation de la conformité de leurs produits et obtenir un marquage « CE » avant toute publicité ou vente. Ils doivent en outre être enregistrés auprès de l'Agence de régulation des produits médicamenteux et de santé (MHRA).

## VII - Voitures autonomes

Trois aspects juridiques majeurs liés à l'exploitation des voitures autonomes peuvent être abordés :

- La **responsabilité liée aux voitures** : dans les États de Floride, du Nevada et du district de Columbia, la loi impute la responsabilité en cas de défaillance de la technologie autonome au fabricant du système autonome, que ce fabricant soit le constructeur du véhicule ou non. En France, le régime en vigueur impose l'indemnisation en cas d'accident au conducteur, ce qui implique une adaptation de la législation et/ou l'adoption d'un cadre réglementaire pour les véhicules autonomes ; dans l'attente d'un tel cadre, ce sont les contrats signés entre les parties prenantes qui répartiront la charge de la dette éventuelle de réparation.
- La **propriété des données techniques** : la question de la propriété des données recueillies et traitées est essentielle mais n'est pour l'instant pas départagée, impliquant que des décisions au cas par cas, l'octroi de licences d'utilisations de ces données et d'accords entre les parties prenantes seront nécessaires.
- La **protection des données à caractère personnel** : la directive STI impose aux États membres de veiller à ce que les données identifiantes soient protégées contre toute utilisation abusive, encourageant également l'utilisation de données anonymes et restreignant l'usage des données personnelles aux traitements nécessaires au bon fonctionnement des applications et services de systèmes de transport intelligent. Ainsi, la production et le déploiement des véhicules intelligents devront suivre, compte tenu des dispositions de la loi Informatique et libertés mais également du règlement européen

encadrant la protection des données personnelles en Europe, une approche respectueuse de la protection des données à caractère personnel (*privacy by design*).

### Tableau comparatif

<b>Afrique du Sud</b>	La législation sud-africaine n'a pas encore traité le cas de la voiture autonome.
<b>Allemagne</b>	L'Allemagne ayant ratifié la Convention de Vienne, seuls les véhicules dotés de systèmes partiellement autonomes (comme les systèmes d'aide à la conduite) sont autorisés à la circulation sur la voie publique, et le chauffeur doit avoir le contrôle du véhicule. Dans l'état actuel du droit de la circulation, le conducteur est strictement responsable des dommages causés. De fait, le conducteur est également responsable des dommages causés résultant d'une défaillance du système autonome du véhicule. En outre, le régime de responsabilité du constructeur réside dans la constatation de trois erreurs : une erreur de construction, une erreur de fabrication, et une erreur d'instruction.
<b>Belgique</b>	Il n'existe aucune règle spécifique concernant l'usage de véhicules sans chauffeur. Le référentiel juridique est la Convention de Vienne, qui a été transposée dans le Code de la route belge en 1968, qui interdit donc la circulation de véhicules sans chauffeur sur la voie publique. Cependant, le gouvernement fédéral a approuvé en septembre 2016 un « code de bonnes pratiques d'expérimentation » des voitures autonomes dans des lieux public.
<b>Brésil</b>	La législation actuelle ne contient aucune disposition concernant les véhicules autonomes. Le Brésil a ratifié la Convention de Vienne sur la circulation routière, ce qui implique que le Code de la route brésilien ne contient aucune disposition autorisant les véhicules autonomes. Cependant, l'expérimentation de voitures autonomes sur la voie publique n'est pas expressément interdite.
<b>Chine</b>	Le gouvernement chinois soutient activement le développement de la R&D et l'expérimentation de la voiture intelligente. Cependant, l'usage de véhicules autonomes sur la voie publique n'est pas encore autorisé, et la responsabilité de toute infraction et dommage causé par un véhicule intelligent incombe, s'il n'y a pas de personne dans l'habitacle, à l'opérateur contrôlant le système du véhicule.
<b>Costa Rica</b>	Le Code de la route du Costa Rica ne contient aucune disposition autorisant l'utilisation de véhicules sans chauffeur, ni de police d'assurance spécifique pour ces véhicules. Dans l'état actuel de la législation sur la propriété intellectuelle, les données techniques générées par les voitures intelligentes seront la propriété du constructeur automobile et du fournisseur de composants.
<b>Espagne</b>	Bien qu'aucune législation relative à l'utilisation de véhicules sans chauffeur n'ait été adoptée, la Direction générale du trafic a adopté l'instruction 15/V-113 en 2015 permettant l'expérimentation de prototypes de véhicules autonomes sur des routes ouvertes au public si certains critères sont remplis.

<p><b>États-Unis</b></p>	<p>Les constructeurs ont la possibilité d'expérimenter leurs véhicules sans chauffeur sur des routes ouvertes au public après avoir obtenu en amont une permission officielle. La Californie, le district de Columbia, la Floride, le Michigan et le Nevada ont déjà adopté des lois autorisant l'expérimentation de véhicules sur route. Les lois de ces États exigent cependant la présence d'un conducteur expérimenté à l'intérieur du véhicule.</p> <p>Afin de répondre aux défis de confidentialité et de sécurité soulevés par la collecte d'un grand nombre de données, les constructeurs automobiles ont développé les « Principes de protection de la vie privée des consommateurs pour les véhicules, les technologies et les services » qui déterminent la collecte et l'usage des informations des utilisateurs des véhicules connectés. Sept principes sont mis en avant : transparence ; choix ; le respect du contexte ; la « minimisation » de la collecte de données, la dé-identification et la rétention légitime ; la sécurité des données ; l'intégrité et l'accessibilité ; et la responsabilité.</p> <p>Afin de minimiser les risques liés à la circulation des véhicules intelligents, les constructeurs ont incorporé des principes de sécurité au cœur du processus de conception (<i>safety by design</i>), et ont établi en 2015 l'Auto-ISAC qui permet de faciliter les échanges d'importants volumes d'informations en temps réel sur les faiblesses éventuelles des dispositifs autonomes et des contremesures envisagées. En outre, les constructeurs se sont engagés de longue date aux côtés de nombreuses tierces parties telles que des technologues de la sécurité, des organisations non-lucratives, des programmes gouvernementaux et des groupes de travail, des universités et des initiatives de STEM afin de développer technologies et usages de sécurité pour les véhicules.</p>
<p><b>France</b></p>	<p>L'article R. 412-6-I du Code de la route indique que tout véhicule en mouvement doit avoir un conducteur. Cependant, la loi du 17 août 2015 dite « loi de transition énergétique pour la croissance verte » contient des dispositions introduisant le cadre expérimental visant à promouvoir l'expérimentation et le déploiement de véhicules propres, incluant les voitures sans chauffeur.</p> <p>Dans l'état actuel de la législation, le conducteur du véhicule est responsable en cas d'accident de la route. Cependant, cette législation ne peut être appliquée telle quelle aux accidents causés par des véhicules autonomes car le conducteur n'a pas le contrôle direct du véhicule.</p> <p>La loi relative à la transition énergétique pour la croissance verte du 17 août 2015 habilite le Gouvernement à agir par ordonnance concernant l'autorisation d'expérimentation de véhicules semi-autonomes ou autonomes sur la voie publique. En ce sens, l'ordonnance n° 2016-1057 du 3 août 2016 relative à l'expérimentation de véhicules à délégation de conduite sur les voies publiques autorise l'expérimentation de véhicules intelligents sur la voie publique sous condition de la délivrance d'une autorisation accordée par le ministre chargé des transports, après avis du ministre de l'intérieur.</p> <p>Début février 2017, l'Allemagne et la France ont conjointement lancé une zone transfrontalière d'expérimentation pour la voiture autonome entre la Sarre et la Moselle, sur un tronçon Merzig-Sarrebruck-Metz. Le lancement de cette zone d'expérimentation s'accompagne d'une mise en place de guichets uniques pour les demandes d'expérimentations.</p>



<b>Grèce</b>	<p>La Grèce a transposé la directive 2010/40/EU sur le déploiement des transports intelligents, ainsi que le règlement 2015/758 concernant les exigences en matière de réception par type pour le déploiement du système <i>eCall</i> embarqué dans les véhicules. En 2014, en vertu de la décision ministérielle 145/2014, un groupe de travail a été créé afin d'élaborer une stratégie nationale pour les systèmes de transports intelligents. Dans l'état actuel du droit, l'utilisation de véhicules autonomes n'est pas autorisée. Cependant, la Grèce ayant également ratifié la Convention de Vienne, un amendement de la Convention ainsi qu'un amendement des législations nationales qui ont transposé la Convention dans l'ordre juridique interne seront nécessaires pour s'adapter à la réalité de la voiture sans chauffeur.</p>
<b>Israël</b>	<p>La voiture intelligente étant encore au stade embryonnaire, la législation israélienne ne prévoit pas encore de dispositions concernant ces véhicules. Bien au contraire, la section 29 des règlements 5721-1961 sur les transports indiquent qu'un chauffeur doit avoir les mains sur le volant dont il a le contrôle à tout instant.</p>
<b>Italie</b>	<p>Il semble que le législateur n'avait pas envisagé l'hypothèse d'un véhicule sans chauffeur : de fait, le Code de la route et des véhicules italien ne pose aucune restriction quant à la manière dont peut circuler une voiture.</p>
<b>Japon</b>	<p>Le Code de la route dispose qu'un véhicule doit être conduit par un conducteur maîtrisant et contrôlant sa voiture. De fait, les voitures autonomes ne sont a priori par autorisées à circuler sur la voie publique japonaise. Cependant, il existe un système national de certification des véhicules qui n'exclut pas les véhicules autonomes : si une voiture autonome obtient une certification, elle pourra être expérimentée sur les routes publiques nationales, à condition qu'un conducteur demeure au contrôle du volant, des freins et autres</p>

	dispositifs de la voiture. Il est également possible d'expérimenter une voiture autonome sur route sans recevoir de certification, grâce à une permission accordée en amont par le ministre.
<b>Liban</b>	L'expérimentation et la circulation de véhicules autonomes ne fait l'objet d'aucune disposition législative ni de texte en écriture.
<b>Portugal</b>	Le Portugal ayant ratifié la Convention de Vienne, les mêmes règles de circulation sont appliquées que dans les pays précédemment cités. Ainsi, le Code de la route interdit la circulation de véhicules sans chauffeur, et le chauffeur est responsable de toutes les manœuvres du véhicule. Cependant, de nombreux changements législatifs, au premier lieu duquel l'amendement de la Convention de Vienne, permettront la multiplication des expérimentations et le déploiement des véhicules intelligents.
<b>Royaume-Uni</b>	Le Royaume-Uni ne dispose pas de régulation spécifique concernant les véhicules semi-autonomes ou autonomes – le Royaume-Uni a signé mais n'a pas ratifié la Convention de Vienne. Une révision du cadre juridique national est à l'étude et devrait être présentée au cours de l'année 2017.
<b>Suisse</b>	Le 24 avril 2015, l'entreprise Swisscom, appuyée par le Département fédéral de l'environnement, des transports, de l'énergie et de la communication, a obtenu l'autorisation exceptionnelle de tester un véhicule appartenant à la Freie Universität de Berlin sur des portions de voie publique déterminées. La Suisse n'ayant pas de constructeur automobile national, les données des conducteurs de véhicules autonomes se verront protégés en vertu de la législation européenne en vigueur, notamment la directive européenne pour le déploiement des STI du 7 juillet 2010 ou le règlement général sur la protection des données du 27 avril 2016.

## **ANNEXE 7 : CONTRIBUTIONS**

---

## I. CONTRIBUTION DE RAJA CHATILA : ROBOTIQUE ET INTELLIGENCE ARTIFICIELLE

La problématique de l'intelligence artificielle telle que posée par Alan Turing<sup>1</sup> était de savoir si les ordinateurs pouvaient être capables de « pensée » (*Can Machines Think?*) et il l'a immédiatement traduite par la question de l'imitation de l'homme. La question initiale de Turing portant sur la « pensée » a été traduite par les fondateurs de l'intelligence artificielle par celle de l'« intelligence », ou des « mécanismes de haut niveau ».

Cette manière de poser la question néglige un constat pourtant simple : le cerveau humain a évolué vers ce qu'il est en développant des capacités de perception, d'interprétation, d'apprentissage et de communication en vue d'une action plus efficace.

La problématique de la robotique pose l'ensemble de ces questions. Le robot-machine est soumis à la complexité du monde réel dans lequel il évolue et dont il doit respecter sa dynamique. La notion d'intelligence doit alors être posée de manière à rendre compte globalement des processus sensori-moteurs, perceptuels et décisionnels permettant l'interaction en temps-réel avec le monde en tenant compte des contraintes d'incomplétude et d'incertitude de perception ou d'action. C'est le sens de la définition de la robotique donnée par Mike Brady (Oxford) dans les années 1980 : « *la robotique est le lien intelligent entre la perception et l'action* ». Dans ce sens on peut aussi dire que le robot est le paradigme de l'intelligence artificielle « incorporée », c'est à dire une intelligence matérialisée dans un environnement qu'elle découvre et dans lequel elle agit.

Il est nécessaire d'adopter une vision d'ensemble du robot, en tant que **système** intégrant ses différentes capacités (perception/interprétation, mouvement/action, raisonnement/planification, apprentissage, interaction) et permettant à la fois la réactivité et la prise de décision sur le long terme. Ces fonctions doivent être intégrées de manière cohérente dans une architecture de contrôle globale (architecture cognitive) ; leur étude de manière séparée risque d'aboutir à des solutions inappropriées.

### **Perception, action, apprentissage**

De nombreuses avancées ont été réalisées dans chacune des fonctions fondamentales du robot. Dans les années 1985-2000, la problématique de la localisation et de la cartographie simultanées

---

<sup>1</sup> A. M. Turing (1950) Computing Machinery and Intelligence. *Mind* 49: 433-460.

---

(*Simultaneous Localisation And Mapping*) a connu un développement formidable qui a permis de bien en cerner les fondements et de produire des systèmes efficaces, le point faible important restant le manque d'interprétations plus sémantiques de l'environnement et des objets qui le composent. L'apprentissage profond a récemment fait une incursion considérable en robotique, prenant la place de l'apprentissage bayésien, à la fois pour la perception et pour la synthèse d'actions. Mais la perception en robotique nécessite une *interaction* du robot avec son environnement et non une simple *observation* de celui-ci. L'apprentissage par renforcement est un apprentissage non supervisé qui permet au robot de découvrir à la fois les effets de ses actions, caractérisés par une « récompense » obtenue comme conséquence de l'action, et l'incertitude de ses actions qui n'ont pas toujours les mêmes effets. Le lien entre perception et apprentissage - en particulier apprentissage par renforcement - est essentiel pour extraire la notion d'*affordance* qui rend compte des propriétés des objets en ce qu'elles représentent pour l'agent, et qui associe les représentations perceptuelles aux capacités d'action. C'est cela qui sert de base au robot pour exprimer le sens du monde qui l'entoure.

### **Interaction**

Au-delà du traitement du langage naturel, les problèmes d'interaction et d'action conjointe homme-robot, avec la mise en œuvre de « prise de perspective » est une problématique fondamentale pour permettre une interaction efficace et naturelle entre l'humain et le robot. Ce sujet demande un développement qui associe des recherches en robotique et en Sciences de l'Homme et de la Société (SHS), - en particulier, en sociologie, philosophie, psychologie, linguistique. Le rôle des émotions dans l'interaction est à explorer, bien au-delà de travaux actuels qui se contentent de classer *a priori* des expressions faciales ou de produire des expressions d'émotions artificielles par le robot. L'expression d'émotions par un robot pose des questionnements scientifiques et éthiques sur l'authenticité de ces émotions et sur l'anthropomorphisation qui peut en résulter.

### **Décision**

La prise de décision s'appuie sur des techniques relativement classiques de l'intelligence artificielle. La non-adéquation de la planification déterministe avec les contraintes du monde réel a amené le développement de méthodes basées sur les processus décisionnels dans l'incertain, comme les processus markoviens. Ces approches trouvent cependant leurs limites dans une trop grande complexité. L'apprentissage d'actions en robotique est souvent réalisé par des méthodes d'apprentissage par renforcement et le lien entre planification et apprentissage n'est pas très clair. Des recherches sur de nouvelles approches alliant apprentissage et raisonnement probabiliste avec une convergence rapide sont nécessaires.

---

### **Conscience de soi**

Mais la question initiale de Turing ne devrait-elle pas conduire à l'interrogation suivante (d'ailleurs évoquée par lui-même) : une machine peut-elle avoir une *faculté de conscience d'elle-même* ? Car malgré toutes les recherches en robotique et intelligence artificielle, les résultats, aussi significatifs soient-ils, restent le plus souvent applicables dans des contextes restreints et bien délimités. Ainsi, la perception ne permet pas à un robot de *comprendre* son environnement, c'est à dire d'élaborer une connaissance suffisamment générale et opératoire sur celui-ci (d'où la nécessité d'étudier la notion d'*affordance*), la prise de décision reste limitée à des problèmes relativement simples et bien modélisés. Les principes fondamentaux restent largement incompris, qui permettraient aux robots d'interpréter leur environnement, de comprendre leurs propres actions et leurs effets, de prendre des initiatives, d'exhiber des comportements exploratoires, et d'acquérir de nouvelles connaissances et de nouvelles capacités. Les clés pour permettre la réalisation de ces fonctions cognitives peuvent être le méta-raisonnement et la capacité d'auto-évaluation, deux mécanismes réflexifs.

### **Pluridisciplinarité**

La recherche en robotique pose des questions proches des sciences cognitives, des neurosciences et de plusieurs domaines des Sciences de l'Homme et de la Société (SHS), comme la sociologie, la psychologie et la philosophie. Des programmes interdisciplinaires seront probablement le bon moyen pour aborder les différentes facettes des questions fondamentales.

Deux aspects concernant l'éthique devraient être abordés: d'une part les *méthodologies de conception* éthique de systèmes autonomes, de manière à ce que ceux-ci tiennent compte des valeurs éthiques humaines (par exemple respect de la vie humaine, des droits humains) et de manière à ce que les algorithmes qui les régissent soient transparents, explicables, traçables, et d'autre part l'éthique des machines, c'est à dire comment les décisions prises par une machine peuvent intégrer un raisonnement éthique.

## II. CONTRIBUTION DE PATRICK ALBERT

---

## ANNEXE 8 : RÉPONSES ÉCRITES AUX RAPPORTEURS

### I. RÉPONSES D'AXELLE LEMAIRE, SECRÉTAIRE D'ÉTAT CHARGÉE DU NUMÉRIQUE ET DE L'INNOVATION ET DE THIERRY MANDON, SECRÉTAIRE D'ÉTAT CHARGÉ DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE, AU QUESTIONNAIRE DE VOS RAPPORTEURS

N.B : Le questionnaire de vos rapporteurs a été adressé au Gouvernement en décembre 2016, alors qu'**aucune stratégie nationale en matière d'intelligence artificielle n'avait alors été annoncée**. Les réponses, à fournir avant le 7 février 2017, sont finalement parvenues à vos rapporteurs le 26 février 2017.

#### I. Questions communes

##### **1. Y a-t-il une stratégie nationale en matière d'intelligence artificielle ? Le cas échéant, comment le Gouvernement compte-t-il mettre en œuvre cette stratégie ?**

- L'intelligence artificielle est considérée comme un sujet de première importance en raison de ses impacts potentiels importants : opportunités économiques offertes par un nouveau marché (émergence de nouveaux acteurs, de nouvelles offres et de nouveaux usages, gains de productivité), transformations des modèles économiques d'entreprises, nouveaux besoins de formations initiales et continues, créations et disparitions de certains types d'emplois, apparition de nouveaux sujets d'ordre éthique (degrés d'autonomie possibles en fonction des usages, niveau de responsabilité des systèmes autonomes...), etc.

- Avec l'accélération des performances des technologies qui sont au cœur du développement de l'intelligence artificielle (puissances de calculs de plus en plus importantes à des coûts toujours plus bas, technologies du *big data* de plus en plus matures, augmentation des capacités de stockage, développement de nouvelles méthodes comme l'apprentissage profond), l'élaboration et la mise en œuvre d'une stratégie nationale en matière d'intelligence artificielle est devenue une nécessité (sous l'effet aussi de la compétition internationale).

- Cela est d'autant plus vrai que nous observons une compétition internationale déjà très vive, avec en leaders les pays en pointe que



---

sont la Chine et les États-Unis. La stratégie nationale que nous sommes en train de construire doit aussi répondre à cet enjeu en nous donnant les moyens collectifs d'imposer une certaine vision sociale, économique, politique de l'intelligence artificielle, sans doute différente de celles proposées par ces pays.

- Une première étape a été de s'approprier ce sujet très vaste, en particulier au moyen d'entretiens avec des experts de différents secteurs et de différentes disciplines, afin d'en discerner les enjeux et d'évaluer leur importance. Cette première étape a confirmé la nécessité d'agir sans tarder afin que l'avènement des technologies d'intelligence artificielle ne cristallise pas des craintes qui vont freiner l'innovation et les créations d'emplois et d'entreprises de demain, mais soit, au contraire, porteur d'une volonté de progrès technologique et social.

- La seconde étape, en cours aujourd'hui, consiste à approfondir la réflexion sur les enjeux éthiques, juridiques, technologiques, scientifiques, et socio-économiques de l'intelligence artificielle, puis à proposer et à prioriser, au regard de ces enjeux, des recommandations qui doivent permettre de maximiser les effets positifs du développement et de la diffusion de l'intelligence artificielle en France, tout en amortissant autant que possible les impacts négatifs. À cette fin, en Octobre 2016, la ministre chargée du Numérique et de l'Innovation a souhaité la mobilisation de toutes les parties prenantes dans l'établissement d'une feuille de route pour l'intelligence artificielle qui devra être finalisée fin Mars 2017. L'écosystème de l'enseignement supérieur et de la recherche est bien sûr autant concerné que celui des entreprises et de l'innovation. Axelle Lemaire et Thierry Mandon, secrétaire d'État chargé de l'enseignement supérieur et de la recherche ont ainsi procédé au **lancement officiel du projet # France intelligence artificielle le 20 janvier 2017**. La méthode retenue consiste en un travail coopératif entre acteurs de l'intelligence artificielle (chercheurs, entreprises, économistes, etc.) et de l'administration (DGE, DGRI, DGESIP, CGI, DG Trésor), au sein de groupes de travail rassemblés en trois grandes thématiques clés. Une première étape de restitution des travaux à mi-parcours a eu lieu le 21 février 2017. L'ensemble des livrables attendus des groupes de travail seront remis le 14 mars 2017. La présentation de la feuille de route issue de ces travaux est programmée le 21 mars 2017.

- Une troisième étape consistera à déployer les actions retenues dans la feuille de route, avec l'appui des financements annoncés par le Président de la République le 21 février dernier.

**2. Quels initiatives le Gouvernement compte-t-il prendre pour relever les défis de l'intelligence artificielle ? Pouvez-vous nous les décrire ?**

---

- D'ores et déjà, comme mentionné précédemment, différentes initiatives ont été prises pour relever les défis de l'intelligence artificielle : entretiens avec des acteurs clés du domaine, définition d'objectifs prioritaires (placer la France sur la carte du monde de l'intelligence artificielle, mobiliser la communauté de l'intelligence artificielle en France et la cartographier, mobiliser des financements et construire un soutien complet à l'intelligence artificielle de la recherche amont aux *start-ups*, porter la question de l'intelligence artificielle dans le débat politique et clarifier le discours auprès des citoyens notamment, en prenant en compte les dimensions éthiques, juridiques, technologiques, scientifiques, et d'impact socio-économique), et lancement de l'élaboration d'une feuille de route pour l'intelligence artificielle.

- Depuis plusieurs années, le développement de technologies et d'usages de l'intelligence artificielle fait, d'autre part, l'objet d'un soutien au travers d'appels à projets de R&D qui permettent des coopérations entre laboratoires de recherche, entreprises fournisseurs de technologies et entreprises utilisatrices. Des projets ont ainsi été soutenus dans le cadre du concours d'innovation numérique du programme d'investissements d'avenir.

- Les groupes de travail constitués pour l'élaboration de la feuille de route de l'intelligence artificielle vont proposer des actions et recommandations sur différents sujets, notamment la maîtrise de thèmes de recherche amont, les formations à créer ou à faire évoluer, la question du transfert des résultats de la recherche aux entreprises, le développement de l'écosystème des fournisseurs de technologies de l'intelligence artificielle, ou encore l'accompagnement des transformations sociales induites par l'intelligence artificielle. Ces propositions étant attendues mi-mars 2017, il est aujourd'hui trop tôt pour en donner le contenu précis

### **3. Comment le Gouvernement se positionne-t-il suite à la remise des rapports récents de la présidence des États-Unis et de la Chambre des Communes sur l'intelligence artificielle (octobre 2016) ?**

- Les potentiels, enjeux et risques de l'intelligence artificielle présentés dans les rapports récents de la présidence des États-Unis et de la Chambre des Communes correspondent à ceux communément mentionnés par les experts. Il est intéressant de souligner la place importante des enjeux d'ordre moral et social.

- Par exemple :

- concernant les enjeux d'ordre moral : influence de l'intelligence artificielle sur la liberté de choix des êtres humains, bases éthiques sur lesquelles asseoir les décisions d'une intelligence artificielle,

---

prise en compte des valeurs citoyennes de la société dans laquelle l'intelligence artificielle se développe ;

- concernant les enjeux d'ordre social : concentration des richesses et creusement des inégalités par un déséquilibre accru entre la rémunération du capital (technologique) et celle du travail, concentration du pouvoir, de la connaissance et des choix sociétaux dans les mains des possesseurs d'une technologie devenue incontournable ;

- Concernant les pistes de réflexion et propositions d'actions de l'administration Obama, on peut noter le souhait louable de ne pas laisser les technologies d'intelligence artificielle entre les seules mains d'intérêts privés, qui pourraient confisquer le potentiel économique de l'intelligence artificielle et biaiser le comportement des technologies en fonction de leurs intérêts financiers. L'administration Obama suggère pour cela un effort public massif et une diffusion ouverte des résultats de recherche sur l'intelligence artificielle. Il convient cependant de souligner que les six grands acteurs du numérique nord-américains que sont Google, Amazon, Facebook, Apple, Microsoft et IBM, investissent déjà annuellement en R&D plus de 45 Md\$. L'intégralité de ce montant n'est pas consacrée aux technologies d'intelligence artificielle, mais il illustre la puissance de ces groupes et le fait que les intérêts privés de grands groupes internationaux vont nécessairement influencer le développement de l'intelligence artificielle et, sans doute, rendre difficile l'ouverture des résultats de recherche. Aussi, au-delà de l'effort public massif suggéré par l'administration Obama, un encadrement, sur le long terme, concernant la transparence des processus de décisions et l'usage des technologies d'intelligence artificielle sera nécessaire.

- La démarche conduite par l'administration américaine est, d'autre part, intéressante dans sa nature même : l'État doit s'interroger sur son rôle pour permettre un développement de l'intelligence artificielle profitable à l'ensemble de la société. Différents scénarios sont, en effet, possibles et il appartient à l'État de savoir lequel il veut promouvoir pour prendre les initiatives qui sont adaptées. C'est également dans cette optique que la feuille de route pour l'intelligence artificielle a été lancée.

#### **4. Y a-t-il des opportunités que la France pourra selon vous saisir grâce à l'intelligence artificielle ? Quelles sont-elles ?**

- Les opportunités semblent aujourd'hui grandes ouvertes dans beaucoup de secteurs (droit, éducation, finances, agriculture, transports, etc.). La France se distingue d'ores et déjà particulièrement sur l'innovation dans les domaines de l'énergie et de la santé, portée par la robotique et les objets connectés. Il convient à ce titre de

---

rappeler que les opportunités en matière d'intelligence artificielle sont autant économiques que sociales et sociétales.

- La France dispose également de plusieurs atouts qui peuvent lui permettre de tirer profit du développement de l'intelligence artificielle.

- L'excellence de la recherche française dans des domaines scientifiques et technologiques clés pour l'intelligence artificielle (mathématiques, informatique, sciences des données) est reconnue internationalement. Cette excellence attire les grands groupes internationaux qui implantent sur le territoire national des centres de R&D spécialisés. L'installation de centres de R&D du groupe nord-américain Facebook et du groupe japonais Rakuten en sont des illustrations.

- Au-delà de la dimension scientifique et technique, de nombreux chercheurs français en Sciences Humaines et Sociales s'intéressent également aux impacts socio-économiques, juridiques et éthiques de l'usage des technologies en général, de l'intelligence artificielle en particulier. C'est à la fois un facteur essentiel à la construction et au développement d'une stratégie d'intelligence artificielle conforme aux valeurs de la société que nous défendons, et sans doute un vecteur du rayonnement et de la différenciation de la France en matière d'intelligence artificielle.

- Un environnement favorable à l'émergence et au développement des *start-ups* a été progressivement développé et constitue une réussite indéniable. Il est notamment le résultat de l'existence de laboratoires d'excellence, d'universités et d'école d'ingénieurs de très haut niveau, ainsi que de la volonté de l'Etat d'encourager l'entrepreneuriat et la création d'entreprises, en particulier dans le domaine du numérique avec la French Tech. Le fonds d'investissement ISAI a ainsi recensé plus de 180 *start-ups* en intelligence artificielle en France. Ces *start-ups* contribuent au rayonnement de la France dans le domaine de l'intelligence artificielle, créent des emplois et vont faire naître des leaders et permettre à des entreprises de trouver des relais de croissance.

- La France compte, par ailleurs, de grandes entreprises de services du numérique, de dimension nationale ou mondiale (Capgemini, Atos, Sopra-Steria ...) qui investissent dans les technologies d'intelligence artificielle. Ces entreprises pourront notamment profiter de la croissance prévue du marché de l'intelligence artificielle pour les applications en entreprise (marché qui devrait passer de 200 M\$ en 2015 à plus de 11 Md\$ en 2024 selon le cabinet d'analyse Tractica) et contribueront à l'appropriation de ces technologies par leurs clients, en particulier français.

- Enfin, l'existence de très grands groupes français utilisateurs de technologies, qui sont des références dans leur secteur, constitue

---

également un atout. En intégrant au plus tôt des technologies d'intelligence artificielle, ces très grands groupes, comme Renault, PSA, Airbus, BNP Paribas, vont développer de nouvelles offres innovantes qui dynamisera leur croissance et stimulera le marché français de l'intelligence artificielle.

### **5. Quels sont les principaux risques des technologies d'intelligence artificielle selon vous ?**

- Les technologies d'intelligence artificielle sont communément associées à l'idée de systèmes autonomes, c'est-à-dire capables de prendre leurs propres décisions. En fonction des technologies d'intelligence artificielle, la manière dont la machine va prendre une décision n'est pas toujours formellement comprise et maîtrisée. Deux systèmes d'intelligence artificielle pourraient donc prendre des décisions différentes dans un contexte identique, ce qui présente des risques dans de nombreux cas d'utilisation. La poursuite de travaux de recherche sur les sujets liés à la maîtrise des processus de décision par les technologies d'intelligence artificielle est donc indispensable.

- Un autre facteur important de risques serait une intelligence artificielle n'intégrant pas de paramètres d'ordre éthique dans son processus de prise de décision. La définition de bases d'éthique faisant consensus et les modalités de leur alimentation au fil des évolutions de l'intelligence artificielle, est un sujet à traiter en amont et qui devra faire l'objet d'une attention permanente. La prise en compte des sujets d'éthique nécessitera également de développer des outils permettant de démontrer que ces bases ont bien été intégrées dans les technologies d'intelligence artificielle utilisées. Ces outils devront s'intégrer dans des processus de contrôle qu'il convient également de définir dès à présent, et de faire évoluer en permanence.

- Le développement des technologies d'intelligence artificielle qui va s'accompagner d'une facilitation de leur usage peut, d'autre part, faire peser des risques sur la sécurité des Etats, des organisations et des citoyens. Mises entre les mains d'acteurs malveillants, ces technologies peuvent représenter une capacité importante de nuisance. Une maîtrise française des technologies d'intelligence artificielle (et pas uniquement de leur utilisation) répond donc à un enjeu de souveraineté et de sécurité nationale.

- Enfin, on ne peut s'abstenir d'évoquer les risques d'accroissement des inégalités sociales associés à l'avènement de l'intelligence artificielle - bien que cela n'y soit pas spécifique - entre les citoyens capables d'y accéder et d'en tirer tout le bénéfice, et ceux qui en resteront exclus. La question de l'inclusion numérique doit à plus forte raison rester une priorité.

---

**6. En matière de régulation de l'intelligence artificielle, quelles sont les priorités ? Faut-il aller jusqu'à légiférer ? Si oui, cela doit-il se faire au niveau national, européen ou international ?**

- Les priorités doivent se porter sur les utilisations de l'intelligence artificielle qui posent directement des risques sur la sécurité, la santé et le bien-être des citoyens. La régulation pourrait notamment s'attacher à prévoir les situations dans lesquelles les réactions d'une intelligence artificielle doivent être préalablement connues ainsi qu'à définir les responsabilités liées à l'utilisation d'une intelligence artificielle et aux conséquences de ses réactions.

- Une éventuelle réglementation ou une régulation de l'intelligence artificielle bénéficierait, du fait de la nature et de l'impact de la technologie, d'une échelle la plus large possible : européenne a minima pour les questions éthiques et sociales, pour éviter les pratiques de forum shopping des entreprises, internationale pour les questions d'armements (robots tueurs). Le Parlement européen s'est d'ailleurs saisi du sujet, par l'intermédiaire de sa commission des affaires juridiques, qui a adopté, sur la base d'un rapport, une résolution appelant la Commission à proposer une directive dédiée aux questions de robotiques et d'intelligence artificielle. Les recommandations du rapport (taxe sur le travail des robots, agence européenne de l'intelligence artificielle) mériteraient néanmoins d'être instruites en profondeur, en raison de leur impact social et économique potentiel important.

- Enfin, compte-tenu du manque de recul que nous avons aujourd'hui en réalité, la régulation encadrant l'intelligence artificielle devra certainement prévoir des éléments de souplesse qui permettront à la fois des ajustements en fonction des usages et données réelles, et de ne pas décourager les investisseurs par une régulation *a priori*.

**7. Est-il suffisant de fixer un cadre éthique sans légiférer ? Si oui, à quelles conditions ? Quel contenu envisagez-vous pour ce cadre éthique ?**

- La réflexion autour du cadre législatif permet de porter les questions d'éthique dans le débat public. Des règles d'éthiques admises universellement n'existant pas, il est important que certains sujets, particulièrement sensibles, fassent l'objet d'un débat public permettant une sensibilisation des citoyens aux questions en jeu, le recueil le plus large possible des avis, leur discussion et une prise de décision démocratique garantissant une certaine solidité des résultats obtenus. La question même de ce qui relève de l'éthique en matière d'intelligence artificielle fait aujourd'hui débat. C'est pourquoi l'éclairage de la recherche - notamment en SHS - en la matière est également fondamental. En tout état de cause et comme indiqué précédemment, des processus et outils de contrôle validant la

---

conformité aux bases d'éthique définies collectivement, est une piste opérationnelle.

## **II. - Questions spécifiques pour Madame Axelle Lemaire, Secrétaire d'État chargée du Numérique et de l'Innovation**

### **1. Comment est assuré le suivi des principaux défis que représente l'intelligence artificielle ?**

- Au niveau du cabinet, un conseiller est spécifiquement en charge du suivi de ce sujet et de la cohérence des actions qui sont décidées et mises en œuvre. (Peut-être, si opportun, mettre en avant des contacts réguliers avec certains acteurs ?)

- Au sein de la DGE, administration rattachée au ministère de l'économie et des finances, l'intelligence artificielle est une thématique à part entière prise en charge par le Service de l'Économie Numérique et à laquelle des agents sont dédiés.

- Comme mentionné précédemment, les travaux sur l'élaboration d'une feuille de route sur l'intelligence artificielle ont été officiellement lancés le 20 janvier. En fonction des résultats de ces travaux, un suivi systématisé de certains sujets sera mis en œuvre par le cabinet en lien avec les Administrations concernées.

### **2. Pouvez-vous présenter la « French Tech » et préciser la place qu'occupent en son sein les entreprises mettant en œuvre des technologies d'intelligence artificielle ?**

- La « *French Tech* » est un terme générique pour désigner tous ceux qui travaillent dans ou pour les *start-ups* françaises en France ou à l'étranger. Les entrepreneurs en premier lieu, mais aussi les investisseurs, ingénieurs, designers, développeurs, grands groupes, associations, medias, opérateurs publics, organismes de recherche et universités... qui s'engagent pour la croissance des *start-ups* d'une part et leur rayonnement international d'autre part. Le nombre de *start-ups* en France est estimé entre 7 000 et 10 000.

- « La *French Tech* » désigne également une action publique innovante pour soutenir et promouvoir cet écosystème des *start-ups* françaises dans l'objectif de favoriser leur croissance et l'émergence de nouveaux grands leaders de l'économie française, créateurs de valeur et d'emplois.

- Des acteurs de la « *French Tech* », investisseurs et entrepreneurs français, sous la houlette du fonds d'investissement ISAI, se sont fédérés en 2016 au sein de l'initiative « FranceisAI » pour recenser et

---

promouvoir à l'international les *startups* françaises utilisatrices de technologies IA : près de 200 *start-ups* ont ainsi été identifiées ([franceisai.com](http://franceisai.com)). « FranceisAI » participe activement à l'initiative « #FranceIA »

### **3. Quelles sont les principales opportunités qui vont s'ouvrir pour nos entreprises grâce aux innovations en intelligence artificielle ?**

- Le fonds ISAI recense, en France, plus de 180 *start-ups* dans le domaine de l'intelligence artificielle. Ce chiffre illustre les opportunités de développement de nouvelles activités qu'offre l'intelligence artificielle en France, en particulier du fait d'un potentiel technique et scientifique important dans le domaine (excellence de la recherche et des formations) et des initiatives prises pour favoriser l'entrepreneuriat.

- Le dynamisme de l'entrepreneuriat français dans le domaine de l'intelligence artificielle attire l'attention des investisseurs. Plusieurs levées de fonds significatives ont ainsi eu lieu en 2016, notamment celle de 8,9 M€ de Shift Technology et de 5,6 M€ de Snips. Cette semaine, le fonds d'investissement français Serena Capital a annoncé la création d'une structure dotée de 80 millions d'euros pour prendre des participations dans des *start-ups* de l'intelligence artificielle et du *big data*. Grâce à ces différentes annonces, c'est finalement tout l'écosystème des *start-ups* français qui est mis en visibilité et qui bénéficie d'un renforcement de son accès aux financements.

- L'intelligence artificielle offre aux entreprises plus matures l'opportunité de profiter de relais de croissance en développant des offres innovantes de produits ou de services, en particulier dans des secteurs où la France dispose d'une solide notoriété comme l'ingénierie informatique (nouvelles offres liées à l'automatisation de processus, à l'intégration de *machine learning*), les services collectifs ou « utilities » (nouvelles offres liées à la « Ville Intelligente »), ou encore la sécurité (nouvelles offres liées aux applications de reconnaissance d'images).

- L'intelligence artificielle va également permettre à nos entreprises de renforcer leur productivité et leurs performances (automatisation des processus, analyses prédictives sur le comportement des consommateurs, maintenances prédictives, etc.).

### **4. Quelles initiatives spécifiques votre secrétariat d'Etat entend-il prendre dans les prochaines semaines en matière d'intelligence artificielle ?**

- La secrétaire d'Etat chargée du Numérique et de l'Innovation et le secrétaire d'Etat chargé de l'Enseignement supérieur et de la Recherche ont officialisé vendredi 20 janvier le lancement des travaux



---

sur l'établissement d'une feuille de route sur l'intelligence artificielle. Un point d'étape à mi-parcours a été effectué le 21 février dernier. Les groupes de travail rendront leurs premiers livrables concernant l'analyse et les enjeux des thèmes qui leur auront été attribués. La fin de travaux est programmée le 14 mars, date à laquelle les groupes de travail remettront leurs propositions d'actions. Un évènement de restitution des résultats des travaux est programmé le 28 mars.

### **5. Quelles incitations peut-on envisager à plus long terme en faveur des entreprises pour y favoriser la recherche et l'innovation en intelligence artificielle ?**

- Dans un premier temps, une prise de conscience du potentiel de l'AI et des changements importants qu'elle peut induire est nécessaire. Les travaux de la feuille de route et la communication qui sera faite sur celle-ci vont y contribuer, notamment en portant le sujet de l'intelligence artificielle dans le débat public.

- Concernant plus spécifiquement le fait de favoriser la recherche et l'innovation en intelligence artificielle en entreprise, plusieurs types de mesures peuvent être envisagés comme la création de formations continues sur les technologies d'intelligence artificielle, la création de centres de ressources partagées en intelligence artificielle pour les entreprises, le lancement d'appels à projets de R&D dédiés, etc. Un groupe de travail constitué dans le cadre de la feuille de route sur l'intelligence artificielle est plus particulièrement chargé d'établir des recommandations sur ce sujet et celui de l'appropriation des résultats de la recherche par les entreprises. Ses conclusions permettront de répondre plus précisément à cette question.

### **III. - Questions spécifiques pour Thierry Mandon, Secrétaire d'État chargé de l'Enseignement supérieur et de la Recherche**

#### **1. En quoi l'intelligence artificielle peut-elle permettre de révolutionner et de moderniser l'éducation et l'enseignement supérieur ?**

Comme pour tous les autres domaines d'action de l'Etat (finances, justice, santé, etc.), l'éducation et l'enseignement supérieur produisent leur lot de données dans l'exercice même de leur action. Ces données, lorsqu'elles sont assemblées, peuvent faire l'objet de traitements déductifs ou prédictifs. Elles doivent alors être mises à la disposition des acteurs, avec les outils appropriés, pour moderniser plus que révolutionner les modes d'apprentissage. Cela concerne par exemple le suivi du travail de l'élève ou de l'étudiant, d'anticiper le décrochement scolaire ou de pointer les parties de cours nécessitant un accompagnement supplémentaire. L'approche de l'Intelligence Artificielle permettra ici de mettre en place une

---

personnalisation de l'apprentissage, de tenir compte de profils d'apprenant de plus en plus diversifiés et contribuer ainsi à l'amélioration de la réussite.

## **2. Comment s'articulent la recherche publique et la recherche privée en intelligence artificielle aujourd'hui ?**

La recherche privée s'oriente essentiellement vers les traitements statistiques des données et vers les modalités de reconnaissances des images, de la voix, etc. Cela concerne autant les grands acteurs d'Internet et des réseaux sociaux cherchant ainsi à valoriser les données acquises sur leurs plates-formes, que les jeunes entreprises innovantes désireuses de mettre au point des applications logicielles exploitant données et nouveaux objets connectés. Il s'agit là d'une forme de maturation des concepts développés dans les laboratoires académiques durant les années passées.

En revanche tout le reste des aspects pourtant essentiels de l'intelligence artificielle (notamment le traitement des langues, les différentes formes de logique, la représentation des connaissances, etc.), en interaction avec les SHS, constituent bel et bien le socle de recherche fondamentale nécessaire à l'avancée de la discipline et du domaine. La recherche académique continue donc de travailler sur des objets amont, sur les ruptures et sur l'interdisciplinarité.

En France, la coopération public / privé en matière de recherche s'effectue sur le terrain. Il existe ainsi des partenariats vertueux entre acteurs académiques et privés autour de la fabrication de nouveaux produits, de la création de *start-up*, de l'embauche de jeunes diplômés issus de Master, d'écoles d'ingénieurs ou d'écoles doctorales.

## **3. Pensez-vous que cette place respective de la recherche publique et de la recherche privée doit évoluer ?**

L'un des enjeux majeurs de l'intelligence artificielle est de pouvoir concilier autour de mêmes réalisations, plusieurs approches issues des sous-domaines de l'intelligence artificielle. Il y a donc là une véritable place pour une collaboration efficiente entre recherche publique (socle fondamental) et recherche privée (exploitation de données massives).

Ces complémentarités devront faire l'objet d'une coordination stratégique, résultant de l'analyse en cours menée dans le programme #FranceIA.

Deux remarques toutefois. Les temporalités d'action de ces deux acteurs ne sont pas les mêmes ; la recherche privée aura toujours besoin d'un calendrier de résultats, d'annonces, de mises à jour. Le secteur public peut lui aussi s'emparer d'un volet lié à l'exploitation massive de données, notamment dans les secteurs qui lui sont légitimes comme l'éducation, la santé, le développement régional, sans pour autant abandonner ces terrains à la recherche privée.

#### **4. Quelle place la France occupe-t-elle dans la recherche internationale en intelligence artificielle ? Comment cette place va-t-elle et peut-elle évoluer ?**

La France occupe une place très enviée dans la recherche internationale en intelligence artificielle. D'une part notre dispositif d'enseignement supérieur dans le domaine, notamment au niveau doctoral, est très performant et bon nombre de jeunes chercheurs trouvent des propositions de travail, tant dans le secteur des entreprises que dans le secteur universitaire au niveau international. D'autre part les laboratoires de recherche français occupent tout le spectre des recherches du domaine de l'intelligence artificielle, peu de pays disposent d'acteurs performants sur l'ensemble des sous-domaines de l'intelligence artificielle (ceci est visible dans les comités scientifiques des grandes conférences académiques internationales pour chaque sous-domaine). Enfin les jeunes entrepreneurs français font preuve de vivacité et d'imagination dans les solutions qu'ils proposent à la mise sur le marché.

Les atouts de la France tiennent dans la capacité des acteurs à tirer profit à la fois de l'approche française en termes de mathématiques et de modélisation, et de l'approche française en terme d'ingénierie et de gestion de projet.

Un autre atout français réside dans la pluralité et la grande qualité de la recherche en sciences humaines et sociales. Sociologues, juristes, psychologues, économistes, ethnologues, linguistes, philosophes, etc. sont nombreux, travaillent dans des laboratoires de premier plan et pour beaucoup d'entre eux sont technophiles. Il y a là un gisement pour stimuler et alimenter des programmes de recherche interdisciplinaires aptes à orienter le monde anthropotechnique qui se dessine.

L'évolution de cette position de la France dans la recherche internationale en Intelligence Artificielle doit tenir compte de risques réels comme la fuite de cerveaux à l'étranger, le contrôle des entreprises naissantes par des acteurs extérieurs, etc. Le plan ambitieux #FranceIA doit aussi réfléchir aux conditions qui permettront d'améliorer la visibilité et l'attractivité de la France pour les chercheurs en intelligence artificielle.

#### **5. Quels sont les domaines de recherche prioritaires selon vous ?**

D'une part il est nécessaire que la France conserve et poursuive le travail des recherches menées sur chaque sous-domaine de l'intelligence artificielle. Il est actuellement impossible de prédire si les créations de valeur viendront demain du web sémantique et de ses ontologies, ou de l'analyse de la voix et des émotions, etc. Sur tous ces points et sur l'ensemble des sous-domaines de l'intelligence artificielle, la France a une position en pointe et doit la garder.

---

D'autre part l'intelligence artificielle n'est plus l'affaire des seuls informaticiens, et il faut aussi solliciter des travaux interdisciplinaires d'envergure, notamment pour relier les SHS et les STIC. L'intelligence artificielle développe des solutions prédictives ou de recommandation soit sur la base de scénarios déterminés, soit à partir de calculs basés sur des exemples déjà rencontrés. En revanche rien ne couvre les domaines imprévus, les changements de tendances, les indécisions humaines, les retournements d'opinion. Là où il y a rupture technologique, il y a aussi rupture de mentalité. En même temps que les recherches sur les technologies, il faut produire des recherches sur les intentions, les détournements, les envies. Plutôt que de parler d'acceptation des technologies, il faut envisager d'emblée l'appropriation de ces technologies.

Dans la cadre du plan intelligence artificielle, un groupe de travail est précisément consacré à la cartographie des activités de recherche en lien avec l'intelligence artificielle et l'identification des enjeux de la recherche fondamentale. C'est sur cette base que nous pourrons aussi mieux orienter les efforts.

**6. Pensez-vous que des progrès majeurs puissent être attendus de l'hybridation entre technologies d'intelligence artificielle ?**

Indiscutablement, oui. À court terme il va s'agir de travailler sur des associations à partir de sous domaines actuels de l'intelligence artificielle : vision et audition, robots et décisions, perception et action, représentation des connaissances et traitement de la langue naturelle, etc. Ces associations vont produire des solutions beaucoup plus performantes, rendues possibles par les capacités d'acquisition et de traitements de données. Elles vont aussi pointer de nouveaux verrous qui alimenteront les questions de recherche.

**7. Pensez-vous que les principaux progrès seront réalisés par la conjugaison entre l'intelligence artificielle et l'intelligence humaine ?**

Des progrès seront probablement accomplis par la conjugaison entre l'intelligence artificielle et l'intelligence humaine. Un humain avec des outils adaptés sera toujours plus performant qu'un humain seul ou qu'une machine seule. C'est par exemple le cas en industrie avec le développement de la cobotique. C'est aussi le cas pour tous les métiers du « care » (de la santé, de l'assistance aux personnes) où un diagnostic rendu plus rapidement, plus précisément, devra toujours être accompagné d'une décision, d'une action qui prend en compte l'humain de façon empathique.

---

## II. RÉPONSES À LA CONSULTATION ORGANISÉE PAR LE PARIS MACHINE LEARNING MEETUP EN VUE DU RAPPORT DE L'OPECST

**Nicolas Beaume (15 février 2017)**

Je suis bio-informaticien spécialisé dans l'utilisation des méthodes d'intelligence artificielle (plus précisément d'apprentissage automatique) en génomique. Vu la quantité toujours croissante de données que nous brassons en génomique et la complexité des problèmes abordés, les méthodes d'apprentissages automatiques sont plus qu'utiles, elles deviennent indispensable.

Je pense que ce champs et celui de l'intelligence artificielle en général va devenir un pivot inévitable de nos sociétés. Inévitable ne signifie pas qu'il faut en faire n'importe quoi et c'est pourquoi je trouve cette initiative particulièrement intéressante. Si mon expertise en la matière peut être utile, je serais ravi d'y participer

**Olivier Guillaume, fondateur/CEO de O<sup>2</sup> Quant, société spécialisé dans les dernières techniques en intelligence artificielle (17 janvier 2017)**

Tout comme l'électricité a augmenté les capacités physiques de l'homme, l'intelligence artificielle augmente de manière significative nos capacités cognitives.

Cette évolution majeure a déjà commencé, on la retrouve déjà dans bon nombres d'applications utiles de la vie courante : smartphones, santé, sécurité, industrie, ... et ce n'est que le début. Demain ce sera les voitures autonomes de manière certaine, et bien d'autres évolutions encore aujourd'hui insoupçonnées.

Avec l'augmentation très récente des données informatiques (le *big data*) et des puissances de calculs, nous entrons maintenant dans une autre dimension, ou le champ des possibles explose. Avec ma société O<sup>2</sup> Quant, nous travaillons sur des projets améliorant de manière significative la productivité ou la condition de l'homme : amélioration de la sécurité, aéronautique, industrie, réduction du gaspillage alimentaire, recherche appliqué en santé avec des résultats extrêmement rapides ...

Même si l'intelligence artificielle pourra être utilisé à des fins malveillantes, je suis très confiant dans l'avenir et à ce que ces nouvelles technologies pourront apporter à l'humanité. Notre pays où il fait bon vivre a bien des atouts, notamment en tant qu'un des acteurs mondial majeur dans l'Intelligence artificielle : Nous avons de nombreuses et très bonnes écoles d'informatique et de mathématique ce qui nous procure les data scientists les

---

plus réputés, le dispositif CIR aident les industriels à faire émerger des projets, et bien d'autres atouts ...

Certains gouvernements comme les États-Unis, le Canada ou l'Allemagne croient et investissent dans cette filière d'avenir.

Parce qu'en France nous avons des chercheurs, des entrepreneurs dynamiques et compétents, il y a un terreau fertile au développement de cette ressource stratégique.

Et c'est aussi à vous, femmes et hommes politiques Français, d'accompagner cette évolution dans notre pays du droit et des libertés, en vous investissant de ce sujet.

**Aurélia Nègre (15 janvier 2017)**

Je suis très enthousiaste et pense profondément que l'intelligence artificielle pourra apporter des améliorations majeures dans la vie des gens, en particulier sur les voitures autonomes mais aussi et surtout la prédiction d'incidents/d'activité pour faciliter le travail de la police, des pompiers et des hôpitaux.

**Dominique Péré, Directeur R&D Éditions Lefebvre-Sarrut  
(1<sup>er</sup> décembre 2016)**

L'Intelligence Artificielle est en passe de poser les fondements d'une nouvelle société :

- Les techniques du *machine learning* vont créer des robots logiciels qui remplaceront les tâches dites intellectuelles à faible valeur ajoutée, au même titre que les robots industriels ont fait disparaître les Ouvriers Spécialisés des usines.
    - Les assistants numériques
    - La classification des textes, images, vidéos, audio ....
    - Le support technique et logiciel de 1er niveau
    - La publication d'informations instantanées (météos, sports, résultats d'élections ....)
  - De nouveaux métiers apparaissent :
    - Data scientists
    - Ingénieurs spécialisés en *big data*
    - Exploitation et contrôles des données générées
- De nouvelles interactions et expériences utilisateurs apparaissent :
- Systèmes conversationnels (*Chat bots*)

- 
- Systèmes adaptatifs de préconisation (boutique Internet, claviers numériques ...)

Les données deviennent un enjeu économique majeur. La plupart des gros projets d'intelligence artificielle sont portés par les GAFAMI (Google, Apple, Facebook, Amazon, Microsoft et IBM) et Baidu qui ont la particularité d'être en très grande majorité situés aux Etats-Unis, sauf un en Chine.

La prépondérance des données sur le logiciel est acquise, selon la logique suivante :

« En échange de l'usage de mes logiciels tu me fournis tes données que je peux utiliser à volonté ».

Pour le *machine learning* les systèmes auront de plus en plus de données d'entraînement et parce que cette technologie est accumulative, ces « intelligences vont devenir de plus en plus performantes » donc utiles et bien sûr utilisées.

Maintenant il se profile derrière l'intelligence artificielle classique, une technologie déjà en marche le *deep machine learning*, c'est à dire le *machine learning* entraîné dans des réseaux neuronaux.

La soudaine amélioration de Google Translate est due à la bascule vers le *deep machine learning*.

Cette technologie pose différents problèmes :

- Qui maîtrise les données d'apprentissage ?
- Comment vont évoluer ces systèmes apprenants quand ils vont apprendre avec les données produites par les systèmes de la génération précédente ?
- Quels sont les outils disponibles pour surveiller un « Neural Network » ?
- Quelles garanties aura-t-on d'un fonctionnement non biaisé et bienveillant des intelligences ?

Notre vie quotidienne va être séquencée par cette technologie, il est important que le « Politique » s'empare du problème, le comprenne et agisse en conséquence. Par exemple créer une obligation aux systèmes d'apprentissage automatique d'intégrer de nouvelles données d'apprentissage, pour éviter le phénomène d'entonnoir. Avoir les moyens d'observer la qualité et la pertinence des solutions calculées.

Dans notre métier d'éditeur juridique nous voyons se profiler des « robots juges » qui vont prendre en charge les délits mineurs. Nous voyons aussi des « robots avocats » qui vont être capable de construire des raisonnements juridiques sur les cas les plus simples.

Tant que les résultats des systèmes *deep machine learning* légal sont intermédiés par des humains, on est dans le légal assisté par ordinateur. Cependant, imaginons qu'à la suite d'un manquement grave au code de la route à 3h du matin pendant que le propriétaire dort, la voiture autonome recevant la notification du robot juge va le renvoyer vers le robot avocat du propriétaire.

Nous voyons se profiler une nécessaire éthique de l'utilisation des systèmes à apprentissage profond.

Oserais-je aborder les problèmes d'indépendances et ces systèmes fonctionnants dans la sphère militaire.  
(<https://www.youtube.com/watch?v=45-WR5RcIp0>)

Comment juger les machines ?  
[http://abonnes.lemonde.fr/idees/article/2016/12/01/comment-juger-les-machines\\_5041562\\_3232.html](http://abonnes.lemonde.fr/idees/article/2016/12/01/comment-juger-les-machines_5041562_3232.html)

Comment vont évoluer les rapports sociaux, lorsque nous serons entourés par des machines intelligentes nous assistant quotidiennement et travaillant à notre place ?

Va-t-on faire une redéfinition de la valeur travail, et de la place de l'homme au regard du travail ?

Une véritable révolution sociétale est en marche, pour ma part je souhaite que nous évitions la « *rage against the machine* » des Canuts du numérique, tout autant que l'abandon de notre destin au profit des machines, toutes intelligentes soient-elles ?

### **Tangi Vass (10 novembre 2016)**

L'enjeu à mon sens n'est pas tant les hypothétiques dangers de l'intelligence artificielle dans l'absolu et de la singularité - c'est-à-dire le moment où l'intelligence artificielle dépassera l'intelligence humaine - que la rapidité et la brutalité des impacts socio-économiques de la 3ème révolution industrielle, déjà largement amorcée, dont l'intelligence artificielle est le cœur.

Le cabinet Roland Berger prévoit qu'en France, « 3 millions d'emplois pourraient être détruits par la numérisation d'ici 2025. »

« Tout l'enjeu repose donc sur la capacité de l'économie française à produire les nouvelles activités qui se substitueront à celles où les gains de productivité ont réduit le nombre d'emploi, de manière similaire à la substitution de l'industrie par les services au XXème siècle. »



---

La difficulté de cette transition est accentuée par le fait que dans l'économie de la data, *"the winner takes it all"*, il n'y a pas de place pour des acteurs moyens.

L'échelle nationale n'est d'ailleurs sans doute pas la bonne, comme elle ne l'était pas déjà pour l'aéronautique.

La ressource stratégique de cette 3ème révolution industrielle est la data, le "carburant" de l'intelligence artificielle. L'accès et le contrôle de la data est donc un enjeu géostratégique, comme l'étaient ceux du pétrole dans l'ère précédente. Des cyber-guerres liées au contrôle de cette ressource sont déjà en cours, des guerres traditionnelles sont possibles dans un avenir proche.

*« Nous n'avons pas de meilleurs algorithmes. Nous avons tout simplement plus de données »* Peter Norvig, directeur de recherche chez Google

Lors du meetup d'hier, le représentant d'Amazon a présenté la technologie du moteur de recommandation d'Amazon, récemment ouverte en open source, tout en plaisantant sur le fait qu'évidemment les données d'Amazon n'étaient pas incluses. C'est un peu comme si l'Arabie Saoudite offrait gratuitement des voitures.

Les États-Unis sont aujourd'hui l'hyperpuissance de la donnée.

**Citation de Stéphane Frénot (INSA Lyon) et Stéphane Grumbach (Inria) :** *« Les données personnelles, tant celles directement produites par les usagers, textes, photos, vidéos, etc., que celles indirectement générées bien souvent à notre insu par les systèmes que nous utilisons, sont au cœur de l'économie de la société de l'information, au cœur, donc, de l'économie en général. »*

*« Dans ce domaine, les États européens ont fait l'impasse, ou tout au moins ont échoué à promouvoir les entreprises de la nouvelle économie. En ne construisant pas d'industrie du Web 2.0, l'Europe s'est privée de l'accès à la ressource, y compris à la ressource provenant de ses territoires. Pour des raisons historiques et politiques, l'Europe a peur des données. Elle voit dans la société de l'information une immense menace qu'il convient de circonscrire, et qui semble inhiber toute vraie ambition, sans se rendre compte que moins on développe l'industrie de l'information, plus nos données quittent le territoire. »*

*« La maîtrise de données permet également la maîtrise de certains marchés qui transitent déjà majoritairement dans certains domaines par les outils de commerce électronique américains. Faute de développer cette industrie, il est probable qu'à brève échéance nous achèterons de très nombreux biens et services produits et consommés en France, comme nos billets de train ou notre électricité, à un prestataire étranger qui dégagera une part de la valeur ajoutée et aura le contrôle de la partie la plus rentable de la chaîne industrielle, laissant les parties coûteuses comme les infrastructures à ses partenaires. » .*

**Charly FLORENT (29 août 2016)**

---

L'intelligence artificielle est un domaine extrêmement fantasmé sans être réellement connu pour ses applications concrètes. Une bonne partie de mon travail au jour le jour en tant que Data Scientist dans une grande entreprise est la vulgarisation et l'évangélisation autour de ce sujet : que peut réellement faire la Data Science ?

Les machines, même si elles peuvent "apprendre", sont limitées par les algorithmes qui les animent. L'apprentissage à proprement parler est mathématique, c'est une reconnaissance de schémas, l'intelligence artificielle ne crée pas. Ceci étant, elle peut à partir de ce qui existe déjà reconnaître des schémas que nous n'avions pas perçus au préalable et faire des connexions entre des éléments qui n'étaient pas évidentes. C'est pour cette raison qu'il est possible aujourd'hui de battre des champions d'échec ou de go, pas en inventant de nouvelles stratégies mais bien en décortiquant les anciennes pour sublimer le mécanisme du jeu.

En conclusion, je pense que c'est une excellente initiative que l'intelligence artificielle soit comprises par les hautes sphères du gouvernement car elle n'est pas à négliger. Cette recherche de schémas peut s'avérer extrêmement puissante dans des domaines peu soupçonnés et apporter une amélioration nette à certains services publics (on pense déjà à la santé avec les carnets de santé virtualisés ou à la détection de fraudes fiscales).

**Jean-Armand Moroni - ingénieur Centrale Paris avec une spécialisation en intelligence artificielle (28 août 2016)**

L'intelligence artificielle est de nouveau sur le devant de la scène, après un moratoire de 1990 à 2005.

À l'actif :

- Des succès récents : automobile autonome, catégorisation d'image, reconnaissance de personne, IBM Watson à Jeopardy, Go...
- Une dynamique forte et relativement ouverte : communauté de chercheurs, d'amateurs, de sociétés (les grands du Web), de sites (Kaggle pour le data mining), mouvement open data, etc. L'intelligence collective du Web.
- Des données en quantité suffisante, des algorithmes pour les traiter (en particulier le *deep learning*), des puissances de calcul qui continuent de croître, des outils open source dédiés (bases NoSQL, bibliothèques ex. TensorFlow).
- Des robots avec des capacités physiques qui s'améliorent (même si toujours loin de celles des humains) ; des plateformes accessibles

(Arduino, Raspberry, capteurs et actionneurs) et l'impression 3D simplifient l'expérimentation.

Au passif :

- Toujours pas de conversation possible avec une intelligence artificielle. Toujours pas de sens commun. Les succès remportés ne vont pas au-delà de ce que sait faire un animal, qui n'a pourtant pas accès au langage structuré. Risque d'atteindre le plus haut du "hype cycle" et d'en dégringoler ensuite : un troisième "hiver de l'IA".
- Les algorithmes d'apprentissage actuels semblent beaucoup moins efficaces que ceux mis en œuvre par un humain (ex. la quantité de données nécessaires) : est-on sur la bonne piste ?
- Les *open data* sont hétérogènes ; l'absence de standardisation complique leur croisement ; le semantic Web patine.
- Les retombées commerciales du *big data* pour les sociétés qui y investissent ne sont pas évidentes, en dehors de cas particuliers type Google.
- Très forte domination des États-Unis pour la mise en œuvre, même si la recherche est beaucoup mieux distribuée (ex. Yann LeCun, père français du *deep learning*).

---

## ANNEXE 9 : COMPTE RENDU DE L'AUDITION PUBLIQUE DU 19 JANVIER 2017

### I. INTRODUCTION

#### 1. M. Bruno Sido, sénateur, premier vice-président de l'OPECST

Monsieur le président, Mesdames, Messieurs les députés, Mesdames, Messieurs les sénateurs, Mesdames, Messieurs les professeurs, en ma qualité de sénateur et de premier vice-président de l'Office parlementaire d'évaluation des choix scientifiques et technologiques, j'ai le plaisir de vous accueillir ce matin au Palais du Luxembourg pour une journée d'auditions publiques sur l'intelligence artificielle (IA).

Le président de l'OPECST, M. Jean-Yves Le Déaut, député, vous présentera en détail l'originalité de cette délégation parlementaire bicamérale. Je me contenterai de saluer le travail entrepris depuis quelques mois, et à un rythme soutenu, par la sénatrice Dominique Gillot et le député Claude de Ganay, les deux rapporteurs désignés par l'Office pour répondre à la saisine émanant de la commission des affaires économiques du Sénat sur l'intelligence artificielle.

D'ailleurs, il est à noter que Mme Dominique Gillot espérait depuis plusieurs années travailler sur ce thème. C'est dire que l'Office ne découvre pas aujourd'hui le sujet de l'intelligence artificielle, d'autant que cette audition publique a été précédée de plusieurs autres études ou auditions publiques abordant déjà ce thème, sous un angle ou un autre. Je pense, par exemple, à la récente audition publique sur les robots et la loi, mais aussi à celle sur le numérique au service de la santé ou encore à celle sur l'agriculture et le traitement des données massives.

En outre, j'ai eu l'honneur d'être co-rapporteur de l'étude relative à la sécurité numérique, qui pourrait être beaucoup mieux garantie si les recommandations préconisées par l'Office étaient suivies, notamment celles qui constituent son *vade-mecum* pour la sécurité numérique des entreprises.

L'intelligence artificielle a donc déjà piqué la curiosité des membres de l'Office, qui ne manqueront pas de regarder avec intérêt la vidéo de la présente audition sur le site du Sénat.

Quant à la publicité ultérieure des travaux de ce jour et de leurs retombées positives dans l'étude à paraître, il faut savoir que chaque étude de l'Office, incluant les contributions des intervenants sous forme de comptes rendus, donne lieu à la présentation de ses conclusions devant la commission qui l'a saisi puis, devant le Sénat, à l'occasion d'un débat en séance plénière auquel participe le ministre compétent : Mme Axelle Lemaire, secrétaire d'État chargée du numérique et de l'innovation, nous fera d'ailleurs l'honneur d'être parmi nous cet après-midi.

---

Avant de céder la parole au président de l'Office, je souhaiterais également souligner que l'OPECST ne se contente pas de travailler le plus en amont possible du sujet qu'il traite, mais qu'il essaie également de bien orienter sa réflexion vers les aspects les moins bien explorés.

En l'occurrence, il ne s'agit pas, pour nous, de dresser un panorama de l'historique des différentes vagues d'intérêt suscitées depuis un demi-siècle par l'intelligence artificielle. Elles sont présentées, à chaque occasion, comme une révolution décisive devant tout emporter sur son passage, ce qui a été largement démenti à plusieurs reprises.

Cette fois-ci, il semblerait que divers éléments se conjuguent pour que la prochaine vague soit plus puissante que les précédentes et induise un changement de société qui pourrait dépasser nos perspectives actuelles d'imagination.

Dans ce contexte, il me semble que les rapporteurs de l'Office ont fort opportunément choisi de dresser une cartographie mettant plus particulièrement en valeur un aspect original, à savoir l'approche éthique des problèmes posés par l'émergence accélérée de l'intelligence artificielle, sans oublier les interconnexions ultérieures qui ne manqueront pas de se nouer avec des domaines inattendus. Mais je n'en dirai pas plus pour laisser à la fois au président et aux deux rapporteurs le plaisir et l'honneur de vous présenter l'Office et de vous exposer leur démarche.

Je vous remercie de nouveau de vous être déplacés et d'avoir mobilisé toutes vos capacités cognitives pour rendre notre intelligence du sujet la moins artificielle possible.

## **2. M. Jean-Yves Le Déaut, député, président de l'OPECST**

Mesdames, Messieurs, je vous remercie également d'être venus ce matin. Le sujet que nous allons aborder est en prise directe avec une actualité brûlante. L'impact du développement de l'intelligence artificielle sur nos sociétés est au cœur des discussions du forum de Davos ; ce thème récurrent est évoqué depuis plusieurs années. Mme Mady Delvaux, que nous recevrons cet après-midi, a publié, au titre d'un groupe de travail de la commission des affaires juridiques du Parlement européen, un rapport sur les règles de droit civil concernant la robotique et l'intelligence artificielle, dont les conclusions donnent lieu à discussion dans la mesure où est, notamment, posée la question de la responsabilité des robots.

Moi-même, en tant que rapporteur général des Sciences et des technologies de l'Assemblée parlementaire du Conseil de l'Europe, je prépare un rapport sur les convergences technologiques, l'intelligence artificielle et les droits de l'homme, en le ciblant plus particulièrement sur les problèmes que pose l'intelligence artificielle au regard des droits de l'homme. Plusieurs rapports vont être publiés sur le sujet. Aussi, il est

---

opportun que l'Office s'en soit saisi, en chargeant Dominique Gillot et Claude de Ganay de réaliser cette étude.

Notre manière de travailler est originale - elle a de plus en plus cours à l'Assemblée nationale et au Sénat - en ce qu'elle consiste à organiser des auditions publiques collectives contradictoires où interviennent des parlementaires et des représentants de la société civile, qu'il s'agisse d'universitaires, de chercheurs ou de responsables d'associations. Ces discussions, au cours desquelles des avis différents s'expriment, nous permettent d'envisager des évolutions législatives. L'Office travaille en amont de la législation. Sur des sujets complexes, tels que celui de l'intelligence artificielle, il est évident que les parlementaires ne sont pas armés de la même manière pour les décortiquer. Or légiférer exige de bien comprendre les sujets.

De manière très simpliste, certains déplorent l'absence de parlementaires dans les hémicycles, mais, pour débattre d'un sujet et légiférer, il faut se l'être approprié. Or c'est compliqué. Tel est le rôle de l'Office. D'ailleurs, de plus en plus de collègues nous posent des questions simples. On nous a récemment interrogés sur le compteur *Linky* par exemple ; on nous a aussi demandé pourquoi on ne développe pas des vignes résistantes au mildiou ou à l'oïdium. On examine ces questions et on essaie de trouver des solutions.

L'intelligence artificielle est un sujet majeur, et il sera très difficile de régler la question de la législation à l'échelon national - une proposition va nous être transmise à l'échelon européen. Il convient à l'évidence de traiter nombre de ces sujets à l'échelle internationale. Dès lors que le robot devient intelligent, que la frontière entre l'homme et la machine va s'estomper ou, plutôt, que l'homme et la machine vont continuer à travailler ensemble, qui sera responsable ? Le problème de la responsabilité et de la réparation des dommages se pose. On parle souvent de robots automatisés dans le domaine de la guerre, mais ce ne sont pas les seuls - heureusement ! L'audition publique sur les robots et la loi, que nous avons organisée en décembre 2015 et qui était préparatoire à celle d'aujourd'hui, a prouvé la pertinence d'aménager les règles du droit pour tenir compte de nouvelles formes de responsabilité, avec le déploiement des robots induits, plutôt que de créer un statut et des régimes de responsabilité spécifiques avec une personnalité juridique pour les robots, une solution à laquelle je n'adhère pas personnellement. Mes collègues rapporteurs en parleront.

Dans le domaine de la médecine notamment, la frontière s'estompe entre l'homme et la machine : on est passé de l'homme soigné à l'homme réparé, mais va-t-on passer à l'homme augmenté ? Certains en parlent. Cette seule question pose des problèmes éthiques. La loi doit permettre de résister à des pressions ou à des contraintes, qui imposeraient à des individus, par exemple, de se soumettre à des technologies de nature à améliorer leurs performances ; je pense aux domaines du sport, du jeu, voire à celui du travail. Toutes ces questions doivent donc être posées.

---

Ces nouvelles technologies induisent déjà des changements de grande ampleur. La frontière entre le médical et le non-médical s'atténue également, et il en est de même pour le naturel et l'artificiel. Avec la biologie de synthèse, on recrée le vivant à partir de molécules recomposées. Pour travailler actuellement à l'élaboration d'un rapport sur le *genome editing*, je puis vous dire que l'on voit aujourd'hui les grands progrès réalisés en matière de modification ciblée du génome, avec un abaissement des coûts. Cela pose aussi un certain nombre de problèmes.

La question de l'équilibre entre les innovations, qui sont nécessaires, et les droits de l'homme est donc posée. D'autres problèmes se posent, tels que les fractures numériques ou l'accès inégal à internet ; des fractures mondiales assez criantes sont également à noter. Le professeur Jan Helge Solbakk du Centre d'éthique médicale à la faculté de médecine de l'université d'Oslo résume bien les choses : le progrès n'est pas synonyme d'accessibilité universelle. L'Office en est d'accord, même s'il faut continuer à employer le mot « progrès », pour lequel nous nous battons, il doit être maîtrisé et partagé. C'est l'un des points sur lesquels vous allez vous exprimer.

En conclusion, j'observe que, au travers de l'intelligence artificielle et de la convergence des technologies NBIC, les nanotechnologies, biotechnologies, informatique et sciences cognitives – la fameuse convergence entre le neuro, le bio, l'info et le cogno –, l'homme va pouvoir augmenter ses capacités grâce aux machines. Toutefois, comme je l'ai souligné précédemment, l'homme augmenté risque de succéder à l'homme soigné : on sait aujourd'hui fabriquer des interfaces fonctionnelles entre le cerveau et l'ordinateur.

Toutes ces questions méritent une réflexion prospective, que nos deux collègues ont souhaité mener. Restons modestes dans un premier temps : ne pensons pas résoudre cette question en pleine évolution et en pleine mutation. Mais les échanges avec des spécialistes et des parlementaires participent d'une bonne méthode de travail. Aussi, je suis très heureux que l'OPECST puisse apporter sa contribution, en vue d'adapter peut-être le cadre législatif français. Les recommandations de nos deux rapporteurs seront débattues au mois de mars prochain ; il s'agira d'ailleurs du dernier rapport de l'Office pendant cette législature. Au titre de nos travaux, nous débattons, la semaine précédente, du rapport relatif au *genome editing* et de l'évaluation de la stratégie nationale de recherche en février. Par ailleurs, une audition publique sur le compteur *Linky* nous a été demandée par nos collègues.

---

## II. PREMIÈRE TABLE RONDE, PRÉSIDIÉE PAR M. CLAUDE DE GANAY, RAPPORTEUR : LES TECHNOLOGIES RELEVANT DE L'INTELLIGENCE ARTIFICIELLE

### 1. Mme Dominique Gillot, sénatrice, membre de l'OPECST, rapporteure

Comme vient de le souligner le président Le Déaut, nous avons l'ambition forte d'aider la décision politique en l'appuyant sur la réflexion scientifique et le dialogue entre les scientifiques et les politiques. Cette gouvernance, souvent réclamée et exaltée, est quelquefois difficile à mettre en place du fait de la frontière qui existe entre ces deux mondes. L'OPECST a vraiment vocation à éclairer les parlementaires et, au-delà, les décideurs de la puissance publique. Tel est notre objectif avec le rapport de l'Office sur l'intelligence artificielle, afin de participer à la formation de la décision politique.

Pour ma part, je modérerai quelque peu les propos de mon collègue : l'intelligence artificielle ne signifie pas seulement que l'homme doive se soumettre, car il peut aussi en bénéficier, profiter de nouvelles opportunités. De même, il ne s'agit pas simplement d'être assujéti, on peut aussi maîtriser pour partager des progrès. J'ai une vision plutôt optimiste de l'avancée de la science pour autant que l'homme se donne les moyens de continuer de l'accompagner et de la maîtriser. Le meilleur rendement, c'est l'homme et la machine ; ce n'est pas la machine seule.

### 2. M. Claude de Ganay, député, membre de l'OPECST, rapporteur

Dominique Gillot vient de préciser le cadre de travail de l'Office.

La première table ronde que j'ai l'honneur de présider porte sur les technologies relevant de l'intelligence artificielle. Nous avons choisi ce titre pour évoquer le caractère multiple de ces technologies et préciser tout particulièrement ce que l'on met derrière le label « intelligence artificielle ».

Je demanderai aux intervenants de respecter un temps de parole de sept ou huit minutes maximum, afin d'avoir un temps raisonnable pour notre débat interactif.

Je laisse tout d'abord la parole à Jean-Gabriel Ganascia, professeur à l'université Pierre-et-Marie-Curie Paris-VI, rattaché au laboratoire d'informatique de l'université Paris-VI. Vous présidez depuis quelques mois le comité d'éthique du CNRS. Physicien et philosophe d'abord, puis informaticien, passionné par les sciences cognitives, vous êtes un spécialiste reconnu de l'intelligence artificielle. Vous avez d'ailleurs publié de nombreux ouvrages à ce sujet : le prochain, qui devrait être publié en février, devrait s'intituler : « *Le mythe de la singularité. Faut-il craindre l'intelligence*



---

*artificielle ?* » Nous allons vous écouter afin de répondre de manière argumentée à cette question.

### 3. M. Jean-Gabriel Ganascia, professeur à l'université Pierre-et-Marie-Curie, Paris-VI

Monsieur le président, Mesdames, Messieurs les sénateurs, Mesdames, Messieurs les députés, je vous remercie tout d'abord de m'avoir invité. Pour respecter scrupuleusement le temps qui m'a été alloué, j'entre dans le vif du sujet.

J'ai intitulé mon exposé sur l'intelligence artificielle « *Naissance et renaissances de l'intelligence artificielle* ». Comme l'a relevé précédemment M. Sido, il existe différentes renaissances de l'intelligence artificielle. D'une certaine façon, il y a un caractère cyclique, que j'appelle « un temps tressé ». Les technologies de l'information en général font que des technologies oubliées redeviennent importantes.

L'autre élément important tient à l'article défini singulier : l'intelligence artificielle. Il s'agit d'une discipline scientifique ; on ne peut donc pas parler d'une intelligence artificielle. Cela substantifierait les objets mus par l'intelligence artificielle, ce qui, à mon sens, est incorrect.

Permettez-moi de faire une petite rétrospective. L'idée de faire une machine qui raisonne est très ancienne : cela remonte à Leibniz. Au XIX<sup>e</sup> siècle, Jevons a utilisé l'algèbre de Boole pour faire un piano mécanique, qui faisait du raisonnement de façon automatique. Ensuite, en 1943, avant la construction du premier ordinateur électronique, la cybernétique - ce nom ne sera donné qu'en 1946 - s'est attachée à la simulation du raisonnement sur des machines.

Sans entrer dans le détail, deux articles princeps paraissent en 1943, un article sur la notion de rétroaction et un second sur la notion de réseaux de neurones formels, une idée introduite par deux personnes, Warren McCulloch et Walter Pitts, un jeune mathématicien alors âgé de vingt ans. Ce dernier montre que des réseaux très simples, avec de petits automates organisés sur trois couches connectées entre elles, permettent de réaliser n'importe quelle fonction logique. Il établit alors un parallèle entre ce que l'on connaissait à l'époque du cerveau et l'ingénierie, cette fabrication d'outils. Toutefois, il faut réussir à connecter les trois couches du réseau entre elles, et l'apprentissage de ces connexions fera l'objet des travaux menés dans les soixante ou soixante-dix années qui suivent.

La vraie naissance de l'intelligence artificielle date de 1956 : un programme de recherche a été écrit par deux jeunes mathématiciens âgés de vingt-huit ans, John McCarthy et Marvin Minsky. Ce programme de recherche est primordial - c'est l'équivalent de ce que fut, en son temps, la déclaration de Galilée au début de l'âge moderne lorsqu'il affirme que la

---

nature s'écrit en langage mathématique : l'esprit peut se décomposer et se simuler avec une machine. Je l'ai souligné dans ma diapositive, tous les aspects de l'apprentissage et toutes les autres caractéristiques de l'intelligence peuvent être décrits de façon si simple qu'une machine peut les simuler. C'est ce coup d'envoi épistémologique de l'intelligence artificielle qui est central pour l'âge moderne.

Quelles sont les différences caractéristiques de l'intelligence ? J'ai utilisé une décomposition classique qu'utilisent les spécialistes des sciences cognitives, à savoir cinq fonctions cognitives classiques : les fonctions réceptives, avec l'analyse des formes, des images, des sons, etc. ; la mémoire et l'apprentissage ; les questions de raisonnement et de pensée, qui sont simulées par les machines ; les fonctions expressives - comment parler ? Comment échanger ? - et les fonctions exécutives - les robots dont on a parlé tout à l'heure.

Je parlerai maintenant des renaissances. Quelles sont les techniques ? Elles sont nombreuses. Mais je me contenterai de vous parler de l'apprentissage profond, le *deep learning*. Le cycle des renaissances commence en 1943 avec les réseaux de neurones formels : un réseau à trois couches peut réaliser n'importe quelle fonction logique : c'est une propriété d'universalité des réseaux. Le problème réside dans le fait d'établir des connexions entre la première et la deuxième couche, et la deuxième et la troisième. Ce travail de programmation étant extrêmement fastidieux, on essaie de le faire automatiquement. Marvin Minsky avait essayé de faire ce travail dans sa thèse, sans y parvenir. En 1959, Frank Rosenblatt avait inventé le perceptron, mais ce système n'apprenait les connexions que sur le réseau à deux couches, ce qui était très pauvre.

La nouvelle renaissance intervient en 1986, avec des théories mathématiques nouvelles, la *Back Prop*, la rétropropagation de gradient permet d'assurer l'apprentissage. Des mathématiciens essaient de formaliser tout ce travail avec d'autres techniques. À partir de 2010, on réutilise les réseaux de neurones non plus à trois, mais à treize couches, mais je n'entrerai pas dans le détail. Les résultats sont époustouflants, notamment pour la reconnaissance des formes. Je citerai un exemple marquant, la reconnaissance des visages. Différentes sociétés s'intéressent à ce sujet, la société *FaceNet* de Google ou *DeepFace* de Facebook entre autres : pour la première, le taux de reconnaissance est de 99,63 %, mais il faut apprendre sur 200 millions d'images, ce qui pose un certain nombre de questions, que nous aborderons ultérieurement.

Autre enjeu majeur, voilà moins d'un an, la machine AlphaGo a battu celui que l'on considérait à l'époque comme le meilleur joueur de Go au monde. Certes, il s'agit d'un jeu, mais celui-ci est extrêmement sérieux. Du point de vue scientifique, John McCarthy disait des jeux qu'ils étaient la drosophile de l'intelligence artificielle. C'est en fait ce qui va permettre de tester l'ensemble des techniques, et cela fait appel à de l'apprentissage

---

profond, ainsi qu'à d'autres techniques d'apprentissage, tel l'apprentissage par renforcement.

Pour clore cette présentation, je parlerai des enjeux liés à l'intelligence artificielle : les enjeux éthiques certes, dont nous parlerons ultérieurement, mais aussi les enjeux économiques, qui sont majeurs.

Je citerai une technique qui a suivi toute l'histoire de l'intelligence artificielle, celle des agents conversationnels. Le test de Turing en 1950 imaginait une conversation entre une machine et un homme, la confusion qu'il peut y avoir entre un homme et une machine. Dans les années soixante, Joseph Weizenbaum a aussi imaginé un agent conversationnel. Depuis quelques années, de nombreux travaux ont été publiés sur ce sujet. L'agent conversationnel Tay a défrayé la chronique l'an dernier ; la société Amazon a développé le système Écho, Google l'Assistant, et Apple a fait part de son intention de développer un assistant conversationnel nommé Viv, qui doit prendre la relève de *Siri*. Ils pensent qu'ils vont constituer le segment terminal dans toute une chaîne de valeurs. La publicité sera « *cool* » avec les agents conversationnels : vous pourrez commander une pizza directement de votre bureau. L'enjeu est simple : les personnes qui maîtriseront la technologie de cet agent auront alors un rôle déterminant sur toute la restauration rapide.

J'évoquerai enfin les enjeux politiques, les enjeux de souveraineté.

Si l'on considère les fonctions régaliennes de l'État, à savoir la défense, la sécurité intérieure, les finances et la justice, on voit qu'elles sont toutes transformées par l'intelligence artificielle, car de nouveaux acteurs interviennent.

Pour citer la justice, j'ai été surpris de constater que les États-Unis utilisent des systèmes prédictifs ; cela n'existe pas encore en France. Le site *crowdlaw.org* peut faire de la législation collective en s'adressant à la population. Tout cela fait réfléchir.

Concernant la finance, les États avaient le privilège de battre monnaie ; c'est un peu moins vrai avec l'Europe et cela l'est encore moins avec les monnaies virtuelles.

S'agissant de la sécurité intérieure, dans le cadre de leur programme d'exemption de visas, les États-Unis exigent depuis quelques semaines que soient notés sur le formulaire *Esta* les pseudonymes sur les réseaux sociaux. Cela signifie, d'une certaine façon, que l'état civil est mieux défini par ces réseaux que par l'État. On peut aussi parler de la reconnaissance des visages, qui assure une certaine sécurité. Là encore, l'État ne peut y procéder parce qu'il ne dispose pas des images et n'a pas toutes les capacités pour le faire, même d'un point de vue légal.

Enfin, pour ce qui concerne la défense, dont je croyais qu'elle restait le privilège de l'État, j'ai appris voilà quelques semaines que les grandes entreprises veulent être autorisées à avoir une attitude non seulement

---

défensive sur le cyberspace, mais également offensive, pour se battre contre leurs agresseurs.

**M. Claude de Ganay, rapporteur.** – Monsieur Gérard Sabah, vous êtes directeur de recherche honoraire au CNRS. Vous avez longtemps travaillé au sein du laboratoire d'informatique pour la mécanique et les sciences de l'ingénieur à Orsay, le LIMSI. En tant que membre titulaire de l'académie des technologies, vous avez produit en 2009 une brochure sur l'intelligence artificielle, qui faisait un point très intéressant sur l'histoire de ces technologies. Je vous demande donc de bien vouloir nous dresser le bilan de ces technologies.

#### **4. M. Gérard Sabah, directeur de recherche honoraire au CNRS**

Monsieur le président, Mesdames, Messieurs les sénateurs, Mesdames, Messieurs les députés, je vous remercie de cette invitation. Je suis content de constater que l'audition est ouverte à la presse, même si elle est assez peu représentée : elle sera ainsi informée de nos travaux. Cela lui permettra d'éviter de publier des articles intitulés « Il n'y aura jamais d'intelligence artificielle. L'idée d'intelligence artificielle présuppose qu'il n'existe qu'une seule forme d'intelligence. » Bien sûr, cela n'est pas le cas. Nous avons essayé, dans le cadre des travaux relatifs à l'intelligence artificielle, d'identifier un certain nombre de caractéristiques de cette intelligence. Pour les premières d'entre elles que j'ai listées sur ma diapositive, l'informatique, en général, et l'intelligence artificielle, en particulier, sont capables de réaliser la plupart des éléments notés, et les recherches pointues portent sur la réalisation des derniers éléments de la liste.

Je rappelle très brièvement les principaux outils de l'intelligence artificielle classiques : au tout début, le développement d'arborescences, la recherche dans les arbres, avec des procédures d'amélioration pour accélérer les choses ; le système expert dans les années quatre-vingt ; le développement de logiques non classiques pour essayer de rendre compte de la complexité des raisonnements humains ; la programmation par contrainte, les réseaux de neurones, les algorithmes génétiques, raisonnement par analogie. On le voit bien, différents modes de raisonnement ont été mis en œuvre. Les mécanismes de l'intelligence artificielle sont également très variés.

Trois courants de pensée relativement différents sous-tendent l'intelligence artificielle : l'analogie symbolique, dont Jean-Gabriel Ganascia a parlé, est fondatrice de la discipline – les processus mentaux sont censés se résumer à des manipulations de symboles – ; l'intelligence artificielle distribuée, où la pensée est vue comme un phénomène collectif – le développement de nombreux agents élémentaires peut être synthétisé et

---

produire un comportement intelligent – et la métaphore des réseaux où l'intelligence est considérée comme une diffusion d'activités au niveau des réseaux de neurones.

Je vous ferai part de quelques succès : les échecs, le backgammon, le jeu de Go où le champion du monde a été battu, le développement d'interfaces vocales, et Watson, qui a battu les meilleurs joueurs mondiaux au jeu *Jeopardy*.

À cet égard, je souligne que Watson disposait de quinze téraoctets de mémoire vive et utilisait près de trois mille processeurs. Cet ordinateur est donc loin d'être similaire à ceux dont nous disposons. IBM a dit que, si le programme avait été mis sur un ordinateur de bureau, les logiciels auraient mis deux à trois heures pour répondre à une question. Le jeu consistait à trouver une question dont la réponse lui avait été communiquée. À la question « *Un écrivain dont le héros est Hercule Poirot* », il fallait répondre : « *Qui est Agatha Christie ?* » Watson disposait de deux cents millions de pages d'encyclopédies, de dictionnaires, de livres, d'articles de journaux, de scénarios de films, etc. Comme ses adversaires, il n'avait pas accès à internet, mais il avait toutes ces données en mémoire.

Actuellement, l'apprentissage profond lié aux données massives est à la mode. On dispose de quantités de données absolument extraordinaires en la matière. Comme l'a montré Jean-Gabriel Ganascia, l'apprentissage profond, qui consiste à utiliser de nombreuses couches intermédiaires, permet de mettre en évidence un certain nombre de concepts élémentaires utiles pour la reconnaissance d'images ou la compréhension de textes, par exemple. Cela a été rendu possible par le développement de la puissance de calculs des machines et par l'utilisation de nouvelles puces graphiques, utilisées aussi au niveau du calcul.

Quelles sont les limites ?

Dans ce type de mécanisme, le réseau produit effectivement quelque chose, mais n'a pas de possibilité d'explication. On ne comprend pas forcément pourquoi cela marche. Cela peut être opposé aux travaux relatifs aux mécanismes d'apprentissage à partir de très peu d'exemples. En 2015, dans la revue *Science*, un article a montré comment on pouvait réaliser des apprentissages à partir d'un ou deux exemples.

Le programme *DeepDream* créé par Google reconnaît, disaient-ils, une image de chat après plusieurs centaines de milliers ou de millions d'analyses d'images. Que veut dire reconnaître un chat ? Il ne faut pas imaginer que l'ordinateur a compris ce qu'était un chat ou qu'il a mis en évidence ce concept. En fait, il met dans la même classe toutes les images de chat. Si l'on demande à l'ordinateur d'autres caractéristiques du chat, il ne pourra évidemment pas y répondre.

Je formulerai quelques réflexions futures, en insistant sur deux points.

---

La question de l'hybridation. Le fantasme actuel de l'auto-apprentissage ? Actuellement, on a l'impression que l'hybridation va tout résoudre, mais l'intelligence artificielle ne se résume pas à l'apprentissage machine, en particulier aux techniques de réseaux de neurones profonds, d'apprentissage profond.

L'intelligence artificielle, comme prothèse, vise à ajouter aux compétences humaines et conduit à se demander dans quelle mesure l'homme va pouvoir éventuellement être « augmenté » par les techniques de la machine.

Plus loin, il y a la notion de conscience, avec la capacité à se représenter soi-même et à raisonner sur ses représentations : on pourrait se demander dans quelle mesure les machines pourraient acquérir ce type de capacités.

J'ai essayé de résumer les avantages et les inconvénients des deux approches des systèmes symboliques de l'intelligence artificielle classique et les réseaux de neurones profonds : difficultés à généraliser pour les systèmes symboliques, généralisation naturelle pour les réseaux de neurones, etc. Essayer de faire collaborer ces deux approches présente manifestement un certain nombre d'avantages, dont certains peuvent être regroupés.

Je ne m'attarderai pas sur l'hybridation entre l'homme et l'intelligence artificielle, car Laurent Alexandre, qui doit intervenir cet après-midi, en parlera beaucoup mieux que moi.

Les hybridations entre les systèmes symboliques et les systèmes neuronaux : le but est d'obtenir de meilleures performances globales, comme je l'ai dit précédemment, avec des champs d'application plus larges, afin de permettre l'acquisition de nouvelles connaissances à partir de sources différentes et pas toujours cohérentes.

Comment y parvenir ?

Il faut, par exemple, essayer d'adapter la structure du réseau avec des algorithmes génétiques ; introduire des modules symboliques dans les réseaux, afin de permettre l'utilisation de logiques floues, de systèmes de règles. Diverses questions se posent, que je n'ai pas le temps d'évoquer ici : comment intégrer ces différents modules ? Comment interagissent-ils entre eux ?

Par ailleurs, concernant la conscience : selon Gerald Edelman, prix Nobel en neurobiologie, les fonctionnalités nécessaires à une véritable intelligence sont celles qui sont fondées sur l'inconscient et qui permettent l'émergence de la conscience chez l'homme. Pourquoi cette assertion ne serait-elle pas vraie pour les machines ?

Souvent, la première question que l'on peut se poser est la suivante : un robot ou un système pourra-t-il être conscient ou peut-on reproduire la conscience humaine dans une machine ? À mon sens, ce n'est pas la bonne question. Il faut se demander : parmi les fonctionnalités que l'homme

---

attribue à sa conscience, lesquelles pourront-elles être mises en œuvre dans des futurs robots ? La notion de conscience est un concept si multiforme qu'il n'est pas possible d'en donner une définition précise. L'idée est donc d'imaginer les différentes fonctionnalités liées à cette notion. L'intelligence artificielle ne cherche pas à reproduire un être humain ; il existe d'autres moyens...

Dans le cadre d'un groupe de travail que j'ai animé au sein de l'académie des technologies, nous avons essayé d'identifier dans notre rapport ces différentes fonctionnalités. Sur ma diapositive, j'ai indiqué en vert les différentes fonctionnalités que l'on sait déjà mettre en œuvre ; en orange, celles pour lesquelles on commence à avoir quelques idées, que l'on sait presque faire ou que l'on pourra faire bientôt et, en rouge, celles qui semblent en dehors de nos compétences actuelles. On le voit, un certain nombre de fonctionnalités sont réalisables dès maintenant ou le seront bientôt.

Que faut-il de plus pour avoir une intelligence artificielle forte ?

Il faudrait que ces mécanismes soient capables de prendre du recul, de changer de mode de raisonnement, de trouver des analogies entre des univers différents, de développer des théories, de les vérifier par l'expérimentation et d'avoir envie de faire quelque chose. Qu'est-ce que cette notion d'envie, qui peut être une motivation pour agir ?

Dans un monde ouvert, se posent des problèmes de relations. Dans un monde fermé, on sait préciser les limites de ces techniques. Il faut également tenir compte de la vision stratégique à long terme des « GAFA » – Google, Apple, Facebook et Amazon –, qui semblent pour l'instant prendre le contrôle sur tous ces aspects. Il importe que l'Europe se positionne en la matière.

Alors sur cette intelligence artificielle forte : est-ce de la science-fiction ou n'est-ce qu'une question de calendrier ? Il n'y a jamais eu d'impossibilité prouvée scientifiquement, ni de preuve de possibilité non plus. Toutefois, il faut le souligner, l'intelligence artificielle n'est pas une technique comme les autres au sens où elle débouche sur une certaine autonomie et qu'elle a des possibilités d'apprentissage, d'auto-modifications, etc. Le problème fondamental qui se pose concerne la validation : comment peut-on garantir les propriétés d'un tel système ? Et comment parvenir à conserver le contrôle face aux robots éventuellement dotés de ces capacités ?

**M. Claude de Ganay, rapporteur.** – Je vous remercie pour ce bilan.

Monsieur Yves Demazeau, vous êtes directeur de recherche au CNRS à Grenoble. Mais vous êtes surtout président de l'Association française pour l'intelligence artificielle, et c'est à ce titre que nous vous avons invité. Vous allez pouvoir nous dresser un tableau de la recherche française dans le domaine de l'intelligence artificielle.

## 5. M. Yves Demazeau, président de l'Association française pour l'intelligence artificielle

Monsieur le président, Mesdames, Messieurs les sénateurs, Mesdames, Messieurs les députés, je vous remercie de votre invitation. À vrai dire, je suis d'abord directeur de recherche au CNRS et, pendant mon temps libre, président de l'Association française pour l'intelligence artificielle (AFIA).

Je vous présenterai l'AFIA et vous exposerai les trois objectifs que nous poursuivons depuis 2011, en essayant de dresser un panorama de l'activité de recherche dans le domaine de l'intelligence artificielle en France, du point de vue des chercheurs.

L'intelligence artificielle, discipline née il y a 60 ans, existe donc depuis longtemps. Il s'agit d'une force tranquille qui continue de progresser.

Elle s'intéresse à l'imitation du comportement humain, par l'usage de l'informatique. L'objectif est de créer des systèmes dont le comportement s'apparente à celui de l'être humain.

Les thèmes scientifiques développés à la dernière *International Joint Conference on Artificial Intelligence* (IJCAI) qui s'est tenue à New York en juillet 2016, étaient les systèmes multi-agents, l'intelligence artificielle et le web, l'analyse combinatoire et la recherche heuristique, l'apprentissage automatique, la planification, le contexte humain, et d'autres encore. Lorsque j'ai commencé à travailler sur ces sujets, il y a près de 35 ans, l'intelligence artificielle regroupait tous ces thèmes au sein de cette même conférence IJCAI. L'enjeu était alors surtout de comprendre comment construire une machine intelligente composée de tous ces domaines. Cette question s'étant avérée très complexe, le problème a été décomposé. Et si la lumière est projetée de temps en temps sur tel ou tel domaine, l'enjeu ultime reste l'intégration de ces différentes parties.

La recherche en intelligence artificielle, comme toutes les autres recherches, suit des cycles, du fait des découvertes et des couvertures médiatiques. Ainsi, on a beaucoup parlé à un moment donné d'intelligence artificielle à propos des systèmes experts, on parle beaucoup actuellement d'apprentissage profond. Ce focus actuel est dû à des résultats spectaculaires, qui ne sont possibles maintenant que par la conjonction de la disponibilité de masses de données, et de la capacité de calcul des machines qui ne fait que croître. Si ces progrès n'avaient pas existé, le *buzz* actuel pourrait tout aussi bien concerner plutôt l'autonomie, en particulier en ce qui concerne les véhicules autonomes. Alors, sur quoi le focus sera-t-il mis dans quelques années ?

En France, l'AFIA a identifié dix thèmes de recherche, organisés en à peu près autant de communautés, sinon de conférences :



- 
- apprentissage automatique ;
  - extraction et gestion des connaissances ;
  - interaction avec l'humain ;
  - reconnaissance des formes, vision ;
  - représentation et raisonnement ;
  - robotique et automatique ;
  - satisfaisabilité et contraintes ;
  - sciences de l'ingénierie des connaissances ;
  - systèmes multi-agents et agents autonomes ;
  - traitement automatique des langues.

Au passage, l'apprentissage profond, qui a réveillé l'intelligence artificielle ces dernières années, ne correspond selon nous qu'à 50 % de l'apprentissage automatique, qui est lui-même l'un de ces dix thèmes. Il y a donc aussi tout le reste... Cet éclatement en sous-disciplines, qui se sont forgées chacune une identité ces trente dernières années nous pénalise aujourd'hui, et nous avons donc tout intérêt à les faire interagir entre elles.

Certes, la vision peut paraître un peu académique, mais 95 % des membres de l'AFIA sont issus du monde académique, contre trop peu d'entrepreneurs et d'industriels encore. D'où vient ce décalage ?

Depuis soixante ans, les États-Unis ont bien compris leur intérêt stratégique en matière d'intelligence artificielle. Créateurs de la discipline - au moins du point de vue occidental - ils nous imposent leurs vues et pillent nos forces françaises en intelligence artificielle. C'est très visible dans les conférences internationales comme IJCAI, mais c'est aussi semble-t-il le cas ici. Les industriels invités à la présente audition l'illustrent également : je vois surtout des représentants des GAFAMI.

Suivant les acceptions, l'intelligence artificielle peut représenter jusque 50 % de l'informatique. Ainsi, la moitié des *Lecture Notes in Computer Science* (LNCS) aux éditions *Springer-Verlag* est consacrée à l'intelligence artificielle - (*Lecture Notes in Artificial Intelligence* ou LNAI)-. En France, l'intelligence artificielle n'atteint certes pas ce niveau de reconnaissance ! Au temps de ma thèse de doctorat il y a trente ans, mon directeur de recherche, Jean-Claude Latombe, avait réussi à faire organiser par le ministère de la Recherche d'alors, un programme de recherches concertées (PRC) sur l'intelligence artificielle, doté d'un budget d'un million de francs, ce qui représentait alors 20 % de l'effort de recherche du gouvernement en informatique. Depuis, l'intérêt gouvernemental et des organismes publics pour l'IA n'a fait que baisser, l'IA est souvent bien ignorée, ou plutôt enfouie tellement profondément qu'elle n'est plus visible.

Du point de vue de l'enseignement, on a également trop souvent considéré ces dernières années que les outils et techniques de

---

programmation en intelligence artificielle faisaient partie des acquis en algorithmique. Dans certaines villes, l'intelligence artificielle est tellement banalisée que l'enseignement a disparu des formations universitaires, ce qui est grave et le sera encore plus dans le futur si rien n'est fait.

Il faut dire que régulièrement, on nous demande des mots et des challenges nouveaux. On nous dit qu'en intelligence artificielle, on fait toujours un peu la même chose, mais le challenge de l'intelligence artificielle, à savoir l'imitation du comportement humain, est toujours le même et il est bel et bien là, vivace comme aux premiers temps. Alors, en dépit de la réduction des moyens, la communauté résiste néanmoins au risque d'éclatement.

Les sous-communautés de l'IA se sont créées voilà une trentaine d'années parce que le problème d'imitation de l'intelligence humaine était un peu trop compliqué. Mais il faut réussir à faire interagir ces communautés et à intégrer les résultats de leurs recherches et de leurs avancées. Face au risque d'éclatement de la communauté en IA et au besoin d'assurer des rencontres entre les chercheurs, une association, dont l'objet est de « promouvoir et de favoriser le développement de l'intelligence artificielle en France », a été créée en 1993, à l'occasion d'une conférence IJCAI à Chambéry. En 2016, nous comptons jusqu'à 359 membres, dont encore bien trop peu d'industriels. Nous avons élaboré un plan 2011-2019, ayant pour objectifs de désenfourir l'intelligence artificielle et de réunir les communautés spécialisées et d'interagir avec les autres communautés.

Pour ce qui concerne le fait de désenfourir l'intelligence artificielle : nous tenons chaque année une conférence nationale, et organisons deux journées, dont l'une, en octobre, s'intitule « Perspectives et défis de l'intelligence artificielle (PDIA) ». En 2015, le thème en était « Les apprentissages ». En 2016, les « impacts sociaux de l'intelligence artificielle ». L'autre journée que nous tenons, tous les mois d'avril, est un « Forum industriel de l'intelligence artificielle (FIIA) ». L'année dernière, il s'agissait surtout du lancement du collège industriel de l'AFIA, où les membres ont discuté de l'investissement de l'intelligence artificielle dans leur environnement. Nous publions des bulletins trimestriels, depuis 2011, et avons une [page web](#) et des réseaux sociaux. Les dossiers sont accessibles en libre-service sur le *web*, avec des cartographies (agronomie, innovation, robotique, éthique, mégadonnées, réalité virtuelle, médecine, EIH, jeux vidéo, aide à la décision, etc.). Nous commençons aussi l'établissement d'un fonds de fiches car nous sommes très sollicités sur ce sujet et allons faire des efforts pour améliorer la compréhension de l'intelligence artificielle et de ses enjeux au plus grand nombre.

S'agissant de la réunion des communautés spécialisées, une plateforme intelligence artificielle (PFIA) se tient chaque année en juillet. Ayant identifié une dizaine de communautés spécialisées, qui pourront donner lieu à autant de collèges thématiques, cette plateforme permet de se retrouver. En 2017, la PFIA aura lieu à Caen du 3 au 7 juillet, vous y êtes

---

cordialement invités. Deux collèges thématiques sont actifs à ce jour : l'un sur les « Sciences de l'ingénierie des connaissances » et l'autre sur les « Systèmes multi-agents et les agents autonomes ». Nous espérons lancer trois nouveaux collèges cette année. Par ailleurs, notre collège industriel se charge du forum industriel de l'IA ainsi que de la conférence sur les applications pratiques de l'IA au sein de la PFIA. Depuis deux ans, nous réalisons aussi des compétitions d'intégration au sein de la PFIA, qui constituent des mécanismes incitatifs pour que les collègues des dix collèges interagissent autour d'un objectif commun, et intègrent leurs technologies respectives. Pour l'instant, il s'agit de compétitions académiques. En 2017, la compétition porte sur l'intelligence artificielle dans les jeux interactifs. Mais nous espérons que ces compétitions, à l'avenir, aient par exemple pour objectif de résoudre des problèmes de nature industrielle qui seraient soumis par l'un des membres de notre collège industriel.

Enfin, nous cherchons à interagir avec les autres communautés, *via* des journées communes avec d'autres sociétés savantes et associations, d'organisations patronales (exemple de la journée commune avec le MEDEF le 23 janvier 2017), des groupements de recherche (GdR) du CNRS, mais aussi des pôles de compétitivité. Nous avons planifié une journée avec la société de philosophie des sciences (SPS) le 2 février, une journée avec l'AFIHM sur l'interaction homme-machine le 17 mars, une journée EIAH et IA avec l'ATIEF en juin, une journée « jeux informatisés » avec le pré-GdR AFAIA du CNRS en juillet, et, enfin, une journée éthique avec le COMETS du CNRS et la CNIL puis une journée recherche opérationnelle avec la ROADEF en septembre. Nous organisons également le 27 avril notre forum FIIA et le 6 octobre notre journée PDIA. Enfin, nous sommes consultés régulièrement, par exemple par l'Académie des technologies, par différents instituts comme l'IHEDN, par des municipalités et associations qui s'inquiètent de l'impact de l'intelligence artificielle, et par la presse, qui relaie les progrès en intelligence artificielle, avec un accent sur les États-Unis.

Pour conclure et envisager l'avenir de l'intelligence artificielle, nous pensons que cette dernière restera une force tranquille durant les soixante prochaines années car nous sommes encore loin d'imiter le comportement humain. Ceci dit, nous estimons que les machines sont, d'année en année, plus intelligentes, alors que l'être humain, lui, évolue moins vite. Je suis très surpris de voir que l'on se réfère toujours aux États-Unis, alors qu'il est nécessaire de considérer ce qui se passe en Asie, en particulier en Chine et au Japon.

Aux États Unis, l'industrie automobile investit beaucoup en intelligence artificielle et l'administration de Barack Obama a fait un rapport pour se préparer à l'intelligence artificielle, mais les gouvernements japonais et chinois ne sont pas en reste. C'est en Chine que l'investissement public semble le plus conséquent. En mai 2016, le nouveau plan national a ainsi intégré un programme en IA sur trois ans. J'ai été personnellement fasciné par le lancement du drone Ehang 184, en janvier 2016. Ce drone pèse 200 kg et peut transporter une personne qui indique *via* un écran tactile la

---

destination à laquelle il souhaite se rendre avec une autonomie de 25 minutes. Les Chinois ont beaucoup investi, et ont une vision différente, et à certains égards complémentaire, de celle des Américains. Le Japon a, lui aussi, lancé un nouveau plan d'action en intelligence artificielle sur trois ans. La robotique est dominante dans leurs investissements. Il y a plusieurs raisons à cela parmi lesquelles le fait fascinant que le shintoïsme conduit à croire dans la réincarnation dans des objets. La relation d'un Japonais avec un robot est donc assimilable à celle qu'il peut avoir avec un autre être humain. L'acceptabilité de la robotique au Japon ne sera sans doute jamais atteinte en France. Il faut accepter ces différences et cultiver nos spécificités. Nous avons une vision française de l'intelligence artificielle à valoriser nous aussi.

Au minimum, nous espérons que la France rétablira le poids de l'intelligence artificielle à 20 % au moins de l'informatique, ce qui est bien loin d'être le cas actuellement. La complexité de toutes les thématiques qui la composent est telle qu'un soutien de l'État à la hauteur des ambitions est nécessaire. Il incombe aux responsables politiques de fournir les moyens nécessaires à une intelligence artificielle française digne de ce nom. Il me semble, par ailleurs, essentiel de rétablir l'enseignement de l'intelligence artificielle dans toutes les filières universitaires. 40 % des chercheurs en apprentissage profond avouent ne pas savoir expliquer les raisons pour lesquelles leur technologie donne des résultats. Les machines vont devenir de plus en plus complexes et de moins en moins explicables. Il y a donc là un vrai problème, qui se posera de plus en plus demain.

Il convient aussi de mettre en valeur les réalisations françaises. Environ 500 personnes se sont inscrites à la journée « Entreprises de France et intelligence artificielle » que nous organisons la semaine prochaine avec le MEDEF. Des entreprises françaises y présenteront leurs réalisations en intelligence artificielle. Il faudrait que nos entreprises communiquent aussi bien que le font les GAFAs sur leurs grandes réussites, au lieu de râler sur le coût du travail.

Enfin, il importe d'établir une meilleure coordination entre tous les acteurs de l'intelligence artificielle car si on perd la main au niveau national, il y a clairement un risque de perte de souveraineté. Chaque année, la recherche avance bien, avec différents cycles : on a beaucoup parlé à un moment donné de systèmes experts ; actuellement on parle de *deep learning*.

Les États-Unis ont compris leur intérêt stratégique à l'intelligence artificielle, et ils imposent leurs vues. C'est très visible dans les conférences internationales, en particulier, ainsi que lors des auditions.

Suivant les acceptions, l'intelligence artificielle peut représenter 50 % de l'informatique. La moitié des *Lecture Notes in Computer Science* aux éditions Springer-Verlag est consacrée aux *Lecture Notes in Artificial Intelligence*. En France, on n'en est pas là. Lorsque je faisais ma thèse, mon directeur de recherche, Jean-Claude Latombe, avait réussi à organiser voilà trente ans un PRC-GDR (programme de recherches coordonnées - groupe de

---

recherche), avec un budget d'un million de francs, ce qui représentait 20 % de l'effort de recherche en informatique.

Dans certaines villes, l'intelligence artificielle est tellement banalisée que l'enseignement a disparu des formations universitaires, ce qui est grave. Depuis quelques années, on nous demande des mots nouveaux, des nouveaux challenges. En réalité, le challenge de limitation du comportement humain est encore présent. On nous oppose le constat que l'on fait toujours un peu la même chose et que la technologie est quelque peu enfouie dans le système intelligent. Même si cette situation a éclaté la communauté, celle-ci résiste néanmoins.

Les thèmes de l'intelligence artificielle au niveau international concernent non seulement le *deep learning*, mais également la planification, la modélisation, les connaissances, etc. Face à l'éclatement de la communauté et, simultanément, à la nécessité que les chercheurs en intelligence artificielle se rencontrent, a été créée en 1993, à l'occasion de la tenue de la conférence IJCAI (*International Joint Conference on Artificial Intelligence*) à Chambéry, une association régie par la loi de 1901, dont l'objet est la promotion et le développement de l'intelligence artificielle en France.

Nous avons élaboré un plan 2011-2019, ayant pour objectifs de désenfouir l'intelligence artificielle et de réunir les communautés spécialisées. Des sous-communautés se sont créées voilà une trentaine d'années parce que le problème était un peu trop compliqué. Mais il faudrait réussir à intégrer les résultats des recherches de toutes ces communautés. D'ailleurs, l'intelligence artificielle n'est pas un monobloc. Même si ce domaine fait partie d'une grande part de l'informatique, il doit y avoir une interaction.

En France, l'AFIA a identifié dix thèmes. Certes, la vision peut paraître un peu académique, mais 95 % des 348 membres de l'association sont des académiques, contre très peu d'industriels. L'apprentissage profond est, à nos yeux, une partie de l'apprentissage – moins de 50 %. Nous mettons actuellement un focus particulier sur l'apprentissage profond, mais cela pourrait tout aussi bien concerner le véhicule autonome, dont les recherches accusent un peu de retard. Certes, on peut toujours imaginer les prochains challenges, mais, dans quelques années, la priorité sera donnée à un autre sujet.

Non seulement l'informatique est explosé entre plusieurs dizaines d'associations, mais l'intelligence artificielle déborde sur un certain nombre de domaines. Nous interagissons donc avec d'autres associations.

Pour désenfouir l'intelligence artificielle, nous organisons une conférence nationale, des journées propres, des bulletins trimestriels incluant des dossiers. Nous commençons aussi l'établissement d'un fonds de fiches. Nous sommes en effet très sollicités sur ce sujet ; nous allons donc faire des efforts en termes de communication pour simplifier les choses. Nous avons aussi un portail *web* et des réseaux sociaux. J'indique que les dossiers sont

---

accessibles en libre-service sur le *web*, avec des cartographies. Des équipes travaillent sur des thématiques applicatives données.

Réunir les communautés spécialisées : on est en train de créer autant de collèges thématiques qu'il y a de thèmes, mais on a aussi créé, il y a un an, un collège industriel. À cet égard, nous organisons lundi prochain une journée commune avec le MEDEF.

Par ailleurs, entre quatre cents et quatre cent cinquante personnes sont attendues l'été prochain pour assister à la plateforme Intelligence artificielle. De plus, des journées communes sont organisées avec d'autres associations, des organismes, des pôles de compétitivité, des organisations patronales, des académies, des organisations gouvernementales, les municipalités, des organes de presse écrite et orale.

Une journée commune avec la société de philosophie des sciences est aussi prévue en février. Et bien d'autres rencontres sont programmées, avec d'autres associations, des groupements de recherche du CNRS ou tout organisme qui le demanderait.

Quoi qu'il arrive, on espère que l'intelligence artificielle continuera à être une force tranquille durant les soixante prochaines années. On est très loin d'être arrivé à étudier la limitation du comportement humain. On le constate, les machines sont chaque année de plus en plus intelligentes, alors que l'être humain ne va pas beaucoup évoluer. Il est nécessaire de considérer ce qui se passe en Asie. À cet égard, je suis très surpris de voir que l'on se réfère toujours aux GAFA, notamment. Pour ma part, je regarde ce qui se passe en Chine et au Japon. Vous le savez, le gouvernement Obama a fait un rapport de prospective pour se préparer à l'intelligence artificielle. Mais, en mai 2016, le gouvernement japonais a lancé un grand plan dans ce domaine. La Chine a également engagé en juin dernier un programme national en la matière pour les trois prochaines années.

J'espère que la France rétablira le poids de l'intelligence artificielle à 20 % au moins de l'informatique, ce qui est bien loin d'être le cas actuellement. La complexité de toutes les thématiques est telle que si le soutien de l'État n'est pas à la hauteur des ambitions, il sera alors quasiment sans effet. À vous de voir si vous y mettez les moyens.

Il me semble essentiel de rétablir l'enseignement dans les formations universitaires. Certes, on parle beaucoup actuellement de l'apprentissage profond. Mais sachez que 40 % des chercheurs en la matière ne savent pas expliquer les raisons pour lesquelles cela marche. Les machines vont devenir encore plus intelligentes et si aucune formation n'est dispensée en intelligence artificielle, il arrivera un moment où l'on comprendra encore moins ce qui se passe. Se pose donc là un problème.

Il convient de mettre en valeur les réalisations françaises. Je suis très heureux de voir que cinq cent dix personnes se sont inscrites à la journée que nous organisons la semaine prochaine en commun avec le MEDEF. On montrera aux entreprises françaises les réalisations en intelligence artificielle.

---

Cela me semble important. Il faudrait que celles-ci communiquent aussi bien que le font les GAFAs sur leurs grandes réussites, au lieu de râler.

Ensuite, il importe d'établir une meilleure coordination entre les acteurs de l'intelligence artificielle – on peut toujours le souhaiter. Je vous l'ai dit, si l'on perd la main au niveau national ou au niveau scientifique, il y a clairement un risque de perte de souveraineté. Il faut sans doute considérer l'intelligence artificielle comme une technologie de souveraineté.

**M. Claude de Ganay, rapporteur.** – Monsieur Bertrand Braunschweig, vous êtes directeur du centre de l'Institut national de recherche en informatique et en automatique (INRIA) de Saclay et représentez le président-directeur général de l'INRIA, Antoine Petit, aujourd'hui retenu. L'INRIA a publié l'été dernier un Livre blanc sur l'intelligence artificielle, qui fait le point sur l'état de la recherche à ce sujet au sein de votre institut.

#### **6. M. Bertrand Braunschweig, directeur du centre de l'INRIA de Saclay**

J'ai effectivement coordonné la réalisation de ce premier Livre blanc sur l'intelligence artificielle, qui présente les équipes de l'INRIA actives dans les différents domaines de l'intelligence artificielle. L'Institut est organisé en équipes-projet, au nombre de deux cents environ, sur nos huit centres de recherche, et qui regroupent chacune une vingtaine de chercheurs en moyenne. Un tiers de ces équipes touche, de près ou de loin, au domaine de l'intelligence artificielle, une proportion qui a significativement augmenté ces dernières années, avec la montée en puissance de l'apprentissage-machine. Plus d'un millier de communications en conférence et plus de quatre cents articles dans les revues consacrées à l'intelligence artificielle ont été publiés par nos chercheurs au cours des dix dernières années.

Ce livre blanc mentionne quelques faits marquants qui motivent l'intérêt actuel pour l'intelligence artificielle, et rappelle aussi qu'il y a des questions de société à traiter. L'INRIA a pris des initiatives en ce sens depuis plusieurs années, en créant, au sein de l'alliance Allistene, la commission de réflexion sur l'éthique de la recherche en sciences et technologies du numérique (CERNA), dont le premier rapport public traite de la recherche en robotique, ainsi qu'en se dotant d'un organe interne, le comité opérationnel d'évaluation des risques légaux et éthiques, qui traite les cas pratiques soumis par nos équipes de recherche. En cette année où nous célébrons les cinquante ans de l'INRIA, nous entendons contribuer au débat public qui doit exister sur l'intelligence artificielle.

L'année 2017 est aussi celle de la préparation de notre plan stratégique scientifique pour la période 2018-2022. Ce plan sera organisé

---

autour de défis scientifiques motivants, dont plusieurs porteront, soyez-en certains, sur l'intelligence artificielle. Nous finissons, en ce moment, un nouveau Livre blanc sur le véhicule autonome et connecté : un des défis du plan devrait porter sur ce sujet.

Le Livre blanc est structuré en huit sous-domaines, sur lesquels nos équipes sont actives et qui couvrent une bonne partie de la recherche en intelligence artificielle. On y trouve, entre autres entrées, les défis génériques de l'intelligence artificielle, l'apprentissage automatique, l'analyse des signaux - vision et parole, notamment -, les connaissances et le *web* sémantique, la robotique et les véhicules autonomes, les neurosciences et la cognitive, le traitement du langage, la programmation par contrainte pour l'aide à la décision.

Je m'en tiendrai à formuler quelques remarques sur certains secteurs qui me paraissent particulièrement importants. En matière d'apprentissage automatique, la révolution de l'intelligence artificielle est essentiellement venue de progrès scientifiques considérables en matière de traitement du langage et de la parole, de vision, de robotique et, surtout, d'apprentissage automatique. En cette dernière matière, même si des résultats remarquables ont été obtenus, il reste encore de nombreux défis propres à motiver longtemps les chercheurs.

Je pense, en particulier, à l'apprentissage non supervisé, sans intervention d'un oracle humain, à l'apprentissage sous contrainte, par exemple pour la gestion de la vie privée, à l'apprentissage de causalité, indispensable pour construire des systèmes prédictifs, ou encore à l'apprentissage continu et sans fin, pour des systèmes destinés à opérer vingt-quatre heures sur vingt-quatre, sept jours sur sept.

Autre défi, l'interaction avec les humains.

Les systèmes d'intelligence artificielle étant appelés à interagir avec des utilisateurs humains, ils doivent être capables d'expliquer leur comportement, de justifier les décisions qu'ils prennent, faute de quoi, ces systèmes ne seront pas acceptés, par manque de confiance. J'ajoute que les systèmes d'intelligence artificielle ont besoin d'une certaine flexibilité et doivent être capables de s'adapter à différents utilisateurs et différentes attentes. Il importe donc de développer des mécanismes d'interaction favorisant une bonne communication entre humains et systèmes d'intelligence artificielle.

J'en viens au défi de l'ouverture à d'autres disciplines. Sachant que l'intelligence artificielle sera souvent intégrée dans un système comportant d'autres éléments, les spécialistes du domaine devront collaborer avec ceux d'autres sciences de l'informatique - modélisation, vérification, validation, visualisation, interaction machine -, ainsi que d'autres disciplines comme la psychologie, la biologie, les mathématiques, mais aussi les sciences humaines et sociales - économie, ergonomie, droit. Le mieux serait que les chercheurs possèdent une double, voire une triple compétence, afin de faire sauter les barrières disciplinaires, qui rendent malheureusement difficile la réalisation



---

de systèmes complets. J'ai bon espoir que les instituts Convergence, financés par les investissements d'avenir, y contribuent. En attendant, il est indispensable de composer des équipes pluridisciplinaires pour aborder les problèmes sous tous les angles.

Composante incontournable des systèmes critiques, la certification des systèmes d'intelligence artificielle ou leur validation par des moyens appropriés constitue également un véritable défi. Alors que la vérification, la certification et la validation des systèmes classiques sont déjà des tâches difficiles, même s'il existe des outils exploitables, l'application de ces outils aux systèmes d'intelligence artificielle comportant notamment des systèmes d'apprentissage automatique est une tâche encore plus ardue, à laquelle il convient de s'attaquer si nous voulons utiliser ces systèmes dans des environnements tels que ceux de l'avion, des centrales nucléaires, des hôpitaux...

J'en arrive à l'interaction entre l'apprentissage et la modélisation, sujet qui nous tient particulièrement à cœur. Alors que les scientifiques passent beaucoup de temps à concevoir des modèles - physiques, mathématiques, symboliques -, les algorithmes d'apprentissage ignorent encore assez largement ces modèles, et travaillent sur des données brutes. Il faut, à notre sens, briser cette barrière entre apprentissage et modélisation, pour établir une interaction dans les deux sens, afin d'améliorer les performances de l'apprentissage automatique grâce aux connaissances synthétisées dans des modèles, ce qui permettra, en retour, d'améliorer ou de spécialiser ces modèles, grâce aux résultats de l'apprentissage. Cela peut requérir une interface entre les couches numérique et symbolique : la tâche n'est pas facile, mais c'est un peu le Graal dont la poursuite doit nous occuper.

Parmi les autres défis importants, sur lesquels je pourrais revenir au cours du débat, il faut également compter la gestion des données multimodales provenant de différents capteurs, la compréhension de scènes et d'environnements, en particulier pour les robots et les véhicules autonomes, et, toujours pour les robots, la prise en compte de l'incertitude et des données incomplètes ou disponibles seulement à une certaine fréquence. Je pense également, pour le domaine du *web* sémantique, à la connexion entre les ontologies, qui en constituent le moteur, et les données stockées dans les bases, qui en sont le carburant.

Le plus grand défi, pour conclure, est non pas technique, mais stratégique. La France excelle sur tous ces sujets en raison de la qualité de ses scientifiques. Tant pour les approches mathématiques - statistique, modélisation, optimisation, apprentissage - que pour la modélisation et la représentation des connaissances, nous bénéficions d'une longue tradition de raisonnement logique et de cartésianisme. Pour preuve, les chercheurs français suscitent de plus en plus la convoitise des entreprises internationales, qui offrent des niveaux de salaire et de financement extrêmement motivants. Mettre nos capacités au service du développement

---

économique grâce au transfert technologique doit permettre à nos entreprises de prendre des parts d'un marché international en forte croissance. Nos entreprises, grandes et petites, ont besoin de ces technologies ; nos chercheurs en intelligence artificielle sont parmi les meilleurs au monde : il faut tisser plus de liens entre ces deux mondes. Cela pourrait être la mission principale d'un grand plan national en intelligence artificielle. Demain sera d'ailleurs lancée la préparation de la stratégie nationale en intelligence artificielle, sous le patronage des deux ministères concernés, celui de l'industrie et celui de la recherche. L'INRIA estime que notre pays doit se montrer ambitieux et faire de l'intelligence artificielle une vraie priorité. Au vu des investissements que consentent plusieurs grands pays, c'est au moins un milliard d'euros sur dix ans qu'il faut mobiliser, aides publiques et contribution des acteurs économiques confondues ; je pense aux entreprises, mais aussi aux collectivités territoriales, qui pourraient fournir des sites d'expérimentation grandeur réelle. Tel est le principal défi que l'INRIA vous propose de relever.

**M. Claude de Ganay, rapporteur.** – Monsieur David Sadek, vous êtes directeur de la recherche de l'Institut Mines-Télécom. Enseignant-chercheur et spécialiste en intelligence artificielle, vous avez mené une carrière atypique, puisque vous êtes d'abord entré chez Orange, dont vous êtes devenu, il y a dix ans, le directeur délégué à la recherche avant de revenir à la recherche publique.

### **7. M. David Sadek, directeur de la recherche à l'Institut Mines-Télécom**

Permettez-moi d'abord d'évoquer les publications de l'Institut Mines-Télécom, notamment son cahier de veille sur l'intelligence artificielle, qui offre une sorte de patchwork de points de vue sur le sujet. Créé conjointement avec la Fondation Télécom, ce cahier témoigne de la diversité de nos activités, qui touchent au traitement des données massives – en la matière, plusieurs chaires industrielles sont portées par l'Institut –, à l'apprentissage, à la reconnaissance des formes, à la compréhension de scènes, à la représentation des connaissances, aux neurosciences informationnelles – domaine nouveau qui vise à caractériser l'information mentale et corticale –, à la robotique, aux réflexions éthiques, enfin, sur l'intelligence artificielle.

Si je centre mon intervention sur la question de l'apprentissage, sur laquelle les orateurs m'ayant précédé ont également mis l'accent, je n'en estime pas moins qu'il faille se garder de réduire l'intelligence artificielle à un seul de ses pans.

L'intelligence artificielle est un vaste domaine, pluridisciplinaire, dont l'unité tient en ceci que l'on y cherche à faire pratiquer à des machines des compétences cognitives qui sont le propre de l'humain. Comme je le dis

---

souvent, l'intelligence artificielle, c'est la didactique des machines : il s'agit de parvenir à expliquer à une machine comment faire des choses que l'homme fait très bien. Cela pose un ensemble de défis, au nombre de six ou sept.

Bertrand Braunschweig a évoqué des défis génériques ; je me situerai à un échelon plus primitif. Gérard Sabah a parlé de la représentation du sens : ce n'est pas parce qu'une machine parvient à classer une image de chat dans un groupe d'images similaires qu'elle a compris ce qu'était un chat. Certes, on avance dans le traitement sémantique de l'information, et d'autres approches essaient de tendre vers le sens au moyen de ce que l'on appelle des ontologies, mais il n'en demeure pas moins que la représentation du sens reste encore un défi pour l'intelligence artificielle.

Autre défi, la notion de sens commun.

L'être humain se caractérise par sa capacité à raisonner sur la base d'informations approximatives, en menant des raisonnements symboliques du type « en général » : « En général il fait beau en Bretagne », assertion au reste discutable... Il use de raisonnements par défaut, mais aussi de raisonnements par analogie : « Paris est à la France ce que Washington est aux États-Unis », par exemple. Mais, dans le domaine de l'intelligence artificielle, de tels processus ne sont pas encore assimilés, y compris par les approches que l'on met aujourd'hui en avant.

Autre gageure, évoquée par Jean-Gabriel Ganascia, le sujet des agents conversationnels : l'interaction, le dialogue avec la machine. C'est en soi un concentré de défis. Il n'est d'ailleurs pas anodin que le fameux test de Turing fasse appel à du dialogue en machine. Sont en jeu des compétences qui supposent une compréhension du langage naturel, une interprétation en contexte, qui fait sortir du simple jeu des questions-réponses etc. N'allez pas croire qu'un logiciel comme *Siri* a résolu la question ; on en est loin. Avoir avec une machine un dialogue évolué reste un défi pour l'intelligence artificielle, quelle qu'en soit l'approche. Outre les compétences que je viens de citer, il faut aussi y ajouter une capacité de production de raisonnement, de langage, d'introspection. Gérard Sabah a rappelé que l'on est loin encore de comprendre ce qu'est la génération de conscience – elle reste d'ailleurs, dans la sphère de l'humain, indécidable pour tout autre que soi –, ajoutant que ce qui doit nous importer, c'est bien plutôt le rôle causal des fonctions attribuées à une machine.

Sans aller donc jusqu'à traiter la question de la conscience, celle de l'introspection, c'est-à-dire la capacité, pour une machine, à raisonner et à réfléchir sur ses propres connaissances et ses propres comportements, reste encore un défi pour l'intelligence artificielle.

Vient ensuite ce que l'on appelle l'intelligence émotionnelle, à savoir la capacité à caractériser et à reconnaître des émotions.

À quoi il convient encore d'ajouter la capacité d'expliquer les comportements, de rendre compte. C'est là un point très important, ainsi que

---

l'a souligné Gérard Sabah, car un système d'intelligence artificielle, pour être acceptable, doit être capable d'expliquer ce qu'il est en train de faire et pourquoi. Une machine, autrement dit, doit être au fait de ce qui régit son fonctionnement.

Autre point important, ce que l'on appelle la « prouvabilité » des systèmes, soit la capacité à montrer qu'un système intelligent fait ce que l'on attend de lui, et seulement ce que l'on en attend. Alors que l'on se soucie beaucoup, aujourd'hui, des règles éthiques, il serait malvenu que des systèmes intelligents soient utilisés à des fins autres que celles pour lesquelles ils ont été initialement conçus. Or des comportements contraires à l'éthique peuvent émerger de manière imprévue dans un système dont les propriétés n'auraient pas été prouvées, au sens que je viens de définir, *ex ante*. Je cite souvent l'exemple d'un système de bataille navale utilisé, du temps que j'étais jeune chercheur, par le département de la Défense américain, pour entraîner les pilotes. On s'est aperçu que, si le système gagnait tout le temps, c'est parce qu'il détruisait les bateaux touchés de sa flotte, pour qu'ils ne la ralentissent pas ! On avait pourtant simplement indiqué au système qu'il fallait maximiser la vitesse de la flotte, et il a fait de lui-même cette inférence simple qu'un bateau touché la ralentissait. Voilà ce qui peut advenir lorsque la validité d'un système intelligent n'a pas été prouvée. La recherche de l'objectif de maximisation peut conduire à des actions non éthiques.

J'insiste, pour conclure, sur l'hybridation des approches, de nature à faire avancer la recherche. Les approches, qu'elles soient neuro-inspirées, connexionnistes, stochastiques ou, au contraire, cognitivistes ou symbolistes ont toutes à apporter dans la compréhension, la modélisation et la mise en œuvre des comportements intelligents. Nous avons des équipes excellentes, en France, dans deux ou trois de ces pans de l'intelligence artificielle, et nous devrions nous investir pour que ces équipes travaillent ensemble. Nous sommes à l'aube d'une ère nouvelle. Cela exige de faire la part des choses, de trier entre le bon grain et l'ivraie. Les approches par l'apprentissage, notamment le *deep learning*, sont très performantes pour tout ce qui relève de la perception – reconnaissance vocale, traitement des signaux pour la reconnaissance des visages –, tandis que ce qui a trait à la mobilisation des comportements intelligents, à la conduite de dialogues en machine, et qui passe par les notions d'attitude mentale et de raisonnement symbolique, reste largement à défricher : la combinaison de l'approche cognitive et des approches faisant appel à la classification serait, en cette matière, fructueuse.

**M. Claude de Ganay, rapporteur.** – Monsieur Jean-Daniel Kant, vous êtes maître de conférences à l'université Pierre-et-Marie-Curie Paris-VI. Spécialiste de la technologie, peu connue, des systèmes multi-agents, vous vous intéressez à l'hybridation entre technologies d'intelligence artificielle.

---

### 8. M. Jean-Daniel Kant, maître de conférences à l'université Pierre-et-Marie-Curie Paris-VI

Je vous remercie de votre invitation et de l'initiative de l'OPECST. C'est une excellente idée de nouer le dialogue entre chercheurs et parlementaires, et j'espère que cette occasion qui nous est donnée de nous rencontrer sera suivie de bien d'autres.

Je vais m'efforcer de vous exposer ce que sont les systèmes multi-agents. Le *deep learning*, le *machine learning*, les robots sont la face visible de l'intelligence artificielle, qui comporte cependant d'autres aspects, évoqués par David Sadek et Bertrand Braunschweig. L'intelligence artificielle, de fait, se décline au pluriel. L'une de ses approches se situe du côté des données, tandis que l'autre se tourne plutôt vers les comportements. La première approche vise, grâce au traitement d'un très grand nombre de données, à faire émerger des comportements liés à l'intelligence humaine. Elle est beaucoup utilisée pour la reconnaissance de formes, mais aussi en aide à la décision. Les banques commencent ainsi à utiliser de tels systèmes pour décider à quels clients accorder un prêt. Les assurances s'y lancent également, pour détecter les fraudes.

Les réseaux de neurones - l'appellation est métaphorique - sont très puissants, mais ce sont des boîtes noires. Les spécialistes, comme Yann LeCun, disent, honnêtement, qu'ils ne savent pas vraiment comment cela fonctionne. Il y a là une question politique, une question citoyenne. Qui voit son banquier rejeter sa demande de prêt, s'il s'entend répondre, quand il demande les raisons du refus, que c'est parce que le neurone n° 39 et la connexion 48 étaient inférieurs à 0,4, est en droit de s'offusquer, et pourrait bien se pourvoir en justice.

Vous le savez, l'Union européenne a été réactive en ce domaine, puisqu'un règlement et une directive adoptés récemment concernent l'explication des décisions automatisées. Bientôt, tous les systèmes devront être capables de justifier leur comportement. Cela sera sans nul doute un frein à l'utilisation, dans l'aide à la décision, des systèmes à réseaux de neurones.

L'autre approche, qui n'est pas gouvernée par les données, passe par la modélisation des comportements humains. Elle s'efforce, dans une démarche pluridisciplinaire, faisant appel à des psychologues, des économistes, des sociologues, de comprendre les mécanismes qui régissent les comportements humains, pour tenter de les modéliser dans des programmes informatiques.

Un système multi-agents est fait de programmes autonomes, dotés de certaines capacités cognitives, qui peuvent être très simples ou très complexes. Ces programmes, que l'on appelle « agents », communiquent entre eux, interagissent, pour former une sorte d'intelligence collective. Cela est très utile lorsque l'on veut modéliser des systèmes humains, sociétés ou économies.

---

L'autre question qui vaut d'être posée est celle des objectifs de l'intelligence artificielle. J'en vois deux. L'un est de faire des machines intelligentes, autonomes, destinées, même si on ne le présente pas ainsi, à remplacer les êtres humains. Quand on crée une voiture autonome, plus besoin de chauffeur. Dans tous les cas où l'on considère que l'être humain sera moins performant que la machine, ou dans les cas où l'on ne peut employer qu'un robot – pour une expédition vers Mars, par exemple –, cela fait sens. Mais la société, le monde politique doivent s'interroger : est-ce le cas dans tous les domaines ? D'autant que remplacer les humains par des machines n'est pas sans incidence sur le marché de l'emploi, c'est mettre des gens au chômage, et c'est bien pourquoi Barack Obama évoquait l'idée de revenu universel.

Le second objectif de l'intelligence artificielle est non pas de remplacer l'homme, mais de lui apporter une assistance. Quand on dialogue avec *Siri*, on obtient de l'aide. En viendra-t-on un jour à ne plus dialoguer qu'avec des machines ? C'est une autre histoire.

Quoi qu'il en soit, mes recherches sont plutôt orientées vers cet objectif d'aide à la décision. Quelles en sont les applications potentielles ? Elles peuvent concerner, par exemple, l'économie. C'est ainsi qu'avec l'économiste Gérard Ballot nous avons développé un simulateur multi-agents du marché du travail français. Grâce à une modélisation à 1/2 000<sup>e</sup> de ce marché, nous sommes en mesure d'évaluer des politiques publiques. Nous avons, par exemple, évalué la loi El Khomri, et les résultats, assez surprenants, font apparaître que les choses sont beaucoup plus compliquées que ce qui se disait alors. Cette évaluation a donné lieu à bien des réactions sur le *web*, les gens se demandant pourquoi les ministères n'utilisaient pas de tels systèmes pour faire leurs études d'impact. Certes, être totalement prédictif sur une matière aussi complexe que le marché du travail n'est pas facile, mais il reste que les systèmes multi-agents sont une voie.

Nous avons également créé des applications dans le domaine social pour étudier la diffusion de l'innovation dans une population de consommateurs, ou bien la dynamique des opinions et des attitudes – nous réfléchissons actuellement à un projet destiné à étudier la radicalisation sur internet.

Créer des systèmes fonctionnant comme un miroir de l'activité humaine est un moyen, à mon sens, pertinent pour rendre compte de leur complexité. On peut également hybrider, et étudier les interactions entre l'humain et ces technologies.

C'est une voie prometteuse, mais qui demande des moyens. En France, hélas, nous en manquons. Ce serait pourtant le moyen de garantir notre souveraineté, face à des pays comme les États-Unis ou la Chine. J'insiste aussi sur l'utilité de la recherche pluridisciplinaire, qui n'est pas assez défendue.

---

**M. Claude de Ganay, rapporteur.** – Monsieur Benoît Le Blanc, vous êtes directeur adjoint de l'École nationale supérieure de cognitique. Vous êtes spécialiste en intelligence artificielle et assumez actuellement une mission pour le ministère de l'enseignement supérieur et de la recherche, dont vous nous direz sans doute quelques mots.

### 9. M. Benoît Le Blanc, directeur adjoint de l'École nationale supérieure de cognitique

Beaucoup a déjà été dit sur les technologies d'intelligence artificielle, un sujet dont on débat dans les médias. La maturité de ces technologies, issues de laboratoires de recherche, est-elle suffisante pour que des entreprises engagent des moyens destinés à développer massivement des produits qui vont bouleverser notre quotidien ? Telle est la question.

À titre personnel, je dois dire que je suis impressionné par le programme Watson et les capacités de développement informatique que met à disposition IBM, au travers de son logiciel *Bluemix*. Un tel logiciel permet de développer des applications de « *cognitive computing* », pour, par exemple, transformer le langage en texte, ou inversement.

À lire rapidement les journaux et sites *web* d'actualité, on pourrait considérer qu'il existe, dans le paysage de l'intelligence artificielle, deux groupes d'acteurs, l'un constitué par les chercheurs « à l'ancienne » que nous sommes, l'autre par de jeunes découvreurs de la *French Tech*, qui lancent des *start-ups* sur tous les sujets, en y injectant de l'intelligence artificielle.

Or les technologies d'intelligence artificielle sont pour certaines destinées à des usages par le public et, dès lors qu'intervient le facteur humain, les choses sont beaucoup plus complexes qu'il n'y paraît. De fait, les programmes d'intelligence artificielle sont destinés à converser avec les personnes. Nous avons, en France, beaucoup de talents en sciences humaines et sociales, des laboratoires très performants, peuplés de chercheurs très technophiles, qui peuvent nous apporter beaucoup. Cela suppose, et j'insiste, d'être capable de s'affranchir des barrières disciplinaires qui nous enferment dans certains modes de pensée. Nous sommes passés, entre la fin du XX<sup>e</sup> siècle et l'aube du XXI<sup>e</sup> siècle, des sciences de l'information aux sciences de la connaissance, lesquelles supposent d'être capable de donner un contexte et du sens à l'information.

La cognitique est ainsi à l'informatique ce que la connaissance est à l'information. Il s'agit dans la cognitique d'adapter la technologie aux capacités, aux limites, aux préférences humaines, pour en rendre l'usage plus simple. L'École nationale supérieure de cognitique, au sein de Bordeaux INP, développe un programme d'enseignement en ingénierie lié à notre programme de recherche sur la cognition et aux entreprises passionnées par le sujet et hébergées sur notre campus. Nous travaillons sur l'expérience utilisateur, soit la manière dont les gens interagissent avec les technologies et

---

au-delà, nous développons des programmes du futur autour du partage des connaissances et de l'hybridation humain-machine au service de la cognition. Autrement dit, il s'agit de réfléchir, d'une part, à la manière dont les gens vont partager les connaissances entre eux, et, d'autre part, vont intégrer à leur personne même une composante technique.

Si j'observe la manière dont nous avons mené chacun préparé et mené nos exposés ce matin, je constate que certains utilisent un diaporama pour appuyer leur discours, quand d'autres préfèrent s'en remettre à des notes manuscrites, imprimées ou bien lues sur l'écran de leur ordinateur. Nous sommes tous informaticiens et, pourtant, nous restons, mentalement, pluriels. Nos approches de la technologie sont diverses. De la même manière, c'est en mettant au creuset de l'interdisciplinarité sciences de l'information et de la cognition et sciences sociales et humaines que l'on créera des synergies.

Par exemple, nous butons, à l'heure actuelle, sur la capacité à relier réseaux sémantiques et réseaux neuronaux. Cela signifie que les modèles mentaux ancrés chez les chercheurs sont profondément différents. Certains sont « symbolistes » – j'en fais partie, comme plusieurs autour de cette table – et pensent que l'on crée de l'intelligent en manipulant des symboles, d'autres sont « émergentistes », et travaillent sur les réseaux de neurones en pensant que l'on crée de l'intelligent par association de petites unités en configurations : il faut pouvoir travailler ensemble, même s'il est difficile de comprendre les présupposés et les postures intellectuelles de chacun.

L'intelligence artificielle et la cognition sont des domaines de recherche transdisciplinaires, au même titre que la santé publique, les études sur le climat ou l'alimentation ; tous ces domaines appellent des expertises larges. Mais le problème est que l'on appréhende la transdisciplinarité par le biais de chercheurs attachés à une discipline. Il n'existe pas d'instance qui permette d'évaluer sereinement et positivement les travaux que mènent des chercheurs quand ils changent de section, de bain disciplinaire, si bien que le faire représente un sacrifice de carrière. Il faut parvenir à lever ce frein.

La transdisciplinarité doit nous guider vers des recherches axées sur la performance combinée entre le cerveau humain et le calcul de la machine. Plus que d'une acceptabilité des machines, je parlerai d'une appropriation, qui se fera lorsque les gens tireront bénéfice de l'usage de leur propre machine.

J'ai eu le privilège de discuter avec deux personnalités qui m'ont marqué : Anatoli Karpov, il y a quelques années, lors de l'une de ses conférences, où il expliquait très clairement que, dès lors que les machines battent les humains aux échecs, l'avenir du jeu d'échecs passe par la confrontation de joueurs utilisant des machines, et, plus récemment, Fan Hui, champion d'Europe de jeu de go qui a alimenté la machine par laquelle le champion du monde, Lee Sedol, a été battu. Fan Hui, appelé à tester cette machine, dit s'être retrouvé face à elle comme face à un mur. La machine l'a battu, et il a, raconte-t-il, éprouvé une terrible honte. Cela a été pour lui



---

d'autant plus difficile à vivre qu'il a dû garder le secret jusqu'au combat contre Lee Sedol, mais il conserve, dit-il, la fierté d'avoir été le premier champion de go battu par une machine. Cette compétition lui a permis de mesurer combien il est difficile de garder confiance face à une machine qui joue sans subir aucun doute, sans aucune psychologie. Cela m'a éclairé sur la capacité des machines à aider les humains à comprendre leur situation. Pour Fan Hui, AlphaGo a tout cassé, ce qui lui permet de réfléchir maintenant à la conduite du jeu, de façon beaucoup plus libre.

Assiste-t-on à un véritable progrès ? Est-ce que la nouvelle vague annoncée de l'intelligence artificielle va submerger notre intelligence ? Je ne le crois pas. Pour cela il faudrait que l'on travaille sur des algorithmes de reconnaissance du fond plutôt que de continuer à faire des algorithmes de reconnaissance des formes, mais cela suppose de se pencher sur le véritable lien qui relie syntaxe et sémantique : c'est le problème auquel on se confronte, depuis des années, tant pour le langage que pour l'intelligence artificielle.

**M. Claude de Ganay, rapporteur.** – Merci de ces exposés passionnants. Place, à présent, au débat interactif.

## 10. Débat

Mme Laurence Devillers, professeure à l'université Paris-Sorbonne et chercheuse au Laboratoire d'informatique pour la mécanique et les sciences de l'ingénieur (Limsi) du CNRS. – Il est en effet passionnant d'être confronté à tous ces défis : ils sont bien souvent les mêmes au CNRS qu'à l'INRIA ou à l'Institut Mines-Télécom. Je vous appelle cependant, en conscience, à oublier un peu les instituts auxquels nous appartenons chacun, pour essayer de réfléchir à ce que nous pourrions faire ensemble. Si l'on veut entrer en concurrence avec les GAFAs ou leur équivalent chinois, les BATX - Baidu, Alibaba, Tencent et Xiaomi -, il faut construire ensemble, autour, par exemple, d'un projet sur l'éthique et les risques numériques, on a besoin d'un institut commun de réflexion à ce sujet. En Grande-Bretagne, on a vu de nombreux projets émerger de cette façon. En France, trop de chapelles différentes coexistent, quand nous aurions besoin de bâtir une vision unifiée des efforts à mettre en œuvre. Qu'en pensez-vous ?

M. Jean Ponce, directeur du département d'informatique de l'École normale supérieure. – Beaucoup ont opposé les méthodes d'intelligence artificielle symboliques et les méthodes de type *deep learning*, attachées aux données, pour souligner le besoin d'hybridation entre l'une et l'autre. N'oublions pas, cependant, que cette hybridation existe déjà dans de nombreux domaines. Je pense, par exemple, au traitement du langage et de la parole, à la robotique, à la vision artificielle – mon domaine – où existe un effort de modélisation important : ce ne sont pas des sous-domaines de l'apprentissage statistique, qui, lui-même, ne se résume pas à l'apprentissage

---

profond. Malheureusement, les outils classiques de l'intelligence artificielle n'ont pas eu l'impact qu'ont aujourd'hui d'autres approches.

Il a été question des besoins de financement. Le problème auquel on fait face tient au fait que nous formons d'excellents jeunes chercheurs que les grandes compagnies américaines se disputent, au point qu'attirer et retenir ces chercheurs dans nos laboratoires est extrêmement difficile, comme l'a souligné Bertrand Braunschweig, du fait des différences de salaire. Nous avons à l'École normale supérieure l'exemple des étudiants de nos deux masters spécialisés. Pour remédier à ces difficultés, il faut, à mon avis, rechercher des partenariats avec l'industrie.

M. Basile Starynkevitch, Institut List, Commissariat à l'énergie atomique (CEA). – Je suis un *geek*, j'aime coder. Je viens de terminer de rédiger une proposition qui m'a demandé trois mois de travail pour un appel qui se clôturait hier. J'ai une chance sur cent d'être retenu.

L'intelligence artificielle a encore besoin d'infrastructures logicielles. Le logiciel libre est une solution ; je suis ainsi membre de l'association April. Mais je ne suis pas un spécialiste en la matière, bien d'autres en parleraient mieux que moi.

Les projets de R&D en intelligence artificielle, parce que l'espérance de vie des *start-up* est de trente ou trente-cinq mois, sont le plus souvent limités à trente-six mois. Il serait souhaitable de prévoir des durées plus longues : s'il faut neuf mois à une femme pour faire un bébé, il ne faut pas un mois à neuf femmes pour en faire un. C'est la même chose pour les chercheurs !

M. Claude de Ganay, rapporteur. – Dans quelle direction les efforts de recherche en intelligence artificielle doivent-ils porter en priorité ?

M. Jean-Gabriel Ganascia. – Les financements, en France, vont à des projets de très court terme. Si les chercheurs rejoignent de grands groupes privés, c'est certes parce que les salaires y sont supérieurs, mais aussi parce que les conditions sont désastreuses dans la recherche publique : ils ont envie d'aller au bout de leurs projets, et, dans le public, cela n'est que rarement possible.

La France aurait tendance à copier les modes de financement des États-Unis ? Non, car les durées sont plus longues outre-Atlantique. Aujourd'hui, il n'y a pas de mémoire dans les institutions de financement de la recherche française...

M. David Sadek. – Un institut d'éthique de l'intelligence artificielle est une idée à explorer. Beaucoup de questions se posent, et des règles collectives seraient utiles. Nous avons besoin d'une approche inter-instituts comme le disait Laurence Devillers.

Sur l'hybridation des approches, très souhaitable également, je veux signaler qu'il y a eu, comme Jean Ponce l'a rappelé, de nombreuses tentatives, depuis longtemps, par exemple sur le traitement de la parole. On

---

la pratique sur certains domaines. En revanche, sur les comportements plus évolués, l'interaction des machines, le dialogue, il reste beaucoup à explorer...

M. Yves Demazeau. – Une dizaine de directions de recherche ont été identifiées. Il y a trente ans, l'intelligence artificielle était une matière intégrée, autour du rêve de machine intelligente, puis on s'est rendu compte de la complexité des choses, et l'intelligence artificielle a explosé en diverses communautés spécialisées. On ne compte pas moins d'une dizaine de conférences spécialisées en intelligence artificielle. Le défi à présent est de développer une science de l'intégration. Un drone en Chine transporte un passager de cent kilos ; un véhicule autonome se déplace dans les rues de villes américaines. Mais ni l'un ni l'autre ne rassemble les meilleures technologies des sous-disciplines. Il reste à les réunir !

**Mme Dominique Gillot, rapporteure.** – Je perçois dans vos propos, pourtant, le souci d'une démarche partagée, l'aspiration à l'interdisciplinarité. C'est un enjeu dont vous avez conscience.

M. Yves Demazeau. – Entre chercheurs, oui, mais pas entre organismes de recherche, ni entre institutions ou acteurs du secteur.

M. Jean-Daniel Kant. – L'offre, dans l'enseignement, est peut-être insuffisante, mais je signale que mon université propose deux masters en intelligence artificielle, l'un orienté sur les données, l'autre sur les comportements. Ils comptent au plus quatre-vingts étudiants, mais ils existent !

L'hybridation doit-elle être une priorité ? Pas nécessairement, car elle ne sera pas facile à faire, en tout cas sur les fonctions de haut niveau où ce sera très difficile. C'est facile sur des fonctions de bas niveau. On a essayé il y a une vingtaine d'années de plaquer des symboles dans les neurones, sans résultat. On peut aussi faire l'inverse... Il faut réfléchir avec les disciplines concernées et reprendre la question, sur l'apprentissage par exemple.

Mme Laurence Devillers. – L'hybridation existe déjà. Je vois le cas des dialogues hommes-machines. Si l'on veut être compétitif, il faut éviter de se poser éternellement des questions théoriques primaires.

M. Igor Carron. – Je suis le P-DG d'une petite *start-up* de *hardware* pour le *machine learning*. Je fais également partie d'un groupe qui organise les *Paris Learning Machine Meetups*, avec à ce jour soixante *meetups*, durant lesquels nous avons accueilli deux cents intervenants, autour d'un réseau de plus de 4 700 personnes. Un autre monde existe en matière d'intelligence artificielle : des personnes sorties d'école depuis un, deux ou trente ans se réunissent, non dans le cadre universitaire, mais au contact direct des recherches en cours, pour comprendre ce qui se passe.

Le besoin de l'État n'est pas de même nature en France qu'aux États-Unis : les discussions de ce matin sont très intéressantes, car l'aspect *push* est important ; on a envie de partager, d'échanger, d'envisager des projets. Mais

---

l'aspect *pull* fait défaut en France. En Amérique, chaque administration a son programme pour associer *start-up* et grands groupes à la recherche du futur, sur la sécurité, sur les risques, etc. C'est le problème des incitations.

Bien des *a priori* qui existaient il y a vingt ans ont disparu, la connexion entre réseaux de neurones et approches symboliques progresse, l'explication des modèles aussi. Je suis allé au *Nips* - la *Conference on Neural Information Processing Systems*, le Davos de l'intelligence artificielle - à Barcelone récemment : j'ai constaté la faiblesse de la représentation française dans certains *workshops* absolument primordiaux, comme l'interprétation des modèles ou les composantes éthiques, par exemple.

Pourquoi n'y a-t-il pas en France une politique de *pull* assumée par l'administration, qui interrogerait les chercheurs sur certains sujets d'avenir ?

**Mme Dominique Gillot, rapporteure.** - Il y a deux ans que je sollicite l'OPECST pour engager ce travail. Hélas, c'est dans l'urgence, juste avant la fin de nos mandats parlementaires, que nous nous penchons sur le sujet. Le secrétaire d'État à l'enseignement supérieur et la secrétaire d'État au numérique présenteront demain le début d'une stratégie nationale pour la recherche en intelligence artificielle : répondez à ces invitations, inscrivez-vous dans les ateliers proposés ! Le Conseil national de stratégie de la recherche vient d'étudier le Livre blanc qui clôture la séquence des cinq dernières années, en matière d'enseignement supérieur et de recherche - la recherche y apparaît bien comme une nécessité fondamentale, la liberté des chercheurs étant mise en avant.

J'ai l'impression que les politiques sont toujours en réponse, jamais en tête de pont - ou en *push* ou, mieux, en *pull*, pour reprendre vos termes. C'est vrai que le politique n'est pas en *pull*. C'est bien là l'enjeu de rencontres comme celle d'aujourd'hui, qui doivent être valorisées, afin que les chercheurs, ceux qui travaillent sur l'avenir, orientent la décision politique. Aux États-Unis, les décideurs ont manifestement plus d'audace, ils prennent des risques, alors que, chez nous, en politique comme ailleurs, le principe de précaution l'emporte...

M. Jean Ponce. - Je veux dire, après avoir écouté M. Demazeau, qu'il existe un bon domaine intégrateur pour l'intelligence artificielle, c'est la robotique. Il doit être poussé et valorisé comme tel. Il y a quelques jours, aux États-Unis, deux cent cinquante millions de dollars ont été débloqués autour de la Carnegie Mellon University, à Pittsburgh, par un effort public-privé, pour financer des projets de robotique.

M. Patrick Aknin. - Je suis directeur scientifique de l'Institut de recherche technologique SystemX, qui fait le lien entre laboratoires et industrie sur des sujets où intervient l'intelligence artificielle, comme les véhicules autonomes, les *smart grids* ou réseaux électriques intelligents, les territoires intelligents, etc. Un frein à l'exploitation de l'intelligence artificielle réside, me semble-t-il, dans les lacunes de l'explication des décisions automatisées, leur « prouvabilité ». Il y a là un défi sociétal. S'y

---

pencher permettrait de déverrouiller des propositions de chercheurs, afin d'aider l'homme dans sa vie quotidienne.

Cédric Villani, qui n'est pas connu comme un pro-connectionniste ou un mentor de l'apprentissage statistique, a parlé pourtant d'interactions dans le secteur : certains chercheurs assurent que, dans quelques années, on aura des preuves et une théorisation de ces approches par apprentissage, par exemple avec la théorie de *Vapnik* qui tente d'expliquer l'apprentissage d'un point de vue statistique. Après l'émergence du *deep learning* dans la période actuelle, on aura des propositions de théorisation. Il faudrait encourager ce type de travaux, pour dépasser les peurs collectives sur ces sujets.

M. Patrick Albert. - Je suis l'un des vétérans dans ce domaine, la *good old-fashioned AI* ou bonne IA à l'ancienne, avec une thèse de doctorat au centre de recherche de Bull, la création d'un premier centre industriel puis une des premières *start-up*, ILOG, dont le nom vient de l'abréviation des termes intelligence logicielle... vendue, hélas !, à IBM, car il n'y avait personne pour la racheter en France. Bref, j'ai une carrière à la fois scientifique et industrielle, et je suis membre du bureau de l'Association française pour l'intelligence artificielle (AFIA).

Pour déterminer des priorités, les critères sont nombreux, scientifiques, économiques, ayant trait à la défense nationale, à la citoyenneté, etc. Nous vivons de plus en plus sous le règne de l'algocratie, avec des algorithmes qui intègrent un apprentissage et qui prennent des décisions, mais sur quels motifs ? Google et Facebook sont ainsi régulièrement accusés de racisme : ils se bornent à répondre que leurs systèmes créent des distributions de probabilité qui fonctionnent. Ce n'est pas socialement acceptable, il faut pouvoir expliquer les décisions. Les entreprises doivent pouvoir justifier leurs choix. Pour l'allocation des prêts, les banques utilisent beaucoup ces outils, mais elles savent expliquer en partie la décision prise.

L'organisation de la recherche est un sujet difficile, il faudra pourtant progresser sur ce point pour mieux développer l'intelligence artificielle. J'en parle d'autant plus librement que je n'appartiens à aucune organisation. Chaque sous-domaine de ce secteur de la recherche a des velléités d'indépendance ; il est compliqué de réunir toutes les communautés. Des efforts sont à accomplir. L'intelligence artificielle est une technologie diffusante, dont les impacts doivent être pris en considération. Il faut aider les différentes communautés à s'organiser. Une organisation la moins centralisée possible serait la meilleure forme, même s'il faut fédérer l'ensemble des travaux et représenter le secteur auprès de l'État et la société.

**M. Claude de Ganay, rapporteure.** - Je vous remercie tous de votre participation à cette première table ronde et je laisse Dominique Gillot présider la deuxième table ronde.

---

### III. DEUXIÈME TABLE RONDE, PRÉSIDIÉE PAR MME DOMINIQUE GILLOT, RAPPORTEURE : DIMENSIONS STRATÉGIQUES EN MATIÈRE DE RECHERCHE EN INTELLIGENCE ARTIFICIELLE

**Mme Dominique Gillot, rapporteure.** – J'invite en premier lieu Jean-Marc Merriaux, Directeur général de Canopé, qui est la prolongation du Centre national de documentation pédagogique (CNDP). Après avoir effectué une carrière de quinze ans à France 5, Jean-Marc Merriaux est toujours très intéressé par les actions destinées à la communauté éducative et s'investit dans l'éducation populaire. Vous avez sept minutes pour dresser un portrait des enjeux éducatifs de l'intelligence artificielle.

#### 1. M. Jean-Marc Merriaux, directeur général de Canopé

Merci pour cette invitation. Bien évidemment, je ne suis ni un spécialiste de l'intelligence artificielle ni un chercheur. Je suis plutôt ici à titre de médiateur faisant le lien entre tous les enjeux pédagogiques attachés aux évolutions technologiques. Canopé, opérateur du ministère de l'éducation nationale, est présent pour accompagner l'évolution des pratiques pédagogiques.

##### *Enjeux de l'intelligence artificielle*

Interrogé sur la prise en charge de l'intelligence artificielle par l'éducation, je me suis efforcé de bien positionner les thèmes du débat, à commencer par la problématique majeure de l'incommunicabilité entre les sciences, les technologies et la société.

L'éducation vise à repositionner l'humain au cœur de l'ensemble des dispositifs. L'humain peut-il changer grâce et par le numérique ? L'autre problématique a trait à la place de l'humain dans une société hyper-connectée : robots, transhumanisme, interactions avec les machines, ubérisation des économies... Nous nous situons également dans un rapport au temps entre la pédagogie et l'ensemble de ces questions technologiques. Malgré le fait que les sciences et technologies n'ont jamais autant accompli pour améliorer la vie de l'homme, l'ensemble de ces technologies nouvelles font peur, ce qu'il ne faut pas sous-estimer.

L'intelligence artificielle aujourd'hui, sait comprendre ce qu'on lui demande, s'exprimer correctement et répondre à des questions simples. Nous avons évoqué *Siri* et un certain nombre d'autres évolutions lors de la première table ronde. Néanmoins l'intelligence artificielle ne sait pas encore comprendre le sens du langage adapté au contexte. On sait que le sens du langage et de l'oralité sont intégrés de manière très forte dans le cadre

---

pédagogique. Demain, l'intelligence artificielle pourra sans doute répondre à des enjeux pédagogiques et inventer son propre langage, expérimenter le monde pour comprendre le sens du langage. L'éducation est aujourd'hui notre seul atout pour réduire l'incommunication entre sciences, technologie et société. Les pédagogues, au premier chef, doivent en avoir pleinement conscience. De ce fait, se pose la question de la place des sciences et des technologies du numérique dans tous les apprentissages pour comprendre une société qui se complexifie.

*Apports pédagogiques de l'intelligence artificielle*

Sur les apports pédagogiques de l'intelligence artificielle, j'ai tenté de dresser un éclairage rapide. De nouvelles pratiques pédagogiques pourraient émerger en dépassant la relation émetteur-récepteur et en transformant l'évaluation. Il est vrai que les enjeux de l'intelligence artificielle nous conduisent à nous interroger sur l'évaluation et, entre autres, sur les moyens de prédire la réussite des élèves. Il nous faut pouvoir intégrer cette dimension. De plus, la répétition est avant tout un acte pédagogique, de sorte que l'intelligence artificielle repose beaucoup sur ces activités de répétition. Par ailleurs, la différenciation des apprentissages constitue un autre aspect essentiel de la pédagogie, la personnalisation devant être adaptée à la diversité des élèves. J'insisterai enfin sur un dernier point, tenant au *continuum* pédagogique entre le temps scolaire et le hors temps scolaire. La présence future, peut-être, de robots aussi bien à l'école qu'au sein de la maison, devra être accompagnée d'outils et d'interfaces pour assurer les usages au sein et en-dehors de la classe. Sur ce point, l'intelligence artificielle peut sûrement apporter un certain nombre de réponses.

J'insisterai essentiellement sur la place et le rôle de l'enseignant du fait de l'ensemble des évolutions évoquées, ainsi que sur la nécessité de l'accompagner dès lors que les nouvelles technologies seront parties intégrantes de la classe. L'intelligence artificielle va transformer le geste professionnel de l'enseignant. De même, la médiation constituera un outil important pour répondre aux enjeux technologiques. Selon moi, les nouvelles technologies ne sont pas conçues pour être en compétition avec les enseignants. L'intelligence artificielle intervient surtout pour compléter le savoir-faire de l'enseignant, en le rendant plus accessible et mieux informé. La posture du médiateur constitue par conséquent un enjeu essentiel de la nouvelle relation à l'élève. La médiation représente un enjeu de réassurance pour l'éducation dans un monde en pleine mutation technologique. Elle renforce la place de l'humain et garantit également le collectif. Le médiateur est un tiers de confiance, ce qualificatif s'appliquant *a fortiori* à l'enseignant. La médiation favorise en outre une meilleure adéquation avec les besoins de chaque usager.

Enfin, j'évoquerai un exemple sur lequel je travaille dans le cadre du projet *e-fran*, qui liera aussi bien des pédagogues que des chercheurs et des *start-uppers*. Nous avons développé *Matador*, qui était à la base un jeu de

---

plateau mais qui a été transformé dans un environnement numérique. Le projet repose sur un monitoring individuel d'apprentissage du calcul mental par chaque élève. Plus l'élève jouera en classe et à la maison et mieux l'enseignant connaîtra ses compétences acquises et non acquises. Nous travaillerons sur le parcours d'un élève qui sera mis en rapport avec tous les autres élèves de son niveau scolaire. L'interaction et l'horizontalité constituent par conséquent des éléments importants de ce type de projet. *E-fran* reposera sur mille cinq cents élèves pendant une année scolaire, avec l'objectif d'analyser plus de sept cent mille opérations chaque année. Selon les profils, des parcours de jeu spécifiques seront proposés à chaque élève. L'ensemble s'appuie sur des chercheurs, aussi bien statisticiens que didacticiens et cognitivistes. Nous travaillons dans cette dynamique en intégrant l'ensemble des questions développées à l'occasion de la première table ronde, même si nos enjeux ne sont pas aussi poussés que certains domaines de l'intelligence artificielle.

Je suis au regret de devoir m'éclipser pour prendre un train. Je peux répondre aux questions maintenant.

**Mme Dominique Gillot, rapporteure.** – Nous devons préalablement achever les exposés. Je passe donc la parole à François Taddéi, qui possède de multiples compétences mais est invité ce jour en sa qualité de directeur du CRI, donc praticien des rencontres des intelligences. Son intervention portera sur le travail qu'il a engagé concernant l'évolution des pratiques pédagogiques liée à l'évolution des technologies.

## **2. M. François Taddei, directeur du Centre de recherche interdisciplinaire (CRI)**

Effectivement, la ministre de l'Éducation nationale m'a confié une mission ayant trait à la société apprenante. Il s'agit de déterminer comment nous apprenons tout au long de la vie, que ce soit dans la petite enfance, dans l'enseignement scolaire supérieur ou au cours de la vie professionnelle.

Aujourd'hui, l'intelligence artificielle et l'intelligence humaine co-évoluent. En ma qualité de biologiste de l'évolution, je constate que nous avons connu un certain nombre de grandes transitions évolutives avec l'apparition de la vie et même avec celle des cerveaux. La connexion de ces cerveaux entre eux délivre une série de messages coordonnés pour progressivement accomplir un nombre croissant de réalisations. Puis l'homme est apparu et a inventé d'autres manières, toujours plus élaborées, de communiquer. Le langage est bien entendu l'une d'entre elles, avant l'écriture, l'imprimerie et l'informatique. Il existe aujourd'hui une accélération de ces transitions qui comportent une certaine part de singularité technologique. Dès lors, nous avons besoin d'inviter des acteurs de tous horizons à réfléchir à cette pluralité de futurs, que l'on peut souhaiter ou non. Cette pluralité d'acteurs inclut les enfants, toujours très



---

intéressés par les thématiques nouvelles. Lorsqu'on donne aux enfants la possibilité de s'interroger, nous constatons qu'ils peuvent aller très loin. Nous avons ainsi développé un programme dénommé « Les savanturiers », à l'occasion duquel des enfants se questionnent et travaillent avec des chercheurs. Assez spontanément, ils posent des questions telles que l'existence avant le *big bang* ou le devenir de l'humain. Les enfants sont fondamentalement attirés par les sujets ayant trait à l'intelligence artificielle et la robotique. Ils se demandent spontanément quelle sera leur place dans ce monde et s'ils co-évolueront avec les nouvelles formes d'intelligence.

L'intelligence collective est présente depuis longtemps dans le vivant, mais passe progressivement à un collectif d'hommes et de machines. Dans le monde des échecs, on parle d'IA-centaures, combinaisons d'hommes et de machines qui battent les hommes mais aussi les machines jouant seuls. Voilà ce qui marche le mieux.

Le besoin de questionnement éthique est en outre présent. Il nous interroge sur la place de l'homme. Les plateformes de recherche participative peuvent mobiliser l'intelligence collective au service de grands enjeux. François Houllier a ainsi rendu plusieurs rapports au terme de la mission sur les sciences participatives qui lui a été confiée. Les intelligences artificielles sont capables d'optimiser un jeu de données mais ne peuvent pas spontanément inventer un autre jeu ou se donner d'autres objectifs. L'école cherche à apprendre à mémoriser le passé, ce que l'intelligence artificielle sait faire, mais il faut aussi inventer. Cette capacité à se questionner est encore une spécificité humaine qu'il faut approfondir toujours plus. Nous devons travailler sur la complémentarité, coopérer pour questionner demain et co-construire ces solutions. Les grands défis planétaires nécessitent toujours plus d'intelligence. Il faut convaincre nos concitoyens que nous ne devons pas arrêter la course, mais plutôt réfléchir aux raisons pour lesquelles nous courons. Nous devons nous poser ces questions collectivement, en ne laissant pas l'évolution technologique s'affranchir des questions éthiques. Aristote disait déjà qu'il existe trois formes de connaissance : *épistémè*, la connaissance du monde (qui a donné l'épistémologie et la science), *technè*, l'action sur le monde (qui a entraîné l'apparition des technologies dont nous discutons aujourd'hui) et *phronésis*, l'éthique de l'action ou sagesse pratique. Nous avons donc besoin de réfléchir simultanément à la coévolution de ces sciences technologiques et de cette éthique individuelle et collective. La réflexion doit être de court et long terme, locale et globale. C'est pourquoi il est nécessaire de former les décideurs sur les grands enjeux. Par conséquent, je me félicite de cette mission parlementaire. La prise de conscience de ces enjeux doit faire l'objet d'un débat toujours plus large et évidemment international. Nous devons impliquer toujours plus d'acteurs dans la réflexion. Notre système éducatif et notre enseignement supérieur devront être pensés pour s'adapter au monde de demain. Les institutions d'intelligence humaine seront tenues de se réinventer à l'heure où les machines vont les concurrencer.

---

**Mme Dominique Gillot, rapporteure.** – La place et la transmission de la connaissance constituent évidemment des enjeux fondamentaux. Je vais passer la parole à Olivier Esper, responsable des affaires publiques de Google France. Vous allez nous présenter les activités de Google et de sa filiale en matière de recherche en intelligence artificielle. Nous sommes évidemment tous très intéressés par ce qui se passe chez Google, notamment par l'appel à la sensibilisation à la responsabilité sociétale de cette grande société.

### 3. M. Olivier Esper, responsable des affaires publiques de Google France

Je vous remercie de m'avoir invité à vos travaux pour y apporter ma modeste contribution, n'étant moi-même pas un chercheur. Je sais que vous prévoyez également de vous déplacer à l'international, j'espère que vous aurez l'occasion de rencontrer Greg Corrado le directeur de la recherche de Google.

**Mme Dominique Gillot, rapporteure.** – Oui c'est bien prévu lors de notre déplacement aux États-Unis la semaine prochaine.

**M. Olivier Esper.** – Très bien. À l'été 2015, les fondateurs de Google ont créé une maison-mère nommée Alphabet pour permettre à différentes entreprises distinctes - et désormais autonomes par rapport à Google - de porter des projets intégrant de l'intelligence artificielle : *Waymo* pour la voiture autonome, *Deep Mind* pour le programme AlphaGo...

Pour sa part, Google concentre essentiellement ses recherches sur l'apprentissage automatique, le *machine learning*, pour concevoir des machines qui répondront intelligemment à des solutions et situations nouvelles. Ces machines sont entraînées avec des jeux de données. La recherche menée par Google correspond à un modèle ouvert, que des ingénieurs extérieurs à Google peuvent utiliser dans d'autres projets. Le constat posé par Google est celui du franchissement actuel d'un cap. De plus en plus d'usages, de traductions concrètes de la recherche existent aujourd'hui dans le monde réel, et sont la résultante de capacités de calcul augmentées, de sources d'informations très riches et de communautés larges de chercheurs et d'ingénieurs.

La plupart des services de Google intègrent de l'apprentissage machine, avec trois objectifs. Le premier d'entre eux est d'améliorer la performance des produits. Par exemple, *Google Translate*, appelé en France *Google Traduction*, reposait initialement sur une traduction mot à mot. En 2016, la nouvelle version de ce service intègre de l'apprentissage machine. Ce changement a abouti à une augmentation des performances inégalée au cours des 10 dernières années. Le second de ces objectifs est de proposer des nouvelles fonctionnalités. Par exemple *GooglePhoto* permet désormais aux utilisateurs d'effectuer des recherches dans leur bibliothèque sur des mots-clés sans avoir au préalable annoté ces photos. Le troisième objectif des

---

chercheurs de Google est de se confronter à des problématiques à grande échelle. Par exemple, Gmail est maintenant capable de bloquer plus de 99,9 % du spam, y compris des variantes de messages jamais encore vus auparavant, grâce à l'utilisation de techniques de réseaux neuronaux.

En dépit des progrès réalisés, les champs de recherche sont encore nombreux. J'en citerai trois identifiés par nos ingénieurs, et qui feront écho à certains éléments que nous avons abordés précédemment. Le premier champ de recherche a trait à la compréhension des modèles. Il s'agit de développer des modèles auto-explicatifs susceptibles d'expliquer un résultat par rapport à une situation spécifique. En deuxième lieu, les ingénieurs Google travaillent sur l'apprentissage machine à partir de petits jeux de données. En la matière la qualité et le volume de ces données sont d'importance. Par exemple, pour reconnaître un chat sur une photo, nous avons besoin de millions de photos de chats pour entraîner les machines. Or la possibilité d'entraîner ces machines sur de plus petits jeux d'échantillons permettrait de mettre en œuvre l'apprentissage automatique sur un plus grand nombre de domaines.

Enfin, le troisième champ de recherche concerne l'amélioration de la compréhension de sémantiques subtiles des modèles. Pour reprendre l'exemple des photos, le modèle est capable de repérer un chat mais ne sait pas encore dire si la photo est triste ou drôle. Une telle contextualisation de l'image n'est pas encore permise. De même, le système ne sait pas encore comprendre qu'une vache volant dans le ciel n'est pas une situation normale.

En parallèle de ces domaines de recherche de nature scientifique, des travaux sont également réalisés sur des enjeux sociétaux et éthiques. Comme d'autres technologies, l'apprentissage machine n'est pas bon ou mauvais en soi car tout dépend de l'usage qui en est fait. Afin de réfléchir aux enjeux sociétaux éthiques, nous sommes membres fondateurs de « *Partnership on AI* »<sup>1</sup> avec Facebook et d'autres entreprises, afin de développer des bonnes pratiques et de maximiser les retombées positives des nouvelles technologies. Ici encore, je citerai trois domaines sur lesquels nous travaillons. En premier lieu sur les risques de sécurité que pourrait poser l'intelligence artificielle, nos chercheurs ont publié l'an dernier un papier en collaboration avec les universités de Stanford et de Berkeley<sup>2</sup>.

Le deuxième exemple de recherches a trait au risque de partialité ou de discrimination si la machine est elle-même entraînée sur un jeu de données biaisé. Dans ce domaine encore, nos ingénieurs ont publié des travaux. Lors de la conférence NIPS à Barcelone, organisée en décembre dernier sur l'intelligence artificielle, deux de nos ingénieurs ont publié un papier pour cerner la problématique et proposer une approche de solutions<sup>3</sup>.

En troisième lieu, l'impact économique de l'intelligence artificielle, c'est-à-dire sa traduction concrète sur les emplois présents et futurs, conduit

---

<sup>1</sup> <https://www.partnershiponai.org>

<sup>2</sup> <https://research.googleblog.com/2016/06/bringing-precision-to-ai-safety.html>

<sup>3</sup> <https://research.googleblog.com/2016/10/equality-of-opportunity-in-machine.html>

---

à réfléchir aux métiers de demain et à ajuster l'éducation en conséquence. Dans ce domaine, les études sont loin de converger vers les mêmes chiffres, mais il importe d'ores et déjà de réfléchir à la formation.

En conclusion au-delà des exemples que j'ai mentionnés, l'objectif de Google est de rendre ses services plus performants et utiles dans la vie quotidienne des gens. L'intelligence artificielle permettra aussi de se confronter aux grands problèmes de notre société. Dans le domaine de la lutte contre le réchauffement climatique, Google a utilisé de l'apprentissage machine pour optimiser l'allocation de ressources de ses *data centers*. Cette mesure a permis une réduction de 40 % de la consommation d'énergie. L'entreprise s'est en effet fixé pour objectif de parvenir dès la fin 2017 à 100 % d'approvisionnement en énergie renouvelable, en particulier grâce à l'utilisation de l'intelligence artificielle. Ces avancées technologiques ont donc un potentiel positif.

**Mme Dominique Gillot, rapporteure.** - Delphine Reyre est directrice Europe des affaires publiques de Facebook, deuxième site *web* le plus visité au monde après Google. Facebook fait un usage croissant des technologies d'intelligence artificielle. J'ai rencontré à deux reprises Yann LeCun, directeur de la recherche en intelligence artificielle de Facebook, qui dirige la recherche mondiale de Facebook mais a installé un laboratoire de recherche en France.

#### **4. Mme Delphine Reyre, directrice Europe des affaires publiques de Facebook**

Merci beaucoup pour votre invitation. Je souhaite juste apporter une petite note personnelle. Avant de rejoindre Facebook, j'ai travaillé pour un grand ministre de la recherche, Hubert Curien, qui entretenait des liens très étroits avec l'OPECST. Je suis donc présente parmi vous avec un plaisir particulier. Olivier a très bien décrit comment les grandes entreprises du numérique abordent les sujets de l'intelligence artificielle. Je ne serai par conséquent pas trop redondante, étant précisé que Yann LeCun sera disposé à répondre à toutes les autres questions par la suite. Je vous propose plutôt de vous apporter un éclairage particulier, car sous l'impulsion de Yann nous avons choisi d'établir notre laboratoire d'intelligence artificielle en grande partie à Paris. Je préciserai comment s'exerce le dialogue entre la recherche publique et la recherche privée sur ces domaines en évolution rapide.

Facebook a investi deux grands domaines d'intelligence artificielle. Un groupe travaille sur la recherche appliquée, en particulier sur des applications concrètes dont l'un des exemples-types concerne les capacités de reconnaissance visuelle. Ces applications permettent à des malvoyants ou des non-voyants d'utiliser Facebook et donc d'accéder au monde des réseaux sociaux. La machine doit par exemple permettre de lire ou de décrire à ces personnes l'image qu'ils reçoivent sur leur fil d'actualité. Le décloisonnement des usages des médias sociaux est ainsi opéré au bénéfice

---

des personnes atteintes de handicaps. Dans son ensemble, le thème de l'accessibilité connaîtra un développement rapide.

Par ailleurs, je souhaitais évoquer plus précisément notre activité de recherche fondamentale, dénommée Facebook *Artificial Intelligence Research* (FAIR), développée par trois laboratoires basés respectivement à Menlo Park dans la Silicon Valley, New-York où enseigne Yann LeCun, et Paris. Le choix de Paris s'est bien entendu opéré sous l'impulsion de ce grand chercheur qui a bien voulu travailler avec Facebook, mais aussi parce que l'excellence de l'école mathématique française est mondialement reconnue. L'INRIA, l'ENS, l'Université Pierre-et-Marie-Curie ainsi que d'autres universités sont véritablement des sources d'excellence, de sorte qu'il nous a paru important de pouvoir travailler avec leurs chercheurs sans nécessairement les faire partir aux États-Unis. D'ailleurs ces chercheurs ne souhaitaient pas véritablement s'expatrier. Ainsi, l'installation du laboratoire à Paris nous a permis de fixer les chercheurs en France, d'en faire revenir d'autres et même d'attirer des chercheurs étrangers. Sur vingt-cinq chercheurs accueillis à Paris, figure en effet une proportion importante de chercheurs étrangers. Telle est la preuve que la France peut être un centre d'excellence et une terre d'attraction pour l'intelligence artificielle.

Ainsi que je l'évoquais, nous travaillons avec la recherche publique et en partenariat étroit avec l'INRIA. Je souhaite également apporter un éclairage sur un instrument très spécifique à la France, les Conventions industrielles de formation par la recherche (CIFRE), qui permettent une interaction formidable entre la recherche publique et la recherche privée. Les thésards et les doctorants accueillis en milieu de recherche privée, auront sans doute vocation à exercer dans la recherche publique. Facebook a également mis en place un programme de dons de moyens de calcul à des universités françaises et européennes, parmi lesquelles l'INRIA, l'ENS et l'Université Pierre-et-Marie-Curie. Nous avons récemment annoncé un certain nombre de dons de ce type en France.

Lorsque nous maintenons des chercheurs en France, ils continuent à participer activement à la communauté scientifique française. Ces chercheurs enseignent et apportent leur contribution à des conférences. Yann LeCun a, à ce titre, ressenti comme un très grand honneur de donner des conférences au Collège de France, dans le cadre de sa chaire d'informatique.

Enfin, la quasi-totalité des travaux effectués par Facebook sont publiés en mode ouvert, de sorte qu'ils sont immédiatement saisissables par tous chercheurs. Il existe donc un véritable effet de démocratisation en intelligence artificielle grâce à ce mode ouvert. La recherche n'est pas réservée à de grands groupes privés et est mise à la disposition de tous.

En termes d'éthique, Facebook est partenaire avec Microsoft, Google, IBM et Amazon du *Partnership on AI*. Il est également important de souligner que nous travaillons quasi exclusivement sur des données publiques.

**Mme Dominique Gillot, rapporteure.** – Vous n'avez pas parlé du crédit d'impôt recherche (CIR).

---

**Mme Delphine Reyre.** - Notre objectif est de trouver avant tout des cerveaux. À cet égard le CIFRE nous satisfait davantage.

**Mme Dominique Gillot, rapporteure.** - Merci beaucoup pour votre exposé très précis. L'intervenant suivant est Laurent Massoulié. Votre expérience est également très intéressante puisque vous êtes directeur d'un centre de recherche commun à l'INRIA et à Microsoft. Ce type de partenariat a vraisemblablement vocation à se développer.

### **5. M. Laurent Massoulié, directeur du Centre de recherche commun INRIA-Microsoft**

Je vous remercie de votre invitation. J'ai rejoint l'INRIA il y a quatre ans pour diriger le laboratoire commun entre Microsoft Research et l'INRIA. Je suis chercheur dans les réseaux et les systèmes distribués. Le laboratoire commun Microsoft Research-INRIA est né il y a dix ans d'une ambition de collaboration étroite entre les directeurs de recherche de Microsoft et de l'INRIA. Le projet a été rendu possible car les chercheurs se connaissaient. Actuellement, le laboratoire héberge une quarantaine de chercheurs travaillant sur six projets, dont quatre sont liés à l'intelligence artificielle.

L'un de ces projets, ayant trait à la vision sur ordinateur et dirigé par Jean Ponce ici présent, traite de la compréhension automatique des vidéos. Un autre projet concerne l'imagerie médicale, avec l'objectif de parvenir à poser des diagnostics personnalisés et à aboutir à une meilleure prévention grâce à l'intelligence artificielle. Enfin nous avons des projets autour des fondamentaux des algorithmes pour l'apprentissage supervisé ou non supervisé, notamment dans le *Cloud*, par exemple, la plateforme de *cloud computing* de Microsoft.

Force est de constater que les partenariats entre un institut tel que l'INRIA et les laboratoires de recherche industrielle comme celui de Microsoft sont non seulement faisables, mais seront de plus en plus nombreux à l'avenir. En réalité, les chercheurs qui font avancer la science dans le privé et le public sont issus des mêmes formations, se connaissent et assistent aux mêmes conférences. Par conséquent, la frontière entre le privé et le public est de plus en plus perméable, et ce dans les deux sens. Dans la conduite de nos projets, nous avons formé de jeunes chercheurs qui, pour un certain nombre d'entre eux, ont été recrutés ensuite par Microsoft. Nous avons également connu des transferts dans l'autre direction. Pour ma part, je suis l'un des trois cas de chercheurs ayant quitté Microsoft pour rejoindre l'INRIA.

L'objectif des partenariats public-privé vise en premier lieu à ce que le privé apporte ses problèmes à la recherche publique. Les chercheurs en vision tels que ceux de l'équipe de Jean Ponce, s'ils peuvent être au contact avec les équipes de Microsoft sur la réalité augmentée, seront ses contributeurs sur le sujet. Dans l'autre sens, il existe un engouement des

---

grands groupes tels que Microsoft sur le fait de travailler avec les chercheurs de l'INRIA, dont l'excellence est reconnue et qui disposent d'une grande latitude pour s'engager sur des objectifs de long terme. Par exemple lorsque Microsoft décide de s'intéresser à l'imagerie médicale, les chercheurs de l'INRIA recrutés ont déjà des dialogues approfondis avec les centres hospitaliers universitaires (CHU) et sont en prise avec les vrais problèmes de l'imagerie médicale.

Je conclurai en affirmant que le partenariat public-privé est un excellent moyen de faire avancer la connaissance dans le privé et l'innovation dans le public. De plus l'attractivité des salaires dans le public restant moindre que dans le privé, ce genre de partenariats peut représenter une solution.

**Mme Dominique Gillot, rapporteure.** - Sans transition, je donne la parole à Dominique Cardon, professeur de sociologie à l'Institut d'Études Politiques de Paris et ancien de Médialab, spécialisé dans la sociologie de la recherche et des mondes numériques.

### **6. M. Dominique Cardon, professeur de sociologie à l'Institut d'Études Politiques de Paris/Médialab**

L'intelligence artificielle est née trois fois, morte deux fois, et réapparaît aujourd'hui dans l'espace public. Elle a sans doute un problème de communication. Pour l'apprivoiser, nous aurions besoin de nous défaire de cette hypothèse d'estompement de la frontière entre l'humain et l'inhumain, et de cette obsession anthropomorphique. L'intelligence artificielle revêt plusieurs formes car les conceptions de l'intelligence sont différentes, il y a donc plusieurs types d'intelligence artificielle. J'observe une tentative des anciennes générations de prendre le train des nouvelles formes d'intelligence artificielle, comme le *deep learning*. Il est important pour la collectivité de chercheurs de ne pas faire de promesses inconsidérées dans cette représentation de la science autonome de la machine.

Par ailleurs, les nouvelles techniques étant massivement statistiques, la question des données est très centrale. Les enjeux posés à la société sont à la fois éthiques et concernent la complémentarité entre recherche publique et recherche privée. Il faut se demander à partir de quelles données les modèles sont aujourd'hui construits. Les données appartiennent en effet aux grandes entreprises privées, les GAFAs en particulier (Google, Apple, Facebook, Amazon). On voit apparaître de nouvelles formes d'articulation entre recherche publique et recherche privée, sous l'impulsion des *meetups* en particulier, comme celui qu'organise Igor Carron

Comprendre, critiquer et apprivoiser sont les principaux enjeux. Le premier de ces enjeux, épistémologique, vise à comprendre le fonctionnement de l'intelligence artificielle et à éduquer la société aux nouveaux objets. Il faut en faire à la fois un enjeu scientifique et éducatif, ainsi qu'il l'a été rappelé. À ce titre, il est nécessaire d'étudier l'informatique,



les mathématiques et les statistiques. Google possède une manière de classer l'information qui lui est propre, Facebook a également une pratique spécifique, elle aussi discutable. Il est par conséquent nécessaire de faire comprendre à nos sociétés les visions du monde qui sont derrière ces représentations pour savoir les critiquer. Les inquiétudes sur ces questions sont nombreuses. Il existe des biais.

La dynamique actuelle autour du Conseil national du Numérique et de l'INRIA correspond à un besoin de notre société de se saisir des enjeux des algorithmes, en dialogue avec les chercheurs, pour éventuellement vérifier les classements d'un certain nombre d'opérations ou le respect de principes de diversité. Il faut qu'il y ait des débats. Sur les monopoles, sur la fixation des prix, sur le respect du principe de diversité ou sur les bulles d'informations, il faut que la société soit partie prenante de ces discussions. Le concept de loyauté des algorithmes vise à mettre en place des techniques informatiques pour vérifier la manière dont les utilisateurs comprennent ce que fait l'algorithme pour eux. Les travaux de Benjamin Edelman à ce sujet méritent d'être cités.

Apprivoiser les différentes formes d'intelligence artificielle suppose de cesser de penser les intelligences artificielles comme des machines autonomes. C'est une image dont il faut se défaire car nous en sommes encore très loin en vérité. Il faut se désenvoûter de ces croyances pour mieux vivre avec ces nouvelles technologies.

**Mme Dominique Gillot, rapporteure.** – Merci d'avoir posé la question de la détention des données, qui suscite beaucoup de fantasmes, et de la capacité à les engranger. Le monopole pose question à notre société et à nos usagers. Nous terminons avec Gilles Babinet, dont le témoignage est précieux car il a été dès 2011, le premier président du Conseil National du Numérique et il est aujourd'hui *Digital Champion* auprès de la Commission européenne, il pourra profiter de son intervention pour nous parler de cette nouvelle mission.

### **7. M. Gilles Babinet, entrepreneur, digital champion auprès de la Commission européenne**

Bonjour, et merci de votre invitation à participer à vos travaux. J'ai quelque peu restructuré mes propos au fur et à mesure des interventions de ce matin. Je ne possède aucune expertise poussée en intelligence artificielle, même si ce sujet me passionne en tant que néophyte. Pour autant, je souhaite le rapprocher des enjeux productifs et d'enseignement en général, aussi bien dans le primaire que dans le supérieur. Depuis une dizaine d'années, une résurgence de l'intelligence artificielle a été constatée, notamment sous l'impulsion de personnalités telles que Yann LeCun et d'autres encore. Les applications concrètes sont disponibles depuis peu. Des intelligences artificielles sont déjà partout, dans notre quotidien. Nous avons tous été



---

confrontés à une intelligence artificielle *via nos smartphones* ou les feux rouges dans les villes. Cette intelligence artificielle faible, qui fait partie du quotidien de chacun de nous, devrait à ce titre être davantage enseignée et être davantage connue.

L'impact productif constitue une préoccupation majeure. Vous connaissez le grand débat actuel qui porte sur la baisse des gains de productivité au sein de l'OCDE. En France, il s'agit d'une priorité puisque nous sommes passés sous la barre des 1 % depuis deux ans. Or il est manifeste que les technologies d'intelligence artificielle sont capables d'appréhender des environnements d'incertitude élevée et de créer de la productivité dans bien des domaines : recherches documentaires, recherches juridiques et jurisprudentielles, transports de personnes... Je pense par exemple aux clercs de notaires ou aux conducteurs de véhicules. Il est par conséquent essentiel d'aider le plus largement possible à la diffusion de ces techniques, dans la mesure où des enjeux de compétitivité entre les nations s'expriment à cet égard. Il faut à la fois faire de la recherche fondamentale et de la recherche appliquée.

Il convient de mettre en place des modèles d'apprentissage et de diffusion de la science les plus pertinents possibles. Aussi bien en termes épistémologiques que d'enseignement, le numérique implique des nouvelles formes de transmission des savoirs.

Nous avons tout d'abord besoin d'interdisciplinarité. Je ne reviendrai pas sur les notions et les enjeux de pluridisciplinarité très largement évoqués précédemment, mais j'évoquerai une nouvelle fois la thématique des mondes ouverts.

Les mondes ouverts interviennent aujourd'hui dans une dynamique qui ne peut plus être ignorée, avec une efficacité bien supérieure au système de recherche verticale. Je suis très sensible au fait que nous disposions en France d'institutions telles que l'École Normale Supérieure (ENS) qui, en termes de ratios sur le nombre de prix Nobel, est en tête de tous les classements. Pour autant, il s'agit d'une école très élitiste donc limitée en nombre d'élèves. Nous pouvons saluer les travaux essentiels réalisés au sein de l'ENS dans le domaine de l'intelligence artificielle, mais ces travaux ne me semblent pas rendre compte des dynamiques du monde à venir. La plateforme d'*open source* *GitHub*, qui réunit à elle seule plus de seize millions de codeurs, produit trente-sept millions de projets informatiques, dont entre quarante et soixante mille projets dans le domaine de l'intelligence artificielle équivalant à plus de cent cinquante milliards de lignes de code. Or il n'y a aucune réflexion stratégique de l'enseignement supérieur et de la recherche pour massifier les dynamiques rencontrées sur ces plateformes et bâtir des ponts avec ces systèmes de science ouverte. Il faudrait systématiser au sein des institutions de recherche françaises et de l'enseignement supérieur des modèles reposant sur des systèmes d'*open source*.

Pour ma part, je suis proche de *La Paillasse*, laboratoire de sciences ouvertes dans le domaine du vivant, que j'accompagne de jour en jour pour

---

l'aider à subsister. Les subventions dont il bénéficie sont assez limitées. En réalité, je constate qu'il existe une incompréhension absolument totale des institutions en charge des enjeux de recherche, dans un monde où la recherche devient de plus en plus futuriste alors que précédemment elle était incrémentale. En effet, la recherche du XX<sup>e</sup> siècle s'effectuait en blouse blanche, pour inventer par exemple le moteur à explosion qui a sans cesse été amélioré par la suite. Puis est apparue au XXI<sup>e</sup> siècle la personne ayant recueilli le plus de critiques positives dans toute l'histoire de l'automobile pour son véhicule Tesla, alors que sa société n'a que dix ans. Il ne faut pas se leurrer : les techniques de l'innovation mises en œuvre au sein de Tesla s'inspirent du monde du numérique, de l'*open source*, de la pluridisciplinarité et également de contributions externes.

Ces modes de fonctionnement sont très incompris d'éminentes personnalités qui dirigent les unités de recherche ou les universités les plus prestigieuses en France. Il s'agit d'une interpellation de ma part, car je crois important de s'intéresser à ces domaines. En effet, même si l'expertise mathématique continuera d'être essentiellement académique, une relation étroite entre algorithmie et code est mise en œuvre pour faire fonctionner les intelligences artificielles. Autant je peux comprendre que l'algorithmie nécessite des chaires académiques et des laboratoires de recherche, autant la partie code est davantage profane. Tout le monde peut contribuer à innover dans le domaine du code.

En outre, il m'apparaît indispensable de nous intéresser à ces systèmes d'innovation ouverte - et je tiens ces propos avec ma casquette européenne - dans la mesure où le niveau d'investissement de la Chine et des États-Unis dans le domaine de l'intelligence artificielle est devenu considérable. La Chine serait même passée devant les États-Unis, avec la recherche conduite par Baidu qui représenterait à elle seule le tiers du total des dépenses de recherche en intelligence artificielle des États-Unis. De ce fait, nous ne pourrions plus entrer en compétition sur la quantité, mais sur le modèle de science que nous mettrons en œuvre, ce sur quoi nous devons concentrer nos efforts. Or selon un rapport de 2016 de l'OCDE, la France est le pays dans lequel la collaboration entre l'entreprise et l'université est la plus faible. Certes nous pouvons, comme la représentante de Facebook l'a fait, nous féliciter de l'existence des CIFRE et du CIR, mais la transdisciplinarité doit également être encouragée.

J'achèverai mon propos sur la nécessité de créer des ponts entre le public et le privé et entre le primaire, le secondaire, le supérieur et la recherche. Je ne crois pas au futur de la recherche académique au sens strict, car les contributions extérieures seront essentielles à la réussite. Une personne autodidacte ayant appris à coder depuis l'école primaire pourra fournir des apports conséquents dans ce domaine. Il faut une prise de conscience politique sur ce point sous peine d'être relégué à un rang secondaire.

---

**Mme Dominique Gillot, rapporteure.** – Vous rejoignez les premiers intervenants sur le rôle de l'éducation et la formation des esprits à une plus grande agilité pour permettre éventuellement les formations autodidactes. Vous avez, les uns et les autres, apporté des informations qui vont susciter le débat.

## 8. Débat

**M. Igor Carron, entrepreneur, fondateur du Paris Machine Learning meetup.** - J'ai entendu les représentants de l'INRIA ainsi que les intervenants de Google et Facebook, nous tenir des propos très intéressants. En revanche une chose n'a pas été dite. *Scikit-Learn*, bibliothèque logicielle libre dédiée à l'apprentissage automatique, est un outil de création française, utilisé par environ deux cent mille personnes, des ingénieurs du monde entier. Cet outil, utilisant l'*open source* et en partie financée par l'INRIA, n'a pas été mentionné. Lui, tout comme un certain nombre d'autres utilisés par nos *start-ups* et notre tissu économique, sont absents des discussions. Je parle à la représentation nationale sur ce point. *Scikit-Learn* est réalisé en grande majorité en France, et est utilisé par une importante communauté à l'heure actuelle. D'ailleurs cet outil pourrait être accessible à des élèves dès la seconde. Il me semblait par conséquent important de rappeler ici que des projets existent, qu'ils sont sous les radars, y compris de personnes qui parlent d'intelligence artificielle avec autorité, et qu'ils éprouvent de grandes difficultés à survivre.

**Mme Dominique Gillot, rapporteure.** - Le cas de *Scikit-Learn* me rappelle celui d'*OpenClassrooms*, qui lui aussi est invisible.

**M. Igor Carron, entrepreneur, fondateur du Paris Machine Learning meetup.** - Oui, mais le modèle d'*OpenClassrooms* se rapproche plus de celui des MOOC (*massive open online course* ou formation en ligne ouverte à tous).

**M. Bertrand Braunschweig, directeur du Centre de l'INRIA de Saclay.** - Merci à Igor Carron d'évoquer *Scikit-Learn*, dont une partie est développée à l'INRIA de Saclay. La communauté est bien consciente de son importance et se montre soucieuse de son maintien. Je voulais vous rassurer sur le fait que *Scikit-Learn* est plutôt en bonne santé aujourd'hui.

**M. Pavlos Moraitis, directeur du laboratoire d'informatique de l'Université Paris-Descartes.** - Les financements public-privé nécessitent un meilleur développement en France. Les recherches universitaires doivent davantage bénéficier de financements par les grandes entreprises. À titre d'exemple, l'un de mes collègues d'Harvard m'a récemment raconté qu'il lui était tout à fait possible d'appeler un représentant de Google au téléphone, pour obtenir dès le lendemain un financement sur un projet.

---

**M. Olivier Esper, responsable des affaires publiques de Google France.** - Je tiens à signaler que nous offrons des dispositifs similaires en Europe.

**M. Olivier Guillaume, Président O2 Quant.** - Nous constatons que l'intelligence artificielle est uniquement un outil d'aide à la productivité en participant à l'augmentation cognitive de l'homme. En France, nous avons les meilleurs cerveaux en recherche. Comment pourrait-on favoriser l'émergence de l'intelligence artificielle dans l'innovation française et européenne et permettre l'amélioration de nos sociétés ?

**M. Jean-Daniel Kant, maître de conférences à l'Université Pierre-et-Marie-Curie Paris-VI.** - J'ai trouvé l'analyse de Dominique Cardon passionnante et je voudrais conseiller de regarder l'excellente mais très déprimante série *Black Mirror*, qui met en scène des projections très réalistes sur une intelligence artificielle mal gérée. Mon objectif n'est pas de produire des intelligences artificielles autonomes, mais je tiens à souligner que ces intelligences existent déjà de fait.

**M. Jean Ponce, professeur à l'École Normale Supérieure.** - Pas vraiment...

**M. Jean-Daniel Kant, maître de conférences à l'Université Pierre-et-Marie-Curie Paris-VI.** - Les *Google Cars* en font partie. Il y a aussi le cas de conseils d'administration où siègent des intelligences artificielles. Je n'ose pas imaginer les projets de recherche et de défense que nous ne connaissons pas, et dans lesquels l'intelligence artificielle autonome est également présente. Certes, il ne s'agit pas de la meilleure façon de répondre positivement aux peurs, mais je pense qu'il y a un certain nombre de gens qui souhaitent, pour une série de raisons, développer l'intelligence artificielle autonome. Il ne faut, par conséquent, pas en nier l'existence.

**M. Patrick Albert, PDG de SuccessionWeb.** - J'ai une question qui me semble fondamentale à l'intention des parlementaires ici présents. Quels sont les objectifs que nous pouvons nous donner en tant que pays ? Alors que le titre de la session est « *dimensions stratégiques en matière d'intelligence artificielle* », nous voyons que trois intervenants représentent des multinationales basées aux États-Unis. Certes nous pouvons nous réjouir qu'ils viennent en France travailler avec nous, mais il est tout de même inquiétant que ce soient eux qui nous conseillent, surtout à l'heure où il s'agit de rendre à nouveau grande l'Amérique. Les brevets déposés sont largement américains. Je pense par conséquent que nous devons travailler à définir nos objectifs nationaux.

**Mme Dominique Gillot, rapporteure.** - Merci pour cette question qui resitue le sujet en son cœur.

**M. Yves Demazeau, président de l'association française pour l'intelligence artificielle (AFIA).** - Je souhaite signaler que nous organiserons lundi prochain avec le MEDEF une journée « *Intelligence artificielle et entreprises de France* », au cours de laquelle seront présents des

---

groupes tels que Renault, Veolia ou Dassault pour évoquer l'impact de l'intelligence artificielle sur leurs activités.

**M. Jean-Claude Heudin, directeur de l'Institut de l'Internet et du Multimédia.** - Nous disposons aujourd'hui d'outils de financement de la recherche mais je tiens à insister sur la complexité des dossiers, qui représente un vrai frein à la compétitivité. On passe plus de temps à monter les dossiers et dans les contrôles que dans le travail de recherche lui-même. Aujourd'hui les *start-ups* ne peuvent pas se permettre de dépenser autant d'énergie à travailler sur les dossiers. Le financement obtenu avec des mois de retard représente un véritable problème.

**Mme Dominique Gillot, rapporteure.** - Ces travers ont été largement commentés mais on assiste à une simplification des procédures, le futur livre blanc de l'enseignement supérieur et de la recherche l'illustrera.

**Mme Amal Taleb, vice-présidente du Conseil National du Numérique.** - Je travaille pour SAP et je suis membre du Conseil National du Numérique. Je souhaite partager une question sous forme de constat avec vous, sur laquelle nous devrions réfléchir collectivement. L'intelligence artificielle doit investir le tissu industriel classique. La France a la chance de posséder un beau tissu industriel. L'Allemagne a récemment fait le même constat lors de la conférence franco-allemande. Pour autant, la représentation nationale ne pourra faire l'économie de la manière dont l'intelligence artificielle va intervenir directement sur les *process* de production *B to B* me semble-t-il, en mettant les entreprises, les grandes comme les petites, dans la discussion. Il faut de l'intelligence collective et garder à l'esprit le fait que nous sommes engagés dans une industrie du futur, dans laquelle l'intelligence artificielle devra être prise en compte de façon importante. Certes nous avons la culture *B to C* de l'usage, des consommateurs et des usagers, mais il faudra de manière beaucoup plus globale introduire en mode *pull* l'intelligence artificielle comme un moyen de gagner des parts de marché.

**Mme Dominique Gillot, rapporteure.** - Différentes stratégies ont été mises au point par des comités qui ont réfléchi longuement aux points que vous évoquez. Sur la stratégie nationale de la recherche, les mesures prises il y a un an sont déjà rediscutées eu égard à la rapidité avec laquelle les sujets et les questions prégnantes émergent. Effectivement, nous devons mener un débat entre les représentants de la recherche fondamentale académique et cette poussée très forte des jeunes *start-uppers*, qui réinvestissent des connaissances acquises il y a parfois longtemps et mises à jour de manière autodidacte. De jeunes diplômés tentent également l'expérience de la *start-up* sans choisir la recherche fondamentale. Nous ne pouvons pas non plus exiger que les choses s'effectuent de manière instantanée. L'ensemble de vos remarques sont essentielles et seront consignées dans notre rapport. Un bouleversement devra manifestement être opéré dans notre société sur la transformation de la connaissance en

innovation, tout en n'omettant pas de tenir compte des impératifs de la décision politique, qui doit être prudente. Nous sommes en France.

**Mme Amal Taleb, vice-présidente du Conseil National du Numérique.** – Effectivement vous avez parfaitement raison, il nous faut résoudre ce hiatus. L'un des moyens est d'investir le tissu industriel beaucoup plus classique. Il se trouve que nous avons beaucoup de chance, puisque que nous avons des logiques et des réflexions similaires à celles de l'Allemagne. De manière différente, ce pays intègre les entreprises dans son tissu. Il nous faut impérativement inclure les grands groupes, les PME plus classiques et les *start-ups* afin de créer cet écosystème qui permettra de développer l'intelligence artificielle et sans doute, de faire jouer les intelligences collectives. Ces associations pourraient également nous permettre de dépasser les blocages financiers.

**M. Gilles Dowek, directeur de recherche à l'INRIA et professeur attaché à l'ENS de Paris-Saclay.** - La difficulté à trouver des financements ne vaut pas seulement pour le secteur privé, mais également pour des organismes publics tels que le CEA.

**M. Jean Ponce, professeur à l'École Normale Supérieure.** - Nous n'avons pas, à l'ENS ou à l'INRIA, de réelles difficultés à trouver des financements de la part d'industriels car nous en rencontrons chaque semaine. Nous sommes privilégiés à ce niveau. Les réseaux de chercheurs (RC) sont aussi un bel outil à cet égard. En revanche, j'explique à ces partenaires que nous sommes une petite équipe en recherche fondamentale, qui ne dispose pas du temps et des compétences nécessaires au travail d'encadrement. Le statut de chercheur n'est pas fait pour cela, et pour ma part je n'aurais pas le temps d'encadrer un trop grand nombre de post-doctorants, de l'ordre de la dizaine par exemple.

**Mme Dominique Gillot, rapporteure.** – Nous avons évoqué le plan de carrière d'un chercheur et les différentes étapes à passer, qui nécessitent des allers-retours. Votre questionnement a été abordé à travers la stratégie nationale de l'enseignement supérieur : comment valoriser les parcours itératifs d'un chercheur dans le privé, le public, la recherche et la formation ? Aujourd'hui nous n'avons pas les bons outils.

**M. Jean Ponce, professeur à l'École Normale Supérieure.** – Aux États-Unis, des collaborateurs ont pour mission de travailler avec les industriels et de faire le pont avec les chercheurs.

**M. Fabien Moutarde, professeur au centre de robotique de Mines ParisTech.** – Il faut faire évoluer les statuts des chercheurs pour permettre davantage d'allers-retours entre la recherche privée et la recherche publique. J'ai commencé ma carrière chez Alcatel, mais j'ai eu du mal à revenir dans le public.

Sur la partie stratégique, mon sentiment est qu'il est nécessaire de pousser au maximum, par des plans de financement de projets de recherche, la robotisation et l'automatisation de tout ce qui peut l'être. Il s'agit en effet

---

d'un enjeu économique et stratégique pour la France et pour l'Europe. Or une censure a existé sur le sujet par crainte du chômage. Pourtant, l'absence d'automatisation aboutira malgré tout à la perte des emplois que l'on souhaite protéger. Des complémentarités saines peuvent être trouvées entre la robotique et l'humain, pour confier aux hommes les tâches ayant davantage de valeur ajoutée. On peut aussi aller vers un revenu citoyen et conduire à des relocalisations.

**Mme Dominique Gillot, rapporteure.** – Les sciences dures et les sciences sociales doivent davantage travailler ensemble. Il semble en effet souhaitable de supprimer les tâches répétitives et difficiles, pouvant entraîner à la longue un handicap. Il est vrai que dans le débat actuel de la société, cette dimension du bénéfique pour les personnes dont les métiers vont être supprimés n'est jamais abordée. C'est pourquoi le sujet de l'éducation est essentiel.

**M. Igor Carron, Paris Machine Learning Meetup.** – En France nous avons un problème avec les données. Alors que certains organismes parviennent à accumuler un grand nombre de données, ils sont immédiatement confrontés aux limites de leur utilisation. Au ministère de l'Éducation, il ne semble pas que des recherches soient menées sur le développement et l'éducation afin d'aboutir à une meilleure efficacité, aucun centre ne collecte les données pour améliorer les politiques éducatives à ma connaissance. Facebook et Google n'étaient que des *start-ups* il y a encore dix-quinze ans, mais se sont développées grâce à l'intelligence artificielle. Pour ma part, j'estime hautement souhaitable que les données de l'Éducation Nationale puissent être utilisées par les chercheurs afin de faire progresser la science de l'éducation. Il nous manque un centre d'expertise à ce sujet.

**Mme Laurence Devillers, professeur à l'Université Paris-Sorbonne/LMSI-CNRS.** – Je souhaite intervenir sur la problématique de l'éthique, au sujet de laquelle les intervenants de Facebook et Google ont mentionné qu'une réflexion était en cours au sein d'un partenariat entre entreprises. Existe-t-il des moyens de mener cette réflexion en collaboration avec les chercheurs ? Comment allez-vous vous ouvrir alors qu'on constate une certaine opacité ? Les impacts d'une telle réflexion porteront notamment sur l'éducation et la protection des données.

**M. François Taddei, directeur du Centre de recherche interdisciplinaire.** – Aujourd'hui, les traces numériques des apprenants ne sont pas disponibles pour faire progresser l'ensemble du système. Pourtant, ces données pourraient être un miroir pour l'enseignant, le chef d'établissement et les parents. Si nous pouvions passer des *data* à l'information, nous progresserions grandement. T. S. Eliot se demandait où est passée la sagesse que nous avons perdue dans la connaissance et où est passée la connaissance que nous avons perdue dans l'information... Nous avons besoin de créer des centres de recherche pour avancer sur ces sujets. La CNIL et l'INRIA dialoguent à ce sujet. Il faut mettre en place des plateformes sécurisées en termes de vie privée, mais avec une vraie valeur

---

collective. Le *Big Data* pourrait par exemple proposer des solutions pour résoudre les difficultés d'apprentissage. L'INRIA s'est beaucoup rapprochée de l'INSERM pour travailler dans le domaine de la recherche en éducation, mais nous ne disposons d'aucune possibilité d'analyser les *data*. Nous savons qu'on consacre 2,24 % de la richesse nationale à la R&D mais nous ne savons pas combien nous consacrons à l'éducation.

**Mme Dominique Gillot, rapporteure.** – Vous soulevez de vraies questions. La recherche en éducation doit se faire à l'université.

**M. François Taddei, directeur du Centre de recherche interdisciplinaire.** – Cependant, la capacité des enseignants sur l'intelligence artificielle et le numérique est très limitée.

**Mme Dominique Gillot, rapporteure.** – Elle est très réduite, en effet. Il faut agir sur les structures de formation des enseignants, en particulier les écoles supérieures du professorat et de l'éducation (ESPE), qui ont remplacé les instituts universitaires de formation des maîtres (IUFM). La définition des maquettes est rigidifiée par les anciennes pratiques. J'ai assisté à un débat absolument incroyable au cours duquel on expliquait qu'un poste réservé à un mathématicien ne pouvait pas être transformé en un poste d'enseignant en informatique. Les carrières sont, à l'évidence, beaucoup trop rigides. À l'Université de Cergy-Pontoise, une recherche est en cours sur l'utilisation des outils de l'intelligence artificielle pour observer comment les enfants fixent leur attention. La difficulté actuelle tient à la résistance des parents à ce que leurs enfants soient l'objet de telles expériences, car l'anonymisation n'est pas au rendez-vous. Il faut donc obtenir une réassurance en la matière.

**M. François Taddei, directeur du Centre de recherche interdisciplinaire.** – Il faut s'assurer que les progrès bénéficieront à chacun des acteurs, y compris aux premiers intéressés. Si les parents ont l'impression que les progrès bénéficient à certains seulement, nous n'y arriverons pas.

**M. Gilles Dowek, directeur de recherche à l'INRIA et professeur attaché à l'ENS de Paris-Saclay.** – Aujourd'hui, le ministère de l'Éducation nationale ne comprend rien aux enjeux qu'il y a à former les jeunes sur l'informatique et l'intelligence artificielle. Je donnerai un exemple sur ce point. Il est coutume de dire que pour enseigner un sujet, quatre éléments sont nécessaires : un élève, un enseignant, du temps à consacrer à cette activité, et un programme. Aujourd'hui, le Conseil supérieur des programmes a fait un bon travail pour élaborer des contenus à enseigner en informatique au collège, mais il manque des enseignants. Or le ministère de l'Éducation nationale ne comprend pas la nécessité d'avoir des enseignants en informatique. Le raisonnement au sein de ce ministère est d'affirmer que l'informatique étant une discipline relativement facile, les jeunes possèdent la matière dans leur ADN. De ce fait, tant que nous resterons arc-boutés sur ce type de conceptions, nous ne pourrons pas avoir un public informé de ce que les algorithmes sont censés faire.



---

**M. Jean-Gabriel Ganascia, professeur à l'Université Pierre et Marie Curie-Paris VI.** – Dans tous les métiers, le numérique est central. Il faut par conséquent lancer des formations mixtes, car c'est nécessaire pour le secondaire comme pour le supérieur. Pour notre part, nous avons élaboré une formation lettres-informatique, car il faut nous préparer à de nouveaux métiers. Nous avons parlé du travail et des risques que la robotisation faisait courir à certains métiers, mais il est possible de compenser les pertes d'emplois par de nouveaux métiers. Pour préparer les personnes à ces métiers, la formation est tout à fait essentielle. À cela, il faut ajouter que la formation dont il s'agit n'est pas nécessairement initiale. La formation tout au long de la vie doit être développée, en transformant les institutions supérieures et les écoles pour parvenir à ce résultat. Il faut également revenir sur le travail, pour lequel la comptabilisation en « temps passé à l'établi » n'a plus aucun sens. Le temps de présence au bureau c'est obsolète. Il est indispensable d'ajouter le temps de formation au temps de travail, étant observé que l'idée de trente-cinq heures passées au travail est tout à fait désuète. Je pense que nous aurons du mal à faire passer ces idées.

Pour rejoindre les propos de Gilles Dowek sur l'administration de l'Éducation nationale, je confirme que le ministère a du mal à opérer ces transformations. J'avais moi-même un projet de sciences participatives des enseignants dans les domaines littéraires, mais ce projet a été mis sur la touche faute de pouvoir être mis en contact direct avec des enseignants par un directeur académique au numérique, qui était aussi peu compétent qu'indisponible.

Enfin, sur les questions d'éthique et d'intelligence artificielle, nous organiserons avec l'AFIA le 5 juillet 2017 une journée qui leur est consacrée.

**Mme Dominique Gillot, rapporteure.** – Juste un petit rappel sur le compte personnel d'activité (CPA), preuve que les parlementaires ont malgré tout fait avancer les choses en matière de formation. Le CPA a été mis en place au 1<sup>er</sup> janvier 2017 pour associer formation et travail. Nous voulons également confier à l'université la formation tout au long de la vie. La loi doit être appliquée, mais elle met du temps pour ce faire.

**M. Gilles Babinet, entrepreneur, *digital champion* auprès de la Commission européenne.** – Nous avons des approches très holistiques, comme en témoignent les exemples de nos réflexions sur l'éthique ou sur l'éducation. Cela nous ramène à l'un des problèmes dont souffre notre pays, qui est celui de la parole politique, handicapée par l'incapacité du corps politique à se projeter dans un futur humaniste, et qui respecte nos valeurs. Il faut tracer ce futur. Trop souvent, les débats tiennent à des enjeux du XX<sup>e</sup> siècle. Or, la révolution digitale ne peut plus nous faire penser que nos sociétés fonctionneront comme au siècle dernier.

**Mme Dominique Gillot, rapporteure.** – L'OPECST a été créé dans ce but entre autres. Je voudrais que Google réponde aux questions qui ont été posées concernant le partenariat sur l'intelligence artificielle. Pourra-t-on faire des choses ensemble ?

---

**M. Olivier Esper, responsable des affaires publiques de Google France.** – Je réponds positivement à la question de Laurence Devillers. Les uns et les autres, avons bien souligné nos modèles ouverts de recherche, qui ont évolué dans ce sens dans la période récente. Le partenariat sur l'intelligence artificielle a bien évidemment pour objet de regrouper la diversité. Son conseil d'administration n'est pas encore finalisé mais des Européens dont Yann LeCun, français, en font déjà partie. L'idée est de mettre en œuvre une réflexion collective.

**Mme Dominique Gillot, rapporteure.** – L'ouverture est une bonne chose mais il s'agit également de travailler avec des partenaires non liés à vous par contrat.

**M. Olivier Esper, responsable des affaires publiques de Google France.** – Sur le site du partenariat sur l'intelligence artificielle figure déjà la possibilité de manifester son intérêt à participer à la réflexion.

**Mme Delphine Reyre, directrice Europe des affaires publiques de Facebook.** – Nous nous engageons auprès de vous à faire part de l'intérêt fort exprimé à être parties à la réflexion. Nous pourrions ensuite imaginer les formes de cette discussion.

**Mme Laurence Devillers, professeur à l'Université Paris-Sorbonne/LMSI-CNRS.** – Je réfléchis avec plusieurs collègues sur la robotique et l'éthique dans le cadre de la CERNA. Nous avons donc des sujets de réflexion qui pourraient intéresser les grands groupes. Nous vous transmettrons nos rapports, mais nous pourrions également imaginer autre chose dans une démarche participative avec d'autres groupes industriels. La réflexion sur l'intelligence artificielle doit être menée de manière collégiale en y intégrant les dimensions éthiques. Je suis impliquée dans la société savante IEEE, qui réunit des centaines de chercheurs du monde entier pour travailler sur des standards à mettre en place ensemble. Il est essentiel de ne pas opérer de façon isolée sur des sujets tels que ceux-ci, qui concernent notre futur. C'est pourquoi nos gouvernements devraient s'emparer de cette question.

#### IV. TROISIÈME TABLE RONDE, PRÉSIDIÉE PAR MME DOMINIQUE GILLOT, RAPPORTEURE : QUESTIONS POLITIQUES, SOCIÉTALES ET ÉCONOMIQUES LIÉES À L'IRRUPTION DES TECHNOLOGIES DE L'INTELLIGENCE ARTIFICIELLE

**Mme Dominique Gillot, rapporteure.** – Nos deux tables rondes de ce matin, qui ont donné lieu à des échanges intéressants, ont permis de cerner ce que sont les systèmes d'intelligence artificielle, dans leur diversité.

Cette troisième table ronde aborde les questions politiques, sociétales et économiques liées à l'irruption des technologies de l'intelligence artificielle. Axelle Lemaire, secrétaire d'État chargée du numérique et de l'innovation, nous rejoindra à seize heures.

Notre calendrier est contraint, mais nous souhaitons tous aller au bout de nos réflexions avant les prochaines échéances électorales.

Monsieur Henri Verdier, vous êtes directeur interministériel du numérique et du système d'information de l'État, adjoint à la secrétaire générale pour la modernisation de l'action publique et administrateur général des données. Vous avez dirigé Etalab, le service chargé de l'ouverture des données publiques, grâce auquel vous avez modernisé le portail d'*open-data* français. Votre regard sur les questions politiques, sociétales et économiques liées à l'irruption des technologies d'intelligence artificielle sera utile pour ouvrir cette table ronde.

##### 1. M. Henri Verdier, directeur interministériel du numérique

L'intelligence artificielle est une vieille histoire qui se renouvelle actuellement à une allure incroyable. Se sont télescopées de manière imprévue une science, une masse de données et une informatique, venue en particulier des jeux vidéo, avec des investissements considérables de la part des acteurs qui s'en servent : Google, IBM, Facebook... Il s'agit d'investissements qui atteignent des niveaux inédits.

Ces acteurs s'inscrivent dans des stratégies post-numériques ouvertes et atteignent d'amples performances. Nous avons récemment appris, par exemple, que l'intelligence artificielle reconnaissait mieux les images que l'œil et le cerveau humain !

Certains orateurs ont assimilé cet enjeu à celui qu'a représenté la course à l'arme nucléaire. La métaphore a ses limites : il ne s'agit pas forcément de fabriquer des intelligences artificielles, mais il s'agit d'en avoir

---

la maîtrise. Certains pays maîtriseront cette technologie, d'autres non, ce qui suscitera un clivage géopolitique.

Pour l'État, le premier enjeu est que notre pays ne devienne pas, comme le craint **Mme Catherine Morin-Desailly**, une colonie numérique des États-Unis. En ce qui concerne le *Big Data*, la France n'a que deux ou trois ans de retard : il est donc encore temps de démarrer. La bataille n'est pas perdue. Notre pays peut compter sur une très grande école de mathématiques, de très grandes écoles d'ingénieur, une culture de l'entrepreneuriat, des industriels de renom, un État centralisé qui pourrait devenir un client. Il faut s'y mettre, avec une politique industrielle moderne. Je relève des initiatives très positives : le travail de l'OPECST sur l'intelligence artificielle, le plan gouvernemental pour l'intelligence artificielle, etc.

Des questions de sécurité se poseront. Je ne suis pas informaticien de formation, mais biologiste. J'ai une vague intuition des propriétés émergentes des systèmes. Dès qu'un système est complexe, son évolution est imprévisible. J'ai beaucoup milité en faveur de la publication des algorithmes. Cependant, l'intelligence artificielle fonctionne comme une boîte noire. Il ne suffit pas de divulguer le code source pour que chacun se fasse une opinion. La Silicon Valley met en place des *Kill Switch*, c'est-à-dire des dispositifs qui se désactivent très facilement.

En matière économique, on se prépare à de grandes mutations, de nouveaux cycles de destruction incroyables, de reconversion et de recombinaison de filières industrielles entières. Le marché du travail sera profondément transformé. Peut-être nous sera-t-il possible d'accompagner ces changements dès lors que cette révolution est prise à la racine.

L'État sera-t-il capable de s'approprier ces nouvelles technologies pour assurer l'efficacité, la simplicité et la justice du service public ? Pourquoi attendre un mois pour savoir si l'on a obtenu une bourse alors que l'on sait, lorsque l'on commande la moindre bricole sur Amazon, à quel moment elle se trouve mise dans la camionnette du facteur ?

Nous n'avons pas une très grande science du *design* des politiques publiques. Souvent, elles ont été construites de manière empirique.

J'ai, par ailleurs, été frappé de constater qu'il était très facile de faire des erreurs très idiotes avec des algorithmes. Par exemple, il est très simple de mettre au point un algorithme permettant de mener des contrôles fiscaux avec une efficacité de 90 %. Règle-t-on le problème ou fait-on courir des risques encore plus grands ? En effet, cet algorithme peut omettre une catégorie entière de fraudes, ce qui n'est pas souhaitable. Il faut donc bien être sûr, lorsque l'on construit un algorithme, qu'il est totalement adapté à la politique que l'on souhaite mettre en place. Or, il est souvent très difficile d'explicitier le but profond d'une politique publique.

On l'a vu avec le *Big Data*, il est assez aisé, faute d'une profonde réflexion, de faire faire des bêtises à des machines qui ne commettent pourtant aucune erreur. Les États-Unis travaillent par exemple à la mise au

---

point d'algorithmes pour aider les juges à prononcer des remises en liberté conditionnelle, mais on y a introduit des statistiques pleines de préjugés ethniques, ce qui conduit à une mauvaise prédictibilité. Les algorithmes donneront donc des résultats à la hauteur de l'éducation qui leur est fournie !

Nous avons le devoir de nous approprier ces nouvelles technologies, mais nous devons aussi nous demander où nous voulons aller. Cela nous amènera assez vite à la question de la « gouvernamentalité algorithmique ». Il est utile de pouvoir prédire grâce aux algorithmes, mais cela comporte des risques. Notamment, celui de la confiscation du pouvoir par des experts. Il est important de faire entrer les vrais débats sur l'intelligence artificielle au Parlement et chez les citoyens. Les algorithmes apprennent et donnent des résultats mais ce sont des boîtes noires. Le besoin de transparence est très grand.

L'invention d'une redevabilité et d'une gouvernance de l'intelligence artificielle est un immense chantier. Si nous voulons rester une démocratie et utiliser à plein ce que proposent ces technologies, il nous faudra faire le lien entre projet politique et progrès technique.

Concernant le plan du Gouvernement pour l'intelligence artificielle, nous avons quelques éléments mais je préfère laisser la ministre en parler.

**Mme Dominique Gillot, rapporteure.** – Merci M. Verdier ! Vous présidez, Madame Marie-Claire Carrère-Gée, le conseil d'orientation pour l'emploi (COE), qui vient tout juste de rendre public un rapport important sur les conséquences de l'automatisation et de la robotisation sur l'emploi. Je vous laisse la parole.

## **2. Mme Marie-Claire Carrère-Gée, présidente du Conseil d'orientation pour l'emploi**

Le conseil d'orientation pour l'emploi est une instance indépendante, placée auprès du Premier ministre. Nous essayons d'élaborer des diagnostics partagés sur les questions d'emploi et de travail. C'est un domaine caractérisé par la pluralité des acteurs et dans lequel le consensus n'est pas toujours spontané.

L'ensemble des acteurs du marché du travail, comme l'ensemble de nos concitoyens, sont confrontés à l'évolution des nouvelles technologies, qui connaissent un extraordinaire développement et dont la vitesse de diffusion est très incertaine. Je précise que je n'ai pas de compétences en intelligence artificielle.

Notre rôle est d'anticiper, autant que faire se peut, les conséquences du progrès technologique en cours sur l'emploi, dans un contexte où le débat public est marqué par des études donnant au pourcentage près le nombre d'emplois menacés et conduisant à un climat très anxieux. Certes, la

---

crainte du chômage est un grand classique à chaque vague d'innovation technologique. Keynes lui-même avait prédit un chômage massif lié au progrès technique. Pourtant, l'histoire montre que, depuis toujours, le progrès technologique s'est accompagné de créations d'emplois.

Malgré cela, des études prédisent une catastrophe à un horizon proche. Je peux citer entre autres l'étude de l'université d'Oxford conduite par Carl Benedikt Frey et Michael A. Osborne qui parle de 47 % des emplois, celle du centre de réflexion Bruegel avec une part des emplois concernés de 52 % et, enfin, celle du cabinet de conseil Roland Berger, spécifique à la France, avec une part de 42 %. D'autres études sont moins pessimistes, avec un chiffre de 9 %, comme le rapport de l'OCDE réalisé par des chercheurs de l'université de Mannheim ou, encore, la note que Nicolas Le Ru a réalisée pour France Stratégie en 2016.

Les analyses existantes sont partielles. Elles s'intéressent uniquement aux destructions brutes d'emplois, et ne font pas la balance entre les destructions et les créations. Elles ne s'attachent ni à la structure de l'emploi ni à sa localisation. Parfois même, elles sont biaisées en ce qu'elles considèrent que la technologie bouscule des métiers dans leur ensemble alors que, de notre point de vue, elle bouleverse d'abord des tâches. Au sein d'un métier, certaines tâches peuvent rester très difficiles, voire impossibles, à remplacer par des technologies. Toutes les personnes qui exercent un métier ne font par ailleurs pas la même chose tous les jours. Le rapport de l'OCDE s'intéresse aux tâches, ce qui est plus pertinent. Les emplois évoluent, y compris dans leur contenu.

Dans notre rapport, nous avons voulu aborder l'ensemble des dimensions – créations, destructions, transformations d'emploi, structures et localisation de l'emploi – et élargir l'approche, non seulement aux emplois susceptibles d'être détruits, mais aussi à ceux qui peuvent évoluer dans leur contenu. Ce constat doit conduire à définir des politiques de formation adaptées. Les études rétrospectives sur la composition de l'emploi montrent d'ailleurs que de nombreux emplois, en France, ont profondément évolué depuis quinze ans dans leur contenu en lien avec la diffusion des nouvelles technologies, quel que soit le niveau de qualification. L'utilisation de technologies numériques est associée à une complexification et une hybridation des compétences requises. Avec l'évolution numérique, un certain nombre de métiers se sont même scindés en plusieurs.

Dans ce cadre, nous avons réalisé une étude quantitative pour préciser le diagnostic, pour savoir quels seraient les métiers concernés. Nous avons émis l'hypothèse que la vulnérabilité des emplois dépend de trois facteurs principaux : l'avantage comparatif de l'homme par rapport aux machines ; les goulets d'étranglement de la frontière technologique ; enfin, la rentabilité économique. Difficile à introduire de façon robuste, ce dernier facteur n'est pas pris en compte dans notre étude, pas plus d'ailleurs que dans les autres études existantes.

---

Nous avons utilisé l'enquête sur les Conditions de travail de la DARES. Nous avons isolé, selon les types de métier, jusqu'à une vingtaine de questions. Nous avons essayé d'apprécier le niveau de vulnérabilité des emplois au regard de leurs conditions concrètes de réalisation. Nous nous sommes refusé à donner un chiffre précis, car il n'existe aucun seuil à partir duquel un emploi serait automatiquement automatisé. Nous avons préféré retenir des ordres de grandeur.

Nous sommes parvenus à trois enseignements principaux. Tout d'abord, moins de 10 % des emplois cumulent des vulnérabilités telles que, dans un contexte d'automatisation, leur maintien serait menacé. Nous avons retenu un seuil d'automatisation de 0,7 : à noter que les résultats auraient potentiellement été un peu différents, avec un seuil de 0,65 par exemple. Les métiers les plus vulnérables sont essentiellement peu ou pas qualifiés, nous en avons listé une vingtaine : agents d'entretien, ouvriers qualifiés ou non, manutentionnaires, jardiniers, maraîchers, caissiers, employés de service...

Ensuite, la moitié des emplois seront profondément, à plus ou moins brève échéance, transformés dans leur contenu, quel que soit le niveau d'automatisation. Cela touche les moins qualifiés, mais aussi des emplois plus qualifiés, dans l'industrie mais aussi dans les services.

Enfin, l'étude quantitative n'était pas le seul sujet de notre rapport. Nous avons pris acte du fait que l'histoire économique montre que l'emploi a continué à augmenter, y compris au cours des vingt dernières années. L'enjeu est la gestion de la période de transition. Nous avons étudié les questions de création d'emplois : ceux induits directement par l'innovation technologique, mais aussi ceux indirectement créés. Même une simple innovation de procédé, si elle fait gagner des parts de marchés à l'international, peut créer de l'emploi. Il y a aussi des mécanismes de compensation : l'évolution de la demande crée des emplois induits.

Une politique publique d'accompagnement doit viser à diminuer le nombre d'emplois détruits et faire en sorte que les mécanismes de compensation se produisent le plus tôt possible et jouent à plein pour éviter les décalages entre destructions et créations d'emplois. Le progrès technologique n'est pas une option que l'on pourrait ou non choisir ; on ne peut l'éviter, et la vitesse d'adaptation est essentielle. Les acteurs économiques et les pouvoirs publics ont leur mot à dire en termes de choix éthiques, sociaux, économiques, fiscaux. Je vous renvoie aux conclusions de notre rapport. Si la première vague du numérique a contribué à délocaliser des emplois existants, voire des fonctions entières dans les entreprises, la vague en cours pourrait, *via* l'automatisation, favoriser des relocalisations d'emplois. C'est un aspect qui doit être pris en compte.

**Mme Dominique Gillot, rapporteure.** - Chirurgien neurologue, ancien élève de Science Po, de HEC, de l'ENA, écrivain, auteur d'ouvrages à succès, chroniqueur pour plusieurs journaux, vous êtes aussi, monsieur Alexandre, le fondateur du site Doctissimo, très consulté, et vous dirigez

---

aujourd'hui l'entreprise DNA vision. L'intelligence artificielle est un sujet qui vous passionne. Nous vous écoutons.

### 3. M. Laurent Alexandre, entrepreneur (DNA vision)

On parle souvent des dangers de l'intelligence artificielle forte, mais les vrais sujets sont la gestion, la gouvernance et la régulation d'une intelligence artificielle faible, sans conscience d'elle-même et pas directement dangereuse pour l'humanité.

Je souhaite insister sur la souverainement numérique et sur le risque de devenir une colonie numérique des GAFAs (Google, Apple, Facebook et Amazon). Il existe un trou noir de valeur en faveur des géants du numérique qui maîtrisent l'intelligence artificielle. En la matière, nous sommes un pays du tiers-monde. Nous exportons des spécialistes, des cerveaux, et nous importons toute la journée de l'intelligence artificielle : soixante-cinq millions de Français importent de l'intelligence artificielle cent quatre-vingt fois par jour sur leur téléphone ! Autant dire que nous avons perdu notre souveraineté numérique.

Nous sommes à un tournant de l'humanité : tel est le slogan du patron d'Intel. Le patron de Google, Sergei Brin, affirme, quant à lui, qu'il va construire des machines qui raisonnent, pensent et feront les choses mieux que l'homme. Ça c'est de l'intelligence artificielle forte. Elon Musk, le patron de Tesla, expliquait le 2 juin dernier, dans un grand élan lyrique, qu'il était urgent de mettre des microprocesseurs dans nos cerveaux avant que l'intelligence artificielle ne nous transforme en animaux domestiques !

Cependant, nous ne sommes pas à la veille d'un HAL 9000 qui signerait la fin de l'humanité, comme dans « *2001, l'odyssée de l'espace* » ! Les enjeux actuels sont des enjeux sociétaux, des enjeux de souveraineté, des enjeux économiques, liés au fait que nous sommes nuls en matière d'intelligence artificielle.

Mais l'intelligence artificielle n'est pas seule. Les GAFAs et les BATX (Baidu, Alibaba, Tencent et Xiaomi) ont éclos. L'idéologie transhumaniste a fait pousser des pulsions démiurgiques, notamment dans les différentes Silicon Valleys ; ce que l'on appelle les NBIC (nanotechnologies, biotechnologies, informatique et sciences cognitives) arrivent à maturité ; la zone Asie-Pacifique est plus transgressive que nous.

Les NBIC sont des technologies exponentielles, donc très imprévisibles. L'intelligence artificielle va être gratuite en valeur relative face à l'intelligence biologique. Cela a des conséquences majeures : quand un bien est gratuit, les substituts sont balayés et les complémentaires sont renchérissés. Or, le travail peu qualifié est substituable, et le travail qualifié, complémentaire. La micro-économie rend ce verdict incontournable.



---

Si je parle de pulsions démiurgiques, c'est que la zone Asie-Pacifique se croit tout permis sur le plan biotechnologique, mais aussi sur le plan de l'intelligence artificielle. Une bonne gouvernance et une régulation internationale seront nécessaires.

Autre phénomène, la « plateformisation » : l'intelligence artificielle sort au robinet des grandes plateformes pour des raisons techniques, car il faut beaucoup de données pour la développer. Une étude récente a montré qu'un mauvais algorithme avec beaucoup de données est supérieur à un bon algorithme avec peu de données. Or les plateformes ont des milliards d'utilisateurs, il n'y a aucun équivalent en Europe, nous n'avons pas de base industrielle comparable. Bref, nous sommes tous les idiots utiles de l'intelligence artificielle, pour paraphraser Lénine qui disait que les bourgeois de gauche sont les idiots utiles de la Révolution. En effet, nous mettons chaque jour dix milliards de photos gratuitement sur la nébuleuse Facebook, Instagram, Messenger et WhatsApp afin d'améliorer l'efficacité des intelligences artificielles des plateformes.

Les technologies NBIC se développent, mais pas toutes au même rythme exponentiel : la biologie et la robotique, c'est lent. La robotique ne fera pas beaucoup bouger le marché du travail. L'intelligence artificielle sans robot va croître exponentiellement et modifiera radicalement des aspects importants du travail. En revanche, l'électronique et l'intelligence artificielle sans support mécanique et biologique vont croître beaucoup plus vite. Le COE sous-estime ce poids croissant de l'intelligence artificielle.

Allons-nous devenir une colonie numérique des GAFA ? Au-delà de l'État, la société civile et l'ensemble du tissu économique doivent comprendre que si nous n'avons pas de plateformes à l'échelle des Chinois et des Américains, nous serons le Zimbabwe de 2080 ! Nous sommes en train de perdre nos ressorts économiques. Il n'y a pas en Europe une miette d'intelligence artificielle et de technologies NBIC, qui constituent pourtant la base fondamentale de création de valeur ! Nous avons une responsabilité historique.

Il s'est créé un trou noir de valeur en faveur des algorithmes et de l'intelligence artificielle, avec un changement radical dans la structure de valeur. Traditionnellement, on avait des *business* verticalisés. Nissan produit des voitures, Danone des yaourts, Gaumont du cinéma et Badoit de l'eau. Aujourd'hui, nous assistons à une horizontalisation des métiers. On voit arriver des conglomérats de l'intelligence artificielle, qui font plusieurs métiers. Baidu, le Google chinois, se lance dans la voiture autonome. L'intelligence artificielle produit cette transversalité. Les cinquante-cinq petits génies de WhatsApp ont créé en quatre ans 23 milliards de dollars de valeur, soit deux fois plus que les cent trente mille ouvriers de PSA depuis un siècle. La paupérisation relative de notre population serait une certitude si nous ne prenions pas le tournant de l'intelligence artificielle !

L'intelligence artificielle induit un changement radical dans l'organisation de la fonction de production. On voit apparaître une part

---

nouvelle de capital, l'intelligence artificielle, un déclin fort du travail et une montée en flèche de la valeur du travail très qualifié, qui est en réalité du capital, payé en actions et en stock-options. L'intelligence artificielle déstabilise le rapport relatif entre le travail et le capital, avec une nouvelle forme de capital à coût nul face au cerveau biologique. Cela ne veut pas dire que le travail va disparaître. En tant que schumpétérien de centre-gauche, je suis même persuadé que nous allons inventer de nouveaux métiers.

Quelles tâches pour quels hommes ? C'est un vrai sujet de réflexion. L'école de la République en France est extrêmement en retard, ce qui n'est pas le cas à Singapour, au Vietnam ou en Corée. La complémentarité neurone-transistor nous semble naturelle, mais c'est en réalité un combat. En 2050, 100 % de ceux qui ne seront pas complémentaires de l'intelligence artificielle seront soit au chômage, soit dans un emploi aidé ! Il est donc important de rénover progressivement l'école et la formation professionnelle. Ayant visité récemment des centres pour apprentis, je puis vous garantir que nous n'en prenons pas le chemin. Notre responsabilité politique est très importante, je ne suis pas sûr que nous nous en rendions bien compte. Nous vivons déjà des crises populistes, ça ne peut que s'aggraver.

Le défi posé par la complémentarité est le suivant : un robot coûte très cher, ce qui n'est pas le cas de l'intelligence artificielle, qui, de surcroît, progresse plus vite ! En outre, les robots remplacent plutôt des métiers peu qualifiés, alors que l'intelligence artificielle non fondée sur des robots peut remplacer des métiers très qualifiés. Je partage l'opinion de Yann LeCun, le patron d'une des deux branches de l'intelligence artificielle chez Facebook : l'intelligence artificielle aura dépassé les meilleurs radiologues - je me permets de le dire en tant que chirurgien, et ce n'est pas une vendetta corporatiste - avant 2030. Se pose donc un problème de reconversion qui dépasse celui des chauffeurs routiers et des emplois non qualifiés : l'intelligence artificielle rendra également nécessaire une réorientation de métiers traditionnels à haut contenu cognitif. D'ailleurs, le lien entre qualification des emplois et risque d'automatisation est beaucoup plus complexe qu'on ne l'imaginait il y a dix ans encore - mais on raisonnait en termes de robots au lieu de raisonner en termes d'intelligence artificielle !

Prenez la médecine personnalisée, qui traite chaque patient en fonction de ses caractéristiques génétiques : elle va tuer le médecin traditionnel - bien entendu, il ne s'agit que d'un raisonnement *ex ante* : les radiologues et les médecins rempliront de nouvelles fonctions. Le malade devient un ensemble de *data* et le médecin un *software*, ce qui n'empêche pas ce dernier d'être empathique ; le séquençage ADN représente trois milliards de bases ADN, vingt téraoctets de données brutes par malade ; chaque individu représente deux millions de mutations par rapport au génome de référence - toutes ne sont pas promesses de maladie, bien sûr -, dont l'analyse exige des ACP, des analyses en composantes principales, à huit cents millions de facteurs. Dans un tel monde, il est clair que même le Dr House, en quinze minutes de consultation, déshabillage et rhabillage

---

compris, sera impuissant. Nous allons donc assister à un transfert de pouvoir médical vers les algorithmes, la médecine étant désormais fondée sur un *big data* dont l'analyse par le cerveau biologique est exclue. Cela ne veut pas dire que le cerveau biologique n'a plus sa place dans cette équation ; mais cette place est profondément transformée. Cela n'est pas sans poser problème : aucun médecin n'étant capable d'analyser vingt téraoctets de données en quinze minutes, les décisions éthiques seront implicites dans l'algorithme d'intelligence artificielle.

Quelle réponse politique peut-on apporter à cette situation ? Je regrette qu'il n'y ait pas davantage de parlementaires présents cet après-midi. Le problème est réel : vous connaissez tous cette courbe qui fait froid dans le dos, celle de l'augmentation des revenus par décile en trois décennies de mondialisation. Tout le monde en a profité, sauf le petit blanc moyennement ou peu qualifié, c'est-à-dire les classes moyennes occidentales traditionnelles, qui ont vu leurs revenus stagner, voire, dans certains pays comme les États-Unis ou l'Allemagne, baisser de 1 à 2 % sur 30 ans. Ce qui est en jeu, en l'occurrence, ce n'est pas tant l'intelligence artificielle, mais, tout simplement, la confrontation entre l'Asie et l'Europe : un docteur en génomique à Bangalore coûte, à l'heure, trois fois moins cher qu'un ouvrier spécialisé qui fait les 35 heures en région parisienne. Cette situation ne durera pas ! Nous avons là l'une des explications à l'émergence du populisme et à la « *trump-brexitisation* », à laquelle personne ne s'attendait. Ce sont des réponses populistes aux angoisses. La réponse politique à ces angoisses n'a rien d'évident ; j'espère profondément que nous saurons garantir à la plupart des gens une formation professionnelle continue qui les rendra complémentaires de l'intelligence artificielle.

On nous propose aujourd'hui une utopie techno-marxiste : le revenu universel de base – je précise que le terme de « marxiste » n'est absolument pas péjoratif à mes yeux. C'est une idée absolument suicidaire. Ceux qui le promeuvent, comme M. Benoit Hamon, sont les idiots utiles d'une intelligence artificielle semi-forte. Si nous décidons de mettre de côté tous ceux qui ne sont pas complémentaires de l'intelligence artificielle en nous contentant de leur donner des jeux et du cirque, nous vivrons, dans 50 ans, *Metropolis*, et, dans un siècle, *Matrix* ! Nous devons nous battre jour et nuit, en réformant l'éducation et la formation professionnelle, pour assurer la complémentarité entre les travailleurs, quel que soit leur niveau de qualification, et l'intelligence artificielle faible. À défaut, nous ferions le choix d'un suicide collectif et ce serait de l'irresponsabilité politique et, au-delà, de l'irresponsabilité de toute la société civile. Il y a là un chantier de plusieurs décennies ; il est regrettable que les politiques, handicapés du mulot qu'ils soient, n'en prennent pas la pleine mesure.

Des contre-pouvoirs pourront-ils émerger ? Comment organiserons-nous la régulation, la gouvernance, la police de l'intelligence artificielle, faible d'abord puis semi-forte ? C'est un débat politique immense. Henri Verdier a beaucoup réfléchi sur ces questions ; je partage en grande partie sa vision. Il faut y travailler dès aujourd'hui. Ce n'est pas gagné ! De tels

---

dispositifs fourniront, demain, des services irremplaçables à nos concitoyens – chacun a pu constater le succès immense et totalement inattendu d'« Écho » et d'« Alexa », lancés par Amazon.

On ne peut qu'être inquiet, à l'image des résultats d'un sondage récent qui montrait que 50 % des Français, et 68 % des jeunes, confieraient volontiers leurs données de santé et leur santé aux GAFAs. Il faudra de la pédagogie pour organiser notre sursaut face aux GAFAs et aux BATX ! Et pour réaffirmer notre souveraineté ! Il est aujourd'hui plus urgent de réfléchir à ces enjeux de souveraineté, à ces enjeux économiques, scolaires et de formation, que de fantasmer sur une intelligence artificielle forte qui, comme le dit très bien Yann LeCun, ne permet pas d'envisager un scénario à la *Terminator* avant 2035.

**Mme Dominique Gillot, rapporteure.** – Nous sommes sans voix !

**M. Laurent Alexandre.** – Je me suis pourtant retenu, aujourd'hui !

**Mme Dominique Gillot, rapporteure.** – J'étais déjà vaccinée !

**Mme Marie-Claire Carrère-Gée.** – Juste une précision : Laurent Alexandre estime que nous avons sous-estimé le risque pour l'emploi du développement des nouvelles technologies. Je suis évidemment moins compétente que lui dans le domaine des technologies ; je suis néanmoins frappée par tant de certitudes, dans un monde qui me paraît, à moi, très incertain, s'agissant à la fois du progrès technologique lui-même et de sa vitesse de diffusion. Le progrès technologique n'est pas un système autonome ; sa vitesse de diffusion dépend de maints facteurs : la rentabilité économique, mais aussi les choix sociaux ou éthiques. On sait depuis des années automatiser les caisses dans les supermarchés ; mais l'automatisation ne se fait pas à la minute, et n'est toujours pas accomplie. Je pense nécessaire que nous restions collectivement très humble sur les prédictions. Les citoyens et les pouvoirs publics ont leur mot à dire sur la vitesse de diffusion des progrès technologiques, lesquels sont, par ailleurs et par définition, difficilement prévisibles. J'ai exposé clairement les limites de notre étude : nous avons travaillé sur la base des frontières technologiques actuelles.

**Mme Dominique Gillot, rapporteure.** – Votre étude n'est nullement mise en cause.

**Mme Marie-Claire Carrère-Gée.** – J'ai présenté les conclusions d'une étude, et non mes certitudes ; à côté de moi se trouve quelqu'un qui professe des certitudes.

**M. Laurent Alexandre.** – J'ai dit que le futur était imprévisible ! Je n'ai pas de certitudes.

**Mme Marie-Claire Carrère-Gée.** – Je sais aussi que certains ont intérêt à entretenir des idées fausses sur la vitesse de diffusion des technologies : il est plus facile de lever des fonds en promettant un progrès technologique majeur sous deux ans qu'en exprimant des incertitudes.

---

**Mme Dominique Gillot, rapporteure.** – Je me félicite d’entendre s’exprimer des avis contrastés ; c’est par le débat que nous avancerons ! J’aurais aimé que Mme la Secrétaire d’État vous entende, Monsieur Alexandre.

Je vais à présent donner la parole à M. Jean-Christophe Baillie. À la suite de vos études, vous êtes devenu, Monsieur Baillie, chercheur en intelligence artificielle au *Sony Computer Science Lab*, puis vous avez participé à différents projets d’entreprise liés à l’intelligence artificielle et à la robotique comme *GoStyle* ou *Aldebaran*. Vous êtes aujourd’hui à la tête de votre propre entreprise, *Novaquark*.

#### 4. M. Jean-Christophe Baillie, entrepreneur (*Novaquark*)

Laurent Alexandre vient de présenter un grand nombre des points dont je souhaitais traiter. Je vais tenter d’ouvrir des perspectives.

Loin de moi toute certitude sur ces sujets ; ce qui me semble intéressant, c’est de lancer des idées, des questions, des points de vue. Je voudrais notamment débattre de l’« utopie techno-marxiste » dont parlait Laurent Alexandre.

Il faut distinguer entre intelligence artificielle forte et intelligence artificielle faible. J’ai récemment écrit un article au titre provocateur, intitulé « Pourquoi AlphaGo n’est pas de l’intelligence artificielle », dans lequel j’appuie cette distinction. L’intelligence artificielle faible, ou ce que certains appelleraient l’informatique avancée, correspond à ce que nous savons faire aujourd’hui : opportunité économique incroyable au potentiel énorme, elle affectera assurément le marché du travail, mais selon une dynamique assez classique de destruction et de création d’emplois, donc sans bouleverser en profondeur le marché du travail comme pourrait le faire l’intelligence artificielle forte. L’un des défis que nous ne savons pour le moment pas du tout relever, c’est celui de l’interaction sociale pertinente, douée de sens, avec d’autres individus, dans un environnement complexe. Les agents d’entretien que je croise le matin accomplissent un travail d’une complexité ahurissante ; les gestes qu’ils effectuent paraissent évidents, mais les problèmes qu’ils résolvent résistent absolument à la robotique, ce domaine qui se frotte à la complexité du réel et, en effet, ne progresse pas de manière exponentielle.

Appelons cela l’intelligence artificielle forte. L’intelligence artificielle connaîtra sans doute un stade intermédiaire avant que nous soyons capables de créer une conscience artificielle. Je pense néanmoins que nous pouvons arriver jusque-là, mais je suis incapable de donner une date – ceux qui se prêtent à ce jeu pratiquent un exercice périlleux ! Des avancées scientifiques profondes sont nécessaires : il y va de l’élaboration d’une véritable théorie de l’intelligence, rien de moins. Et nous ne savons pas s’il y en a pour dix ou pour cent ans ! Dire le contraire, c’est prétendre qu’il eût été possible de

---

prévoir dès le XIX<sup>e</sup> siècle que la théorie de la relativité serait un jour formulée. On ne peut prédire l'émergence d'une idée nouvelle.

Quoi qu'il en soit, le jour où l'on en viendra à une intelligence artificielle forte, on assistera à une modification profonde du marché du travail : les machines que nous fabriquerions seraient capables d'exercer y compris des métiers manuels, grand défi de l'intelligence artificielle.

On peut y voir une catastrophe, la promesse d'une explosion du taux de chômage. Mais ne serait-ce pas là, plutôt, une opportunité pour repenser l'organisation de la société ? Aujourd'hui, notre société est crispée sur la notion de travail, dès l'école. Nous sommes invités, dès l'enfance, à apporter notre pierre à l'édifice commun en trouvant notre place sur le marché du travail. En l'absence de robots, nous n'avons pas le choix ! Quelqu'un doit bien faire le travail, extrêmement important, de l'agent d'entretien. Nous avons besoin d'une société qui valorise le travail et qui, par exemple, à supposer qu'elle mette en place un revenu universel, doit l'établir à un niveau intenable, indécent : si ce niveau est fixé à 2 000 euros, qui acceptera de faire le travail d'entretien dont je parlais chaque matin ?

Pour les besoins de la discussion, on pourrait introduire une distinction entre travail et activité : le travail, c'est ce que vous faites pour subvenir à vos besoins, et que vous arrêtez de faire si vous gagnez au loto – c'est ce que feraient la plupart des agents d'entretien. Par contraste, on peut introduire la notion d'activité : ce que vous aimez faire par passion – si vous gagnez au loto, vous continuez à le faire. Songez, de ce point de vue, à une société dans laquelle la robotique et l'intelligence artificielle forte mettraient fin au travail et autoriseraient la mise en place d'un revenu universel décent, non contradictoire avec la nécessité que les travaux que j'évoquais soient effectués. Dans cette société sans travail, les robots accomplissent ces tâches dont personne, en réalité, ne veut.

Cette société de l'activité ne serait pas une société de l'oisiveté ; simplement, les gens y seraient pleinement libres de décider de l'organisation de leur temps, et y compris, d'ailleurs, de ne rien faire, ou de passer leur vie à la plage en lisant La Pléiade. Une telle activité ne ferait l'objet d'aucune stigmatisation, au contraire de ce qui se passe aujourd'hui, dans une société du travail – une société d'avant l'intelligence artificielle forte –, qui doit inscrire dans sa logique la nécessité de travailler – le chômage, c'est mal ! Une autre logique sociale est imaginable.

Le point clé de cette vision utopique, c'est évidemment l'éducation : il s'agirait de créer des citoyens capables de jouir réellement de ce temps et de cette liberté. Nous n'en sommes pas là ! Si, du jour au lendemain, une grande partie de nos concitoyens étaient mis au « chômage », je ne suis pas certain que les choses se passeraient bien. Il faudra donc créer, *via* l'éducation, un terreau où seront cultivées d'autres valeurs et d'autres visions de la société, ouvrant, à très long terme, une nouvelle ère pour l'humanité.

---

Pour l'heure, on en reste toujours à l'équation « intelligence artificielle forte égale chômage et fin du travail ». C'est mal, dit-on ! Certes, dans la société d'aujourd'hui ; mais il est possible d'imaginer une société différente.

**Mme Dominique Gillot, rapporteure.** – Que faites-vous de la dignité sociale, de l'utilité sociale ?

**M. Jean-Christophe Baillie.** – Il existe certes un besoin communément partagé de se sentir utile ; mais un tel besoin peut très bien se réaliser dans une activité librement choisie, par contraste avec un travail pénible que vous ne pouvez librement abandonner, et qui s'apparente à une privation de liberté. Je ne prône absolument pas l'idée d'une oisiveté généralisée ! J'introduis l'idée d'une activité librement choisie, en accord avec des passions et des intérêts, en un sens très large, nourris par l'éducation : sport, art, culture, accompagnement des personnes âgées par exemple.

**M. Laurent Alexandre.** – Les robots le feront mieux que nous !

**M. Jean-Christophe Baillie.** – Je n'en suis pas sûr ! L'idée que le travail serait nécessairement bénéfique au motif qu'il donnerait aux gens une utilité est une erreur de logique.

**M. Gilles Dowek.** – Ce désir d'utilité sociale est une invention du XIX<sup>e</sup> siècle ! C'est une construction sociale. Casanova n'est utile à rien ; il n'en est pas du tout malheureux.

**Mme Dominique Gillot, rapporteure.** – Je vais à présent donner la parole à M. Jean-Claude Heudin. Enseignant-chercheur, vous avez, dans les années 1990, co-fondé une société spécialisée en intelligence artificielle. Vous avez été expert auprès de l'Union européenne et conseiller scientifique pour la Cité des sciences. Surtout, vous êtes, depuis de nombreuses années, directeur de l'Institut de l'internet et du multimédia, placé au sein du pôle universitaire Léonard-de-Vinci. Vous avez publié de nombreux ouvrages et vous vous intéressez aux questions politiques et sociétales que pose l'irruption des machines intelligentes.

### **5. M. Jean-Claude Heudin, directeur de l'Institut de l'internet et du multimédia**

Je vais répéter des choses qui ont déjà été dites ce matin et cet après-midi ; mais il existe un algorithme, en intelligence artificielle, qui s'appelle l'« apprentissage par renforcement ». Je vais donc modestement tenter de renforcer quelques points.

Ma présentation s'articulera autour d'une idée simple : dans un contexte historiquement très anxigène, où il se dit beaucoup de choses sur l'intelligence artificielle, notre priorité devrait principalement aller à

---

l'éducation et à la formation. Le grand public est pour l'essentiel informé par les *blockbusters* américains, qui véhiculent une vision de l'intelligence artificielle complètement déformée, très alarmiste. En tant que praticien de l'intelligence artificielle, je peux vous dire que la réalité, dans les laboratoires, est beaucoup plus laborieuse ! Je vais donc m'efforcer de remettre de la mesure dans ces débats.

L'intelligence artificielle est une petite partie de l'informatique. Ce dont nous avons parlé aujourd'hui à ce titre renvoie en réalité au *deep-learning*, qui correspond à de spectaculaires progrès récents, mais ne représente que la partie émergée de l'iceberg. Que s'est-il réellement passé depuis 2013 ? Nous connaissons un nouvel « été torride » autour de l'intelligence artificielle et, plus spécifiquement, du *deep-learning*. Avant 2013, nous savions construire des applications à partir des réseaux de neurones, mais les résultats n'étaient pas bons - le taux de réussite, s'agissant de la reconnaissance d'une image de chat, par exemple, n'était que de 50 %.

En 2013 se sont conjugués trois facteurs. D'abord, un peu par hasard, une équipe de chercheurs canadiens a réussi à initialiser des réseaux multicouches, ou « réseaux profonds ». Le deuxième facteur est l'accès aux données, qui explique l'avantage des GAFAs, lesquels possèdent de très nombreuses données - des algorithmes, même médiocres, alimentés de beaucoup de données se révèlent bien meilleurs que de très bons algorithmes dont les données seraient plus pauvres. Troisième facteur : la capacité de calcul des cartes graphiques GPU. Résultat : on sait aujourd'hui construire des systèmes dont les taux de réussite sont voisins de 100 %.

Qu'est-ce que cela change ? Auparavant, pour aborder des problèmes complexes et automatiser la réponse à de tels problèmes, il fallait construire des modèles. Désormais, si l'on dispose des données, on peut les faire apprendre à la machine. De nombreux domaines en sont affectés - je pense à la voiture autonome, aux assistants personnels dans le secteur du commerce et des relations clients, à la santé, à la robotique.

Je reviens sur la distinction entre intelligence artificielle faible et forte, qui me semble un peu légère. Je préfère définir six niveaux : l'intelligence sous-humaine pour des tâches spécifiques - la grande majorité des systèmes conçus actuellement, même dans le cadre du *deep-learning*, reste sous-humaine, y compris la reconnaissance vocale ; une intelligence équivalente à celle de l'humain, pour des tâches spécifiques toujours ; un troisième niveau, supérieur à la plupart des intelligence humaines, à nouveau pour des tâches spécifiques - c'est le stade où nous en sommes avec un logiciel comme AlphaGo ; l'intelligence supérieure à toute intelligence humaine pour des tâches spécifiques, qui nous fait entrer dans l'utopie ; un cinquième niveau supérieur à l'intelligence humaine pour une majorité de tâches - ce n'est pas demain la veille, loin s'en faut ; enfin, l'intelligence artificielle ultime, dystopie ou utopie - pour moi, c'est jamais !

Est-ce un danger pour l'emploi ? Au contraire, rien n'est moins sûr. Nous sommes seulement capables, aujourd'hui, d'établir des corrélations ; en



---

tout état de cause, les pays les plus robotisés ont les taux de chômage les plus bas. Les études se contredisent parfois. Seule certitude : le raisonnement en termes d'emplois détruits et d'emplois créés est binaire, mais l'intelligence artificielle modifie en profondeur un nombre croissant de métiers.

Je précise que la solution n'est de toute façon pas une nouvelle taxe. Ce n'est certainement pas en mettant un boulet aux pieds de nos entreprises et de nos *start-ups* que nous améliorerons notre compétitivité. J'espère ne pas être d'accord avec Laurent Alexandre lorsqu'il dit que nous avons déjà perdu la bataille ! Je suis un indéfectible optimiste : j'ai encore un espoir.

**M. Laurent Alexandre.** - Je n'ai pas dit que c'était perdu ! J'ai dit que c'était « ric-rac », et qu'il y avait le feu au lac !

**M. Jean-Claude Heudin.** - Il y a au moins un point de désaccord entre nous : l'intelligence artificielle est un atout français. Certes, Yann LeCun est parti chez Facebook, mais la recherche française est de tout premier plan, des entreprises dynamiques existent ; il faut de toutes nos forces les encourager, car les enjeux sont réels.

Il existe néanmoins un déficit d'information du grand public ; les informations qui circulent sont très anxiogènes. Je fais presque une conférence par semaine pour tenter d'évangéliser, de rassurer le public et de démystifier le sujet. Un effort est nécessaire au niveau des formations. Les formations de l'enseignement supérieur ménagent très peu de place à l'intelligence artificielle, ne serait-ce qu'une initiation à ses enjeux. Quelques écoles le font, mais elles sont très peu nombreuses - et je ne parle pas des universités. Et les applications de l'intelligence artificielle dans l'enseignement sont malheureusement quasi inexistantes.

Pour paraphraser Michel Serres, notre société connaît une complexification extrême ; nous allons avoir besoin de toutes les intelligences pour relever les grands défis qui vont se présenter à nous. L'intelligence artificielle doit augmenter notre intelligence, et non la remplacer. Je refuse le vieux démon, qui remonte aux Grecs, de l'annihilation du travail au profit d'une société du loisir. Le travail est très structurant ! J'aime la métaphore de l'intelligence artificielle comme « troisième hémisphère » de notre cerveau biologique. Elle est indispensable pour la résolution de problèmes complexes qui, aujourd'hui, dépassent notre intelligence biologique. La réflexion éthique, en la matière, est absolument indispensable.

**Mme Dominique Gillot, rapporteure.** - Vous venez de recentrer notre débat sur la question, déjà abordée ce matin, de l'éducation et de l'information du public, de la construction d'une opinion éclairée.

Je vous pose une question avant de lancer le débat. Quels sont vos liens avec les médias ? Des journalistes spécialisés travaillent-ils sur ce sujet ?

**M. Jean-Christophe Heudin.** - Je suis un habitué de « *La Tête au carré* », sur France Inter. Certains médias, heureusement, font leur travail !

Mais ces sujets restent anxiogènes pour le grand public. Dans l'enseignement supérieur, et même dans les grandes écoles de commerce, très peu de choses sont organisées sur cette question. Il est pourtant indispensable que les jeunes comprennent l'impact de ces technologies sur leurs futurs emplois. Je le dis souvent : à l'ère des données, du temps réel, de l'analyse prédictive, le *marketing* à la papa est mort.

**Mme Dominique Gillot, rapporteure.** – Ces derniers mois, nous constatons néanmoins une prise de conscience sur la nécessité d'échanger autour de ces questions : des conférences et des rencontres, comme celle d'aujourd'hui, sont organisées.

## 6. Débat

**M. Jean Ponce.** – Je suis enseignant-chercheur. Je ne suis ni économiste, ni politique, ni futuriste – Dieu m'en préserve ! Je voudrais réagir à certaines déclarations péremptives de M. Alexandre. On ne peut les laisser passer. À entendre qu'il vaut mieux de mauvais algorithmes et de bonnes données, Yann LeCun, que je connais bien, et qui est le pape de ces questions, frémirait ! Ce n'est tout simplement pas vrai.

Je frémis également en entendant dire que la France est un pays du tiers-monde dans ce domaine, ou que la robotique ne décolle pas à cause de son coût. Ce n'est pas vrai – pensez à la voiture autonome ! Je donne quelques exemples concrets d'investissements qui sont réalisés actuellement dans le domaine de la robotique : Dyson, le fabricant d'aspirateurs, vient d'investir cinq millions de livres dans un laboratoire de l'*Imperial College* de Londres ; Foxconn s'apprête à installer 50 000 robots en Chine ; Toyota a investi 50 milliards de dollars au titre de la recherche en robotique dans la *Silicon Valley* ; le gouvernement américain vient de déployer 250 millions de dollars dans un partenariat public-privé sur ce sujet, à Pittsburgh. Le plus grand succès public en matière de robotique est le robot *Roomba*, ce petit aspirateur qui coûte 100 euros. L'histoire des robots chers qui ne décollent pas, c'est un mythe ! Même chose pour l'intelligence artificielle forte. Il faut faire un petit peu attention à ce qu'on dit.

J'ai apprécié l'intervention de Jean-Claude Heudin ; je reviens simplement sur un détail technique : il est faux de dire qu'en 2013 a été découvert un moyen d'initialiser des réseaux de neurones. On a simplement déployé davantage de données *via* d'importantes capacités de calcul.

**M. Jean-Claude Heudin.** – Nous pourrions en discuter !

**M. Jean Ponce.** – Aujourd'hui, nous ne disposons d'aucune explication théorique des raisons pour lesquelles les réseaux de neurones fonctionnent, c'est-à-dire donnent, dans un certain nombre de domaines, d'excellents résultats.

---

**M. Laurent Alexandre.** – J’ai été mal compris. Je n’ai pas dit qu’il n’existait pas de potentiel économique et technologique, en France, dans ces domaines ! J’ai dit que nous étions importateurs nets et massifs. C’est incontestable ! La balance de ce que nous importons et de ce que nous exportons, en matière d’intelligence artificielle, est totalement déséquilibrée. *Android* est propriété de Google ; *iOS*, d’Apple. Ce sont deux sociétés californiennes, et 99 % de l’intelligence artificielle consommée en France arrive à travers ces deux plateformes.

**M. Jean Ponce.** – *Android* est un système d’exploitation ; pour autant que je sache, ça n’a rien à voir avec de l’intelligence artificielle.

**M. Laurent Alexandre.** – Les applications qu’on utilise tous les jours passent par ces deux plateformes. C’est de l’intelligence artificielle, même si les gens ne l’y rattachent pas spontanément ! Nous sommes importateurs nets d’intelligence artificielle.

Sur la robotique, je n’ai pas été compris. Je n’ai pas dit qu’on ne faisait pas de progrès en robotique. Mais un investissement de 250 millions de dollars, c’est très peu, par comparaison avec les GAFAs et l’intelligence artificielle, où les sommes se comptent en dizaines de milliards de dollars. J’ai simplement dit que les progrès, s’agissant de la partie mécanique des robots, étaient moins exponentiels que ceux qui concernent la partie relevant de l’intelligence artificielle. Car, en termes de coûts, dans un robot, il y a plus de mécanique que d’intelligence artificielle et leur taux de croissance en est donc diminué d’autant. Le prix des pièces mécaniques en titane ne diminue pas de 50 % tous les douze mois.

**M. Jean Ponce.** – Vous plaisantez ! Le fait de parler de pièces mécaniques en titane n’est pas un argument. Je ne mets en doute ni votre enthousiasme, ni votre rhétorique, ni vos connaissances, mais c’est un mauvais argument !

**M. Laurent Alexandre.** – Il y a des métaux rares dans les robots. Prenons le cas d’Atlas. La robotique, parce qu’elle est en partie mécanique, croît moins vite que ce qui est purement immatériel. En termes de coûts relatifs, c’est évident. La partie mécanique ne suit pas la loi de Moore.

**Mme Laurence Devillers.** – Votre vision est très restrictive ! Le champ de la robotique est très large. Certains domaines de la robotique se développent plus vite que d’autres.

**M. Jean Ponce.** – La loi de Moore s’applique uniquement à des aspects mécaniques puisqu’il s’agit des circuits. Elle ne s’applique absolument pas au développement de l’intelligence artificielle, ni, d’ailleurs, à la fabrication électromécanique du robot.

**M. Jean-Christophe Heudin.** – Gordon Moore est le premier à dire que sa loi ne s’applique pas à l’intelligence artificielle.

D’une manière générale, je suis très réservé dès qu’on parle d’exponentielle. Pour reprendre les termes de Stephen J. Gould, je pense que

---

l'histoire des découvertes scientifiques et technologiques relève plutôt de la théorie des « équilibres ponctués », avec de longues phases de tâtonnement et de soudaines avancées.

**M. Jean-Daniel Kant.** – Je voudrais revenir sur le thème de l'impact de l'intelligence artificielle sur l'emploi. Il faut être très prudent sur les projections. La question est complexe, et ceux qui assènent des tendances définitives à vingt ou trente ans prennent le risque de se tromper.

Les observations ne donnent aucun résultat clair. Ce sont, disent certains, les emplois peu qualifiés qui seront touchés ; d'autres leur rétorquent que ce sont plutôt les emplois intermédiaires qui sont menacés par l'automatisation, ce que semble confirmer l'exemple californien. En l'absence d'une intelligence artificielle forte, comme l'a dit Jean-Christophe Baillie, il faut des gens pour entretenir les jardins, garder les enfants ou faire le ménage.

Je compte étudier cette question scientifiquement en utilisant les techniques de simulation avec des systèmes multi-agents, dont j'ai parlé ce matin. Seule certitude : ça va transformer l'emploi – de quelle façon ? On ne sait pas – et il faudra reposer la question de la place du travail dans la société. Ce n'est pas rien.

**M. Benoît Le Blanc.** – Je voudrais apporter trois remarques sur le débat en cours. L'intelligence artificielle donne lieu à de considérables investissements financiers, c'est indéniable. Mais comment savoir s'il ne s'agit pas d'une bulle ? Cette dynamique n'est-elle pas simplement due au fait que certaines personnes ont beaucoup d'argent disponible ? On ne sait pas ce que ça va donner !

Par ailleurs, s'agissant du rapport qu'ils entretiennent avec le travail, des notions de carrière ou d'investissement personnel, les jeunes ont des réactions très inattendues. Ce sont eux qu'il faudrait interroger sur l'évolution du marché du travail.

Enfin, on parle beaucoup, ici, d'exponentielle. La cognition humaine a beaucoup de mal à appréhender certains concepts mathématiques, comme l'exponentielle, mais aussi, par exemple, les notions liées aux faux positifs et aux faux négatifs. Ces choses-là n'existent pas dans la nature !

**M. Olivier Guillaume.** – S'agissant des supermarchés sans caisses, les technologies sont en train d'être mises au point. Elles seront utilisables rapidement. Monsieur Alexandre, vous avez dit que l'intelligence artificielle serait meilleure que les radiologues d'ici à 2030, soit dans treize ans ; mais ce sera le cas, d'ores et déjà, dans trois mois, pour le dépistage des cellules cancéreuses, notamment. Nous travaillons sur ce sujet actuellement.

**Mme Dominique Gillot, rapporteure.** – Ça ne veut pas dire que nous n'aurons plus besoin de radiologues ou de médecins !

**M. Laurent Alexandre.** – Ce n'est pas ce que j'ai dit.

---

**M. Henri Verdier.** – Nous avons beaucoup parlé du travail, je ne pensais pas que nous en parlerions autant. Je le pense depuis longtemps : la révolution numérique sera une révolution organisationnelle et managériale ou ne sera pas. Pendant un siècle, une pulsion de robotisation du travail s’est imposée : obéissance érigée en principe cardinal, instructions très claires et non ambiguës, respect des plans et de la gouvernance, etc. Si le travail consiste à exécuter la pensée d’un comité exécutif, il est robotisable, ou « intelligence-artificialisable ».

Mais si le travail, comme dans la recherche, comme dans les *start-ups*, comme en politique, signifie plutôt l’engagement, la détermination, la créativité, alors il est loin d’être menacé par l’intelligence artificielle. Certaines philosophies du *management* et de l’organisation du travail seront davantage menacées de substitution que celles qui tirent parti du meilleur des humains.

**Mme Marie-Claire Carrère-Gée.** – Je partage entièrement ce point de vue. Au Conseil d’orientation pour l’emploi, nous travaillons sur l’impact quantitatif et qualitatif du développement de l’intelligence artificielle sur l’emploi, mais aussi sur l’organisation du travail et les conditions de travail – ce sera l’objet du second tome. À quelles conditions la nouvelle vague de progrès technologique est-elle de nature à rendre l’activité ou le travail humains plus riche et moins pénible ?

**M. Laurent Alexandre.** – L’impact sera très hétérogène en fonction des secteurs. Mais l’avenir est très difficile à prédire en matière d’intelligence artificielle – n’oublions pas que nous avons connu deux ou trois « hivers » de l’intelligence artificielle, pendant lesquels nous avons fait très peu de progrès.

Certaines tâches nous semblent très complexes, comme l’analyse d’un scanner, mais se révèlent relativement faciles à automatiser. S’agissant du diagnostic du cancer du sein, le radiologue est déjà égalé, voire dépassé. L’échéance de 2030, dont je parlais tout à l’heure, c’est la date à laquelle il est imaginable que le radiologue soit dépassé par l’intelligence artificielle pour tout type d’examen radiologique. Cet horizon est relativement court pour une spécialité qui ne s’imaginait pas du tout menacée il y a dix ans.

Quant à l’intelligence artificielle forte, je n’y crois pas avant plusieurs siècles. Je vous incite à lire un magnifique papier du *New York Times* du 29 juillet 1997, après la victoire de l’ordinateur sur Kasparov aux échecs. Le journaliste interroge un spécialiste de l’intelligence artificielle : « Les échecs, c’est très facile ! Le vrai défi, c’est le jeu de go ; en la matière, je vous le dis, l’ordinateur ne saura pas jouer avant 100 ou 200 ans ! ». Comme vous le savez, il y a 6 mois, Lee Sedol, le champion du monde de jeu de go, a été écrabouillé par le programme informatique.

**Mme Marie-Claire Carrère-Gée.** – Des erreurs de prédiction se sont produites dans l’autre sens !

**M. Laurent Alexandre.** – Certes !

---

**Mme Laurence Devillers.** – La comparaison est nulle et non avenue : on compare un humain avec une machine ; mais derrière la machine se trouvent cent ingénieurs au travail !

**M. Henri Verdier.** – Le couple humain-machine bat toutes les machines et tous les humains ; cela donne des parties bien plus intéressantes !

**Mme Dominique Gillot, rapporteure.** – Nous sommes d'accord !

**M. Stéphane Pelletier.** – Ce sujet est extrêmement complexe, voire insaisissable. Je suis le fondateur de *Dreamstarter*, une plateforme *web* qui a pour objectif de faire sortir la santé et l'alimentation des sphères spéculatives.

Je rebondis sur le thème de la santé. Nous discutons d'intelligence artificielle ; l'objectif de ces débats est de dégager une orientation réaliste, dans un monde extrêmement incertain. Des talents existent, qui essaient de convaincre les citoyens, à défaut des pouvoirs publics, du bien-fondé de leurs études ; je pense notamment à celles du professeur Benveniste, qui ont été reprises par le professeur Montagnier, sur la mémoire de l'eau. Au lieu de faire des supputations sur ce qu'il adviendra dans deux cents ans, peut-être pourrions-nous mieux considérer ceux qui proposent des approches pionnières et avant-gardistes.

**Mme Dominique Gillot, rapporteure.** – L'un n'empêche pas l'autre.

**M. Stéphane Pelletier.** – Sur l'intelligence artificielle, nous devons réussir à déterminer une orientation réaliste et consensuelle. Je souhaiterais que le citoyen soit au cœur des démarches proposées.

**Mme Dominique Gillot, rapporteure.** – C'est l'un des enjeux de cette journée !

## V. INTERVENTION DE LA MINISTRE : AXELLE LEMAIRE, SECRÉTAIRE D'ÉTAT CHARGÉE DU NUMÉRIQUE ET DE L'INNOVATION

**Mme Dominique Gillot, rapporteure.** - De nos trois tables rondes, ressort la nécessité de concerner l'ensemble de la population, de mettre en place un fil conducteur dès la petite enfance et tout au long de la vie, pour permettre à nos concitoyens de bien se positionner sur les nouvelles technologies nées de l'intelligence artificielle. Il est en effet indispensable de faire preuve de mobilité professionnelle et de compréhension des grands enjeux éthiques et sociétaux posés par ces technologies. Nous sommes par conséquent très intéressés par votre intervention, Madame la Ministre, d'autant que nous avons appris le lancement demain d'une réflexion importante pour définir une stratégie nationale pour l'intelligence artificielle.

**Mme Axelle Lemaire, secrétaire d'État chargée du numérique et de l'innovation.** - Sur un sujet aussi complexe et difficile, on voit bien combien les responsables politiques doivent faire confiance aux experts. Merci d'avoir organisé cette journée et cette audition, car le sujet de l'avènement des technologies d'intelligence artificielle dans nos sociétés ne doit pas rester l'apanage des seuls chercheurs et ingénieurs. Il est grand temps que la thématique entre dans la sphère publique car elle se trouve au cœur de notre vie et de notre quotidien.

Le champ d'application de l'intelligence artificielle est de plus en plus vaste, comme j'ai eu l'occasion de le constater lors du *Consumer Electronic Show* (CES) de Las Vegas il y a quelques jours. Un grand nombre d'entreprises présentes là-bas avaient intégré des technologies faisant appel à une certaine forme d'intelligence artificielle, et en particulier au *machine learning*. Les applications sont de plus en plus nombreuses dans le domaine de la santé, de la défense, des relations entre les entreprises et leurs clients. À chaque étape, des commentaires décrivent les risques et les interrogations. L'approche française qui semble prédominer est celle de l'appréhension face au risque, puisque 65 % des Français se disent inquiets du développement de l'intelligence artificielle. Comparativement, 36 % des Britanniques et 22 % des Américains expriment la même crainte. Il existe sans doute des différences culturelles d'approche du risque. Face à ce défi, il n'est pas question que la parole des États soit absente. Le Président Barack Obama a fait publier par la Maison Blanche trois documents sur les conséquences possibles et la nécessité de mieux appréhender le défi de l'intelligence artificielle. Ce faisant, il fait appel à la mythologie de la « nouvelle frontière à conquérir » chère à son pays. Ainsi, l'importance d'investir dans l'intelligence artificielle est comparée au programme Apollo. D'autres États, dont font partie le Japon et la Corée du Sud, commencent seulement à s'intéresser à ces enjeux. À ma connaissance, l'initiative française est la seule

---

en Europe. L'idée part du constat que la France dispose d'atouts considérables pour s'inscrire en leader dans le domaine de l'intelligence artificielle. L'école mathématique française rencontre l'école informatique, de même que la science des données. De manière non formelle, nous avons dénombré près de quatre mille chercheurs français travaillant dans ces domaines et cent formations, dont d'excellents masters. Cette excellence est reconnue à l'international puisque de grands groupes ont ouvert leurs centres de recherche spécialisés dans notre pays. Des entreprises françaises issues ou non du secteur informatique mènent également des travaux de recherche dans le domaine de l'intelligence artificielle. Je pense notamment à CapGemini, Atos, Sopra Steria, Thalès et Safran, mais nous ne sommes qu'au début de l'histoire. L'enjeu pour le tissu économique et industriel français est de réussir à intégrer ce type de technologies sur des modes de production non exclusivement dédiés au numérique ou aux technologies d'intelligence artificielle.

Prenons l'exemple désormais classique du secteur automobile. La carrosserie avec de l'ingénierie civile intègre de plus en plus souvent du logiciel, ce qui soulève des questions éthiques. Cet exemple illustre la transformation que doivent mener nos industries traditionnelles pour produire des biens offrant une valeur ajoutée sur les marchés internationaux.

Le cadre réglementaire et législatif de notre pays a évolué de manière à accueillir de plus en plus ce potentiel de développement par l'intelligence artificielle. Je pense que le choix fait par le gouvernement de mettre la donnée au centre des politiques publiques en créant le statut juridique des données d'intérêt général et la mission de service public de la donnée, doit permettre de proposer ce carburant de l'intelligence artificielle qu'est la *data*. Tout cela ne peut naturellement se développer sans s'accompagner du questionnement éthique et sociétal. D'ailleurs, la loi pour une République numérique confie à la CNIL - dont je salue la présidente Isabelle Falque-Pierrotin ici présente - la mission d'animer le débat public autour des enjeux technologiques, en particulier de l'intelligence artificielle. Le sujet commencera par les algorithmes, mais l'idée est bien d'introduire un débat participatif et collaboratif pour sortir des débats d'experts et aboutir à des choix de société.

J'aimerais ici faire passer le message qu'il existe un formidable potentiel de développement économique et social lié à l'essor de l'intelligence artificielle. Systématiquement cependant, ce développement doit s'accompagner d'un questionnement sur les enjeux éthiques et de modes de gouvernance permettant de définir un cadre réglementaire. Peut-être que ce qui a été fait pour la bioéthique devra être reproduit pour l'intelligence artificielle. Il faut poser les enjeux sans langue de bois et sans jargon. Le but est de prévenir les risques pour se prémunir contre d'éventuels dérapages, et être maîtres de nos choix. Le mot essentiel est en effet la « maîtrise ». Voulons-nous que la France soit uniquement un pays consommateur au niveau international, ou un pays offreur ? Telle est la question.



---

Les enjeux éthiques ne sont pas toujours posés sans caricature. Aux États-Unis, le panel des réponses est très large. Les questionnements portent notamment sur les robots tueurs, dont Bill Gates et Elon Musk réclament l'interdiction. Stephen Hawking participe également à ce débat en demandant une limitation de l'usage de l'intelligence artificielle dans le cadre d'une mission internationale se voyant confier des objectifs précis. Nick Bostrom, chercheur à Oxford, a également beaucoup écrit sur le sujet avec une vision sans doute moins alarmiste. Dans tous les cas où la réalité technologique rencontre le potentiel économique, se pose la question de l'éthique et de l'humain. En termes d'emplois, des adaptations seront nécessaires sans pour autant que se produise la destruction catastrophique d'emplois telle qu'elle est parfois annoncée. Il faut que l'ensemble de notre société s'adapte à la nouvelle donne et à la transformation des tâches. Les plus répétitives seront appelées à disparaître pour se concentrer sur la valeur ajoutée, de sorte qu'il faudra faire monter en compétence l'ensemble de la population française.

Confronté à l'ensemble de ces enjeux, il fallait effectivement que le gouvernement soit au rendez-vous. Le sujet de l'innovation ayant récemment été ajouté à mon portefeuille, j'ai commencé à travailler dès le mois d'octobre à la définition de ce que pourrait être l'intelligence artificielle pour notre pays. Dès demain, aura lieu le lancement des travaux en groupes de travail, aux côtés de Thierry Mandon, mon collègue en charge de l'enseignement supérieur et de la recherche, autour de certains champs prioritaires. Il apparaissait dans un premier temps important de fédérer l'écosystème français, constitué de chercheurs mais également de penseurs sur les questions éthiques et des entreprises. Parmi celles-ci les *start-ups* font sans doute partie des acteurs majeurs de notre pays. Dans un travail de cartographie informel, nous en avons dénombré au minimum cent, dont certaines ont déjà remporté des succès retentissants et des levées de fonds élevées. À l'université, un grand nombre de post-doctorants veulent également être des entrepreneurs. Nous devons par conséquent étudier les passerelles les plus fluides possibles entre la recherche et l'entreprise, dans un enjeu de maîtrise.

Le deuxième axe est celui de l'industrialisation et du transfert des technologies de l'intelligence artificielle vers les secteurs économiques pour en maximiser les retombées sur notre territoire. Le troisième axe de travail aura trait à la définition du cadre réglementaire et institutionnel qui prendrait en considération les enjeux macro-économiques, sociaux, éthiques ainsi que nos priorités en matière de souveraineté et de sécurité nationale.

Tels sont les trois axes qui guideront les travaux des sept groupes, auxquels s'ajouteront d'autres groupes de travail plus spécifiques. Au total, une centaine de personnes se réuniront dès demain. Ces acteurs disposeront d'un délai court pour réfléchir et donner les outils au gouvernement suivant pour alimenter les priorités des chercheurs et des entreprises identifiés. Nous sommes au début de l'histoire, mais la France tient à être présente dans l'écriture de cette page très vaste. Nous devons voir l'intelligence artificielle

---

comme un outil au service de la croissance et de la prospérité sociale et sociétale, pour peu qu'il soit bien utilisé.

**Mme Dominique Gillot, rapporteure.** - La réunion est-elle ouverte à ceux qui n'auraient pas été invités ?

**Mme Axelle Lemaire.** - Nous avons d'ores et déjà dépassé la capacité d'accueil de la salle, mais j'invite tous ceux qui souhaiteraient nous rejoindre à contacter mon conseiller Nathaniel Ackerman. J'aimerais organiser des journées consacrées à l'intelligence artificielle ainsi que des ateliers collaboratifs dans les territoires.

Le lancement de demain nous permettra, à Thierry Mandon et moi-même, d'annoncer la composition du Comité d'orientation. Le groupe de travail « Cartographie » mesurera les forces en présence en France. Les autres groupes de travail seront respectivement consacrés à la recherche et au développement des compétences, à l'industrie et au transfert de technologies vers le secteur économique (robotique, internet des objets, énergie, application des technologies hybrides, applications médicales...) Un sous-groupe réunira en outre les réflexions consacrées aux véhicules autonomes, finances et commerce. Les groupes en charge des considérations macroéconomiques et éthiques, réfléchiront notamment aux thèmes « souveraineté et sécurité nationale », « social et éthique ». Si certains d'entre vous souhaitent se joindre à ces groupes, je les invite à se faire connaître.

**Mme Dominique Gillot, rapporteure.** - Je pense que votre annonce de cette réunion d'ores et déjà à guichet fermé suscitera encore plus de curiosité. Je donne la parole à la salle.

## 1. Débat

**Danièle Bourcier, directrice de recherche émérite au CNRS.** - Avec ma double casquette de chercheuse et de juriste sur les questions liées à l'intelligence artificielle, je précise que je ne m'occupe pas seulement du droit de l'intelligence artificielle, mais je considère que les juristes eux-mêmes sont concernés par l'évolution de leur métier due au développement de l'intelligence artificielle. J'étendrai d'ailleurs le droit à l'administration. En France, il y a une tradition de l'intelligence artificielle dans l'administration. Depuis le temps que nous évoquons la réforme de l'État, nous nous apercevons que nous devons parler différemment aujourd'hui. Par exemple, une loi résout-elle un problème ? Non car tel n'est pas le but de la loi. Je m'arrêterai là, mais la matière est extrêmement riche. L'intelligence artificielle aura non seulement un impact sur le droit mais également sur d'autres domaines tels que les sciences humaines et sociales. La formation au niveau universitaire doit être développée sur ce thème dans nombre de facultés de droit.

---

**M. Olivier Guillaume, président O2 Quant.** - Je suis très heureux de vous entendre sur le lancement de cette stratégie et évoquer les sujets que vous avez précisés. Étant président d'une *start-up* fournissant des briques de technologie en intelligence artificielle pour les grands groupes dont vous avez mentionné le nom, j'ai besoin d'aide pour croître plus rapidement et effectuer le transfert de technologies. Aidez-moi s'il vous plaît.

**M. Jean-Daniel Kant, maître de conférences à l'Université Pierre-Marie-Curie Paris-VI.** - L'intelligence artificielle peut-elle aider le politique à concevoir de meilleures lois ? Peut-être est-ce le cas. Il existe des technologies permettant de simuler l'effet d'une loi sur la société. Le modèle est certes contestable mais il a le mérite d'exister. Cette application d'intelligence artificielle n'est pas souvent citée mais elle concerne le champ politique et le citoyen.

**M. Alexei Grinbaum, CEA Saclay, LARSIM.** - Quelle est votre vision européenne de l'intelligence artificielle ? Un pays à lui seul ne disposera pas nécessairement des données suffisantes pour pouvoir concurrencer les États-Unis.

**M. Pavlos Moraitis, Directeur du Laboratoire d'informatique de l'Université Paris-Descartes.** - Je souhaite appuyer la demande de création d'un groupe de travail intelligence artificielle et droit, car je travaille sur les applications de l'intelligence artificielle dans le domaine juridique.

**M. Patrick Albert, P-DG SuccessionWeb.** - Pourrions-nous transformer le secrétariat d'État en un ministère à part entière ?

**Mme Axelle Lemaire.** - Je n'avais pas anticipé la question consacré à « intelligence artificielle et droit » qui est pourtant fondamentale. J'étais ce matin avec le Bâtonnier de l'Ordre des Avocats de Paris, avec lequel nous avons évoqué le degré d'appropriation des avocats quant à l'évolution de leur métier. Grâce aux algorithmes, des prédictions sur les possibilités de gain d'une affaire seront sans doute possibles, tandis que les avocats apporteront une valeur ajoutée essentiellement humaine sur la compréhension des circonstances de la situation et le choix ou non d'aller au contentieux. Le métier d'avocat va ainsi être transformé par l'intelligence artificielle, qui facilite la prise de décision. Dans la loi pour une République numérique, nous avons tenté de faciliter l'accès aux travaux de recherche et aux publications scientifiques. Les chercheurs se sont ainsi vu ouvrir l'accès aux bases de données administratives en les croisant de façon respectueuse avec les données personnelles. À l'heure actuelle, des travaux sont menés par des économistes pour définir ce que serait un revenu universel. Pour cela, il faut connaître véritablement l'état de la pauvreté en France et pouvoir exploiter différentes données.

J'en viens à répondre à la question posée sur les politiques publiques. Je suis la première frappée par l'irrationalité à la base des décisions politiques. Tout le système fonctionne de façon telle que nous n'intégrons pas les données disponibles pour redonner sa légitimité à la décision politique. L'enjeu est celui de la capacité du politique à se reposer

---

sur les travaux des chercheurs pour jauger de l'efficacité des politiques publiques. Je suis très favorable à la constitution d'un groupe de travail réunissant l'intelligence artificielle, le droit et les sciences humaines.

Par ailleurs, le secteur public doit rester maître de la définition des politiques publiques. Ceci est particulièrement vrai dans les domaines de l'éducation et de la santé. Or lorsque des entreprises privées auront acquis une expertise et un quasi-monopole dans l'offre de services en santé et en éducation, l'hôpital et l'école seront alors mis en cause dans leur capacité à fournir ce type de services publics. Selon moi, un enjeu de légalité est en cause, ce qui passe par une nécessaire montée en compétence des administrations pour intégrer des technologies aujourd'hui totalement absentes des modes de prise de décision. J'ai lancé ce chantier avec la Direction générale de la concurrence, de la consommation et de la répression des fraudes (DGCCRF) à Bercy.

L'intelligence artificielle peut clairement aider le politique, au point que j'ai soutenu un projet d'incubateur d'innovation démocratique. Je viens d'apprendre que Facebook va faire de même avec Sciences Po pour un incubateur d'analyse des politiques publiques.

Monsieur Guillaume, je vous invite à prendre contact avec mon Cabinet dans le cadre de la mission *FrenchTech*.

Enfin, sur la question du secrétariat d'État ou du ministère, il appartiendra au prochain Gouvernement de définir l'architecture de l'organisation interministérielle sur les sujets d'innovation, qui doivent être considérés comme une priorité politique absolue. Les derniers débats des primaires n'ont sans doute pas encore évoqué ces enjeux.

**Mme Dominique Gillot, rapporteure.** - Merci beaucoup Madame la Ministre de vos réponses très précises et argumentées. Il faut affirmer le fait qu'on a besoin d'embarquer tous les publics dans l'appropriation des nouvelles technologies, cela ne doit pas se résumer à un partenariat entre les experts et les pouvoirs publics. Cette préoccupation de démocratisation de la nouvelle manière d'appréhender tous les actes de notre vie quotidienne, doit être partagée avec le plus grand nombre de nos concitoyens pour éviter que le clivage se creuse. Il s'agit d'un enjeu de démocratie pour la France, ayant également trait à sa capacité à exercer sa souveraineté dans un monde en pleine mutation.

## VI. QUATRIÈME TABLE RONDE, PRÉSIDIÉE PAR M. CLAUDE DE GANAY, RAPPORTEUR : ENJEUX ÉTHIQUES DE L'INTELLIGENCE ARTIFICIELLE

**Claude de Ganay, rapporteur.** - Je vous propose d'aborder notre quatrième table ronde consacrée aux enjeux éthiques de l'intelligence artificielle. « *Science sans conscience n'est que ruine de l'âme* », comme l'affirmait Rabelais. L'intelligence artificielle représente à cet égard un ensemble technologique par rapport auquel la réflexion éthique est non seulement nécessaire mais urgente.

Je demanderai aux intervenants de tenter de tenir leur temps de parole. Je donne la parole à Gilles Dowek, qui a une double casquette d'informaticien et de philosophe. Vous êtes directeur de recherche à l'INRIA et professeur attaché à l'ENS Paris-Saclay. Surtout, vous êtes un spécialiste des enjeux éthiques de l'intelligence artificielle. C'est pourquoi nous vous écoutons avec beaucoup d'attention.

### 1. Gilles Dowek, directeur de recherche à l'INRIA, professeur attaché à l'ENS Paris-Saclay

L'éthique des sciences et des techniques, objet des travaux de la Commission de réflexion sur l'Éthique de la Recherche en sciences et technologies du Numérique d'Allistene (CERNA), n'a cessé de se transformer au cours du temps. En premier lieu, nos valeurs elles-mêmes se sont transformées. Nous essayons de résoudre nos conflits de façon plus pacifique que par le passé, tandis que des questions éthiques différentes sont apparues à chaque époque. Notre époque, que nous pouvons faire débiter aux années 1930-40 est celle de la révolution informatique parfois appelée la troisième révolution industrielle. Les deux premières révolutions industrielles étaient celles de la transformation de l'énergie, alors que la révolution informatique est une révolution des techniques de l'information et de la connaissance, une révolution cognitive. En cela, elle est comparable à l'invention de l'imprimerie. Selon Michel Serres, cette troisième révolution est une révolution cognitive et de notre rapport à la connaissance.

Les deux premières révolutions industrielles ont consisté à remplacer nos muscles par des machines. Les révolutions des techniques de l'information remplacent nos cerveaux par des machines. Il n'est pas surprenant dans ces conditions que les questions éthiques soient différentes. À ce titre, nous pouvons recenser des questions générales et des questions particulières.

---

La première de ces révolutions globales tient au fait que naguère, le travail d'un certain nombre de personnes consistait exclusivement à traiter de l'information. C'est le cas, par exemple, des caissiers dans les supermarchés ou des conducteurs de camions. Ces personnes se contentent de traiter de l'information en tournant le volant ou en appuyant sur un bouton de caisse. Tel est également le cas d'autres professions, comme les juristes, les médecins ou les enseignants. Dans mon travail d'enseignant, je ne fais qu'acquérir de l'information pour la transmettre. Les tâches effectuées naguère par des êtres humains peuvent aujourd'hui être épaulées par des machines. La libération du genre humain du travail constitue sans doute une bonne nouvelle, car il est satisfaisant de constater que nous travaillerons moins mais avec une productivité plus grande. De ce fait, tous les systèmes fondés sur la valeur du travail (socialisme, capitalisme...) ne nous seront plus d'aucune utilité pour comprendre comment répartir les richesses au XXI<sup>e</sup> siècle. Des oppositions naîtront sans doute, mais elles ne seront pas similaires à celles que nous connaissions au XX<sup>e</sup> siècle.

La première question éthique qui nous est posée est donc très globale : comment répartir les richesses après la fin relative du travail ?

En deuxième lieu, la petite quantité d'informations que nous sommes capables d'échanger nous avait conduits naguère à inventer des mécanismes d'expression de manière très économe. Par exemple au XIX<sup>e</sup> et au XX<sup>e</sup> siècle, nous ne pouvions donner la parole quotidiennement à des populations isolées au cœur de la France, de sorte que nous avons inventé un système pour leur permettre de s'exprimer une fois tous les trois ou cinq ans par la voie d'un bulletin déposé au fond d'une urne. Aujourd'hui qu'il est possible à chacun de se faire entendre sans l'intermédiation d'un élu, cette notion même de représentation des citoyens n'a plus de signification. Par conséquent, il nous faut inventer de nouvelles institutions permettant à un individu de s'exprimer seul. Nous avons évoqué la consultation qui a été organisée à l'occasion de la loi sur le numérique, mais il s'agit d'un petit pas. Il sera en effet nécessaire d'aller beaucoup plus loin dans la réforme des institutions. Peut-être qu'un jour, le Sénat n'existera plus et qu'il sera remplacé par une meilleure ou une plus mauvaise solution. Il nous appartient d'y veiller en vertu de notre responsabilité éthique.

La troisième transformation très globale concerne l'école. L'école obligatoire a toujours eu pour mission de répondre aux révolutions industrielles de l'énergie. Désormais, la révolution industrielle de l'information nécessite de revoir l'école de fond en comble pour aider les élèves à vivre à l'ère de l'information. Il faut donc compléter les enseignements de physique et de biologie par des enseignements de l'informatique, des humanités ou de l'éthique.

S'agissant des enjeux plus spécifiques de la révolution actuelle, le traitement des données scolaires ou médicales qui concernent les personnes permet de récolter un grand nombre d'informations sur ces dernières. Or les machines ont la capacité de regrouper ces informations sur une longue

---

durée. Les ordinateurs et les systèmes d'intelligence artificielle ont une hypermnésie. C'est pourquoi il faut définir de nouvelles normes pour maîtriser l'ensemble de ces données. Par ailleurs si l'automatisation des traitements médicaux permet aux personnes âgées de gagner en autonomie, cette évolution risque également de les couper de tout lien humain. De ce fait, la mise en circulation d'une machine remplaçant une infirmière nécessite d'inventer de nouvelles formes de liens et de réfléchir au type de tâches que nous souhaitons déléguer aux robots et aux algorithmes.

Une autre nouveauté tient au fait que de longs calculs permettent d'aboutir à des résultats si longs, que nous ne savons pas les expliquer. Par exemple, les ordinateurs calculent tous les jours la température du lendemain sans que nous puissions comprendre comment ils sont parvenus à leur conclusion. En réalité, nous n'avons plus aucun espoir de retrouver notre capacité d'explication, qui était due à notre infirmité pour effectuer de très grands calculs.

Enfin la dernière question éthique et juridique a trait à la création de textes juridiques encadrant les algorithmes et les robots, qui pourraient devenir des personnes morales et des sujets de droit.

Pour conclure, j'observerai que le terme « intelligence artificielle » a été successivement revendiqué par une série de communautés d'informaticiens sans grand-chose en commun, sinon qu'ils demandent à des machines d'effectuer des tâches qui, si elles étaient effectuées par des êtres humains, feraient appel à leur intelligence. Ces différentes communautés et actions de recherche nous ont appris qu'il n'existait pas une, mais des formes d'intelligence artificielle. C'est un mot utilisé pour désigner des choses différentes. De surcroît, la diversité des recherches en la matière ne fait que refléter la diversité des formes d'intelligence que nous attribuons à l'être humain. Nous pouvons bien entendu, continuer d'utiliser le terme « intelligence artificielle » pour définir ces recherches, mais il me semble que nous éviterions nombre de peurs et de fantasmes si nous lui préférons les termes précis qui correspondent à tel ou tel outil : « apprentissage automatique », « reconnaissance des formes », « traitement de la langue » etc. Il apparaît en effet particulièrement important que les différents domaines de l'intelligence artificielle posent les questions éthiques qui leur sont propres, tout à fait différentes les unes des autres le plus souvent.

**Claude de Ganay, rapporteur.** - Sans attendre, je passe la parole à Laurence Devillers, professeur à l'Université Paris-Sorbonne. Vous êtes spécialisée dans les émotions dans les interactions entre les hommes et les machines. Vous animez actuellement un groupe de réflexion sur l'éthique de l'apprentissage automatique au sein de la Commission de réflexion sur l'Éthique de la Recherche en sciences et technologies du Numérique d'Allistene (CERNA).

---

## 2. Laurence Devillers, professeur à l'Université Paris-Sorbonne/LMSI-CNRS

Il y a une très bonne transition à faire après les propos de Gilles Dowek sur la pluralité d'intelligences artificielles. Je vais évoquer les enjeux éthiques de l'apprentissage machine, présent dans de nombreux systèmes d'intelligence artificielle.

Dans le domaine de la robotique, les machines intègrent un grand nombre de systèmes à base d'intelligence artificielle ainsi que d'autres technologies. Au LMSI, laboratoire du CNRS où j'effectue mes recherches, je travaille avec un grand nombre de partenaires sur le projet Romeo 2, consacré à l'interaction avec les robots pour assister les personnes âgées. Pour ce faire, nous menons un grand nombre d'expérimentations sur le terrain et enregistrons des personnes âgées en interaction avec les robots. Nous constatons ainsi l'émergence de multiples sujets éthiques. Pour gérer le dialogue, différentes techniques d'apprentissage machine sont disponibles. Certaines d'entre elles réalisent de très bonnes performances mais sont opaques. Pour une personne en train d'interagir avec ce type de machine, il est absolument impossible de comprendre ce qui se passe. D'ailleurs, je ne pense pas que nous parviendrons un jour à retracer l'ensemble des opérations. Pour autant, il y a une exigence importante d'éduquer davantage la société sur ces sujets. Pour notre part nous, chercheurs dans l'industrie, devons travailler sur la traçabilité et l'explicabilité des systèmes ainsi que sur les évaluations.

L'autre point important, et qui constitue une vraie rupture technologique, tient au fait que ces machines apprennent en fonction de l'interaction avec l'être humain. Il faut donc mettre des garde-fous car ces machines arriveront demain. Le robot Paro, utilisé auprès des personnes âgées dans le cadre de la maladie d'Alzheimer, présente déjà une utilité certaine et crée une forme de lien social. Il est possible de parler à travers une machine, qui peut représenter un médiateur. De plus les personnes qui utilisent les machines, interagiront elles-mêmes avec les patients, ce qui constituera une forme de lien triangulaire.

Je voudrais maintenant assurer une présentation du travail mené au sein de la CERNA en ce moment, qui débouchera sur un rapport en mars 2017 et qui fait suite à un premier rapport sur l'éthique de la recherche robotique en 2016.

Les défis scientifiques de demain supposent de poser les règles morales et éthiques qui assortiront nos implémentations sur les robots. L'un des buts du chercheur en robotique sociale est d'empêcher un déficit de confiance de la part des utilisateurs mais également, de freiner une confiance probable. Les travaux de recherche permettent en effet de constater que les gens projettent sur la machine des intentions anthropomorphiques sur lesquelles il est nécessaire de les détromper. Il est important d'éduquer les utilisateurs sur leurs projections vis-à-vis de la machine, lorsqu'ils ont



---

tendance à lui attribuer davantage de capacités qu'elle n'en a. Toutefois, le fait que la machine apprenne de nous va susciter un attachement et créer du lien. Je travaille beaucoup sur cet aspect de coévolution entre la machine et l'homme, étant rappelé toutefois que des liens d'attachement peuvent également se créer avec sa voiture ou sa montre. Néanmoins, les laboratoires travaillent aujourd'hui sur les interfaces du dialogue, encore très compliquées, et sur les habitudes des personnes pour pouvoir les mémoriser.

Un grand nombre d'initiatives actuelles portent sur l'éthique, ce dont je me réjouis. Je citerai à titre d'exemple celle de l'IEEE ou institut des ingénieurs électriciens et électroniciens, dont le projet de charte de décembre 2016 met en exergue des leviers importants. Il est nécessaire d'éduquer les chercheurs, les journalistes, les industriels et les politiques sur l'éthique. Celle-ci consiste en réalité à définir des règles, à les appliquer à un robot, mettre en œuvre des outils pour vérifier leur respect par ce robot et prévoir un cadre juridique.

L'apprentissage machine qui se trouve au cœur des systèmes est en réalité très opaque car l'utilisateur ne perçoit aucunement les capacités de la machine. De ce fait, il est extrêmement important de comprendre, à différents niveaux, qu'il faut pouvoir évaluer les performances, et de vérifier les données utilisées. On parle beaucoup d'intelligence artificielle, mais en ce qui me concerne je préfère employer le terme de « bêtise artificielle ». En fonction de leur nature, les données fournies à la machine pourront finalement aboutir à un système très désagréable, discriminant ou raciste. Par conséquent, nous devons tous être conscients des risques liés à la manipulation des outils. Lorsque des industriels recourent au *deep learning* pour opérer une classification de leurs clients, j'attire l'attention sur le fait qu'il s'agisse d'une sémantique de surface. Par conséquent, il faut se méfier des performances de ces systèmes, qui vont certes apprendre mais pas toujours avec du sens. L'intelligence artificielle possède des capacités qui lui appartiennent en propre, mais à l'inverse l'homme est doté de sens et d'une intelligence cognitive que n'aura jamais la machine. C'est pourquoi il ne faut pas opérer de confusion ni comparer constamment l'homme et la machine. Les chercheurs travaillent à élaborer un système utile pour les humains, mais nullement à copier l'humain.

Au sein de la CERNA, qui réunit des juristes, des philosophes et des scientifiques en informatique, nous avons mis en évidence les concepts renouvelés d'apprentissage, de responsabilité, d'explicabilité des systèmes et d'évaluation : nous rendrons donc public un rapport sur ces aspects en mars 2017. Par ailleurs, nous avons élaboré des préconisations et des questionnements classés en six thèmes :

- les données des systèmes d'apprentissage ;
- l'autonomie des systèmes apprenants ;
- l'explicabilité des systèmes d'apprentissage ;
- les décisions des systèmes d'apprentissage ;

- 
- le consentement dans le domaine du numérique ;
  - la responsabilité dans les relations humain-machine.

Les mécanismes de contrôle commencent à émerger. Pour élaborer des outils intelligents et autonomes, nous devons nous munir d'outils de vérification et d'évaluation.

La CERNA s'intéresse également à des préconisations générales telles que la formation des enseignants et des étudiants, les plateformes collaboratives de recherche ou la création d'un institut de recherche sur le risque numérique. Sur ce dernier, il importe que la participation soit la plus large possible, et que les grandes institutions de recherche et les industriels en fassent partie.

En outre, j'indique que mon livre à paraître chez Plon le 23 février 2017 sera intitulé « *Des robots et des hommes, mythes, fantasmes et réalités* ».

**M. Claude de Ganay, rapporteur.** - Je passe la parole à Serge Abiteboul, ancien titulaire de la chaire d'informatique du Collège de France, directeur de recherche à l'INRIA, romancier, militant en faveur de l'enseignement de l'informatique et fondateur d'un blog intitulé *Binaire*.

### 3. Serge Abiteboul, directeur de recherche à l'INRIA

J'ai choisi de vous parler d'aspects éthiques dans le cadre des données massives. L'intelligence artificielle est la possibilité pour un logiciel de réaliser une activité qui, chez un humain, demanderait de l'intelligence. Je n'évoque par conséquent pas uniquement le *machine learning* mais l'intelligence artificielle au sens le plus général du terme. Ces logiciels peuvent avoir des actions dont l'importance sur notre vie et nos sociétés est considérable. En corollaire, ils doivent se comporter de manière responsable. À titre d'exemple, un moteur de recherches qui biaiserait ses réponses limiterait notre liberté de choix et pourrait entraîner un déséquilibre du commerce mondial. De même, le lancement d'une fausse nouvelle économique pourrait avoir un impact sur le cours de l'action de la société concernée. Dans ces deux cas, le problème vient de ce que l'information est répercutée sur tous les réseaux et les données sont échangées sans être vérifiées.

Partant de ces constats, nous remarquons que de telles activités intelligentes s'appuient souvent sur la masse de données considérables générées par les réseaux sociaux ou nos téléphones. Grâce à l'intelligence artificielle, nous tentons de créer de la valeur. Les recommandations sur les réseaux sociaux dépendent de la masse de données dont ces systèmes disposent. En définitive, l'accumulation de données et leur analyse ont pris une importance sociétale considérable. Dès lors, il importe que ces données se conduisent de façon responsable.

---

Pour prendre un autre exemple, la décision sur une demande de prêt représente une tâche relativement répétitive sur laquelle un humain pourrait avoir des biais considérables alors qu'un algorithme se comporterait en théorie de façon plus juste. De telles questions d'équité commencent juste à apparaître. Il est indispensable au préalable que le décideur fixe lui-même ce qui est considéré comme juste avant de faire travailler l'algorithme. Une telle décision ne relève pas du rôle de l'informaticien mais revêt un caractère politique.

Cet aspect de vérification est donc essentiel. De même, la transparence est une notion cruciale. Il convient par exemple qu'un opérateur précise à l'intention du public, qu'il s'interdit de vendre des données confidentielles auxquelles il aurait accès. Or ces exigences de loyauté et de transparence ne vont pas d'elles-mêmes dans les systèmes d'intelligence artificielle. Sur une plateforme marchande ou un réseau social si les recommandations proposées à l'internaute sont trop réduites, son point de vue risque d'être modifié. En définitive, les choix à opérer sont éthiques, et non techniques.

Pour ma part je suis optimiste car je pense que les algorithmes peuvent nous permettre d'obtenir de bien meilleurs résultats que par le passé. Il faut cesser de regretter constamment le passé. Par exemple, l'affectation des élèves après le bac fonctionnait très mal avant la mise en place de la plateforme APB. L'utilisation d'un algorithme a abouti à une nette amélioration. Mais il manque encore la transparence. Aujourd'hui, l'algorithme est publié et les critères de choix sont publics. Il devient possible d'améliorer davantage le système dans la transparence.

En définitive, les algorithmes peuvent apporter plus d'équité et de transparence que les humains, à la condition toutefois de consentir des efforts de vigilance. De tels efforts émaneront de l'État et de la réglementation, des associations d'internautes et de la technique. Sur ce dernier point, la technique permettra de gérer les systèmes en les analysant de façon plus efficace, par exemple sur la politique de *privacy*. La technique peut également aider à vérifier si un système se comporte de façon convenable sur le *web*.

Pendant cinquante ans nous avons œuvré avec acharnement pour réussir à travailler sur de gros volumes de données et à les analyser : nous y sommes parvenus aujourd'hui. À mon sens, les années à venir seront déterminantes sur les aspects éthiques. Il faudra tenir compte de ces dimensions dès la conception, pour aboutir à des produits qui seront par exemple *Ethic by design* ou *Fair by design*.

**M. Claude de Ganay, rapporteur.** - Jean Ponce, je vous passe la parole. C'est votre collègue Francis Bach, indisponible, qui nous a proposé de vous inviter. Professeur à l'ENS, passionné de science-fiction, vous vous êtes spécialisé sur les applications permettant de rapprocher les récits de science-fiction des capacités réelles des robots. Vous êtes ainsi devenu l'un des plus grands spécialistes de la vision artificielle, c'est-à-dire des

---

technologies permettant aux machines de se rapprocher des capacités visuelles de l'être humain.

#### 4. M. Jean Ponce, professeur à l'ENS

Contrairement à mes collègues ici présents, je ne suis pas un spécialiste de l'éthique. Personnellement, il me semble que les machines véritablement intelligentes qui en viendraient à se poser des questions éthiques, restent du domaine de la science-fiction. En revanche, j'évoquerai une problématique connexe dans le domaine de l'intelligence artificielle, ayant trait aux attentes et aux risques suscités.

En premier lieu en termes d'attentes, les voitures autonomes ou les robots assistant les personnes fragiles sont des thèmes récurrents de ces dernières années et vont finir par arriver. Les attentes, nées en particulier d'effets d'annonce des grands groupes industriels, suscitent également des craintes et des fantasmes. Cela correspond à un discours marketing qui se comprend, dans la mesure où ces entreprises visent à accroître leurs profits et donc à exagérer les progrès constatés. C'est le scénario de l'intelligence artificielle *Skynet* dans le film *Terminator*. Je ne pense pas que des machines intelligentes prendront de sitôt le pouvoir politique. De même les annonces fréquentes de systèmes obtenant des résultats qualifiés de surhumains, me font en général sourire. Je ne crois pas que les radiologues seront au chômage d'ici deux ans, surtout que les systèmes d'intelligence artificielle produisent encore des erreurs, telles que des faux positifs et des faux négatifs. Je comprends qu'un robot humanoïde comme Atlas fasse peur, il est costaud. C'est une merveilleuse innovation technologique mais cela reste quasiment un gros jouet télécommandé.

À ces attentes peuvent correspondre des déceptions, des craintes et des risques. Un propriétaire de Tesla est récemment décédé au volant de sa voiture. Pourtant dans ces circonstances il faut davantage blâmer les attentes que le résultat. La voiture autonome n'est pas encore là, et des personnes seront toujours présentes pour pallier les éventuels dysfonctionnements. Tesla ou Volkswagen parlent d'aides à la conduite, pas de voiture autonome. Dans les taxis automatiques Uber à Pittsburgh on a maintenu une personne humaine au cas où.

Les programmes d'intelligence artificielle ne sont pas parfaits. Ils donneront des réponses mais avec un taux d'erreur non négligeable. Pour les voitures autonomes, il y aura des accidents mortels, moins qu'avec la conduite humaine mais le niveau d'acceptabilité est différent.

En outre, il est souvent difficile de prédire la fiabilité des résultats de ces programmes et de quantifier les incertitudes. L'accident de Tesla est-il dû au système de navigation ou de décision ? La réponse n'est pas évidente et pose des problèmes de responsabilité et de régulation.

---

Pour autant, en dépit de ces réserves, ma foi dans l'intelligence artificielle est intacte, il ne faut pas en douter.

En matière d'éthique et de données, la protection de la vie privée est d'autant plus difficile que les méthodes modernes sont très gloutonnes en matière de données. De ce fait, les grandes entreprises du *web* disposent d'un avantage stratégique par rapport au reste de l'humanité, même si elles ne sont pas autorisées à partager avec les chercheurs ou à publier les données ainsi récoltées. Cela pose par conséquent la question de l'anonymisation des données. Il conviendra de trouver des solutions d'anonymisation de ces données afin de pouvoir les utiliser.

La question des robots tueurs est un sujet important, qui pose la question de la responsabilité des programmeurs et des décideurs militaires et politiques au-dessus.

Par ailleurs, la question des dilemmes éthiques est difficile à résoudre. Les trois lois d'Asimov sont par exemple très compliquées à mettre en œuvre dans les machines.

En conclusion, je ferai le constat qu'il existe une frontière de plus en plus ténue entre la recherche académique et la recherche privée. Il est impératif que les deux communiquent pour répondre aux besoins de la société.

**Claude de Ganay, rapporteur.** - Serge Tisseron a été l'auteur en 1975 de la première thèse en bande dessinée. Psychiatre et psychanalyste, vous êtes chercheur à l'université Paris Diderot et spécialiste de Tintin et des robots.

### 5. Serge Tisseron, psychiatre, chercheur associé à l'Université Paris Diderot-Paris VII

Comment l'intelligence artificielle est-elle perçue ? Quels sont les aspects problématiques dans cette perception et comment faire évoluer les choses ? Ces considérations, loin d'être anecdotiques, n'ont été que peu évoquées aujourd'hui.

Nous avons beaucoup entendu que l'intelligence artificielle forte n'était pas pour demain. Or il apparaît effectivement que la fantasmagorie actuelle est celle d'une intelligence artificielle beaucoup plus grande que celle qui existe en réalité. Je m'aperçois qu'il ne suffit pas de détromper le public pour le faire changer d'avis. En réalité, l'intelligence artificielle faible angoisse, précisément, parce qu'elle est faible, car de ce fait, elle est manipulée constamment par son constructeur, qui remet à jour quotidiennement le système, chasse les *bugs*, évite les problèmes et donc les procès. L'intelligence artificielle faible est par conséquent perçue comme un moyen par lequel l'utilisateur de la machine n'a pas le contrôle sur cette machine.

---

Un saut fantasmagique prête alors à penser qu'une intelligence artificielle forte et autonome permettra à l'utilisateur de négocier directement avec la machine, même si elle est méchante. Une mythologie autour du libre-arbitre de l'intelligence artificielle s'est donc créée aujourd'hui pour échapper à l'angoisse d'une intelligence artificielle sous contrôle de son fabricant. De ce fait, il m'apparaît important de travailler concrètement sur l'ensemble des repères permettant à l'être humain de garder une différence nette entre la machine et l'humain.

J'évoquerai rapidement deux situations actuelles. Une grande banque (le Crédit mutuel) fait répondre à ses clients par un ordinateur très sophistiqué tel que Watson sans que les appelants soient informés qu'ils ont affaire à une intelligence artificielle. Si ce chemin est poursuivi, les supermarchés seront bientôt équipés de robots-caissiers sans que les consommateurs fassent la différence avec un être humain. Bien entendu cette éventualité est exclue dans l'immédiat, mais il faut anticiper les situations en travaillant d'ores et déjà sur les problèmes concrets. Dans les démocraties, il est nécessaire que le citoyen connaisse toujours l'origine des messages auxquels il est confronté. Dans cet esprit, il est important que toutes les personnes en contact téléphonique avec une machine en soient conscientes. Un message pourrait aviser le client de la façon suivante : *« Vous êtes en contact avec une machine. Si vous préférez parler à un humain, appuyez sur la touche dièse. »*

Le second problème que j'aborderai, vise la situation dans laquelle un individu est en contact avec une machine en le sachant, mais en pensant qu'elle a des capacités équivalentes à celles de l'être humain. Une telle croyance relève de l'anthropomorphisme. Alors qu'il a été si utile au développement de l'être humain, l'anthropomorphisme des machines risque de se retourner aujourd'hui contre l'humain.

L'intelligence artificielle est le concept qui a été retenu, était-ce un bon choix ? Il faudrait se demander pourquoi c'est ce mot qui a été retenu en 1956. Les fondateurs de la discipline n'ont pas vu qu'il allait poser problème ensuite. Il aurait mieux valu parler d'apprentissage automatique ou d'apprentissage logiciel. De même on parle de « machines autonomes » alors qu'il serait préférable d'employer le terme de « machines automatiques », on parle d'« empathie artificielle », alors que mieux vaudrait parler de « simulation d'empathie », on parle, enfin, de « réseaux de neurones » alors qu'on devrait plutôt parler de « réseaux de composants électroniques ». Il faut réviser notre langage, car il risque d'entraîner quiproquos et confusion. Nous assistons de ce fait à une véritable exploitation par les industriels de ce langage anthropomorphe. Les publicités pour les robots au Japon sont très significatives de ce point de vue. Derrière l'idée d'un robot autonome, se joue la volonté d'un certain nombre de commerciaux de nous faire oublier que les machines sont toujours reliées à leurs fabricants pour nous rendre service mais aussi, le cas échéant, nous manipuler.

L'État a un devoir d'informer sur l'ensemble de ces points. Le citoyen doit toujours savoir s'il est en présence d'une machine ou d'un humain, et être protégé des publicités mensongères. Par ailleurs, il me semblerait intéressant que les groupes de travail évoqués par Mme Lemaire comptent également des représentants des diverses religions car ils sont des personnes ressources.

En conclusion, je me réjouis de la somme des travaux menés sur l'intelligence artificielle. Il est très important de protéger l'homme des dangers physiques et matériels que les machines automatiques pourraient lui faire courir. Il est tout aussi essentiel de prévoir des directives sémantiques protégeant l'homme des dangers qu'il pourrait se faire courir à lui-même en raison d'une mauvaise appréciation de la réalité des machines automatiques. Pour coller à l'actualité, je dirai que la décision du Parlement européen de parler de « personnes électroniques » ou de « personnes robots » pour évoquer certains robots sophistiqués vient encore ajouter à la confusion. Tout cela rend difficile un regard réaliste sur ce qu'est l'intelligence artificielle.

---

## VII. CINQUIÈME TABLE RONDE, PRÉSIDIÉE PAR MME DOMNIQUE GILLOT, RAPPORTEURE : DÉFIS JURIDIQUES INHÉRENTS AUX USAGES DE L'INTELLIGENCE ARTIFICIELLE

**Mme Dominique Gillot, rapporteure.** - Cette dernière table ronde va évoquer les défis juridiques posés par l'intelligence artificielle. Mady Delvaux a été retenue à Strasbourg et nous a adressé son intervention, que je vais vous lire.

### 1. Intervention de Mady Delvaux, députée européenne (Luxembourg - groupe S&D), rapporteure du groupe de travail sur la robotique et l'intelligence artificielle

Je tiens tout d'abord à vous présenter mes excuses pour mon absence, malheureusement je ne puis me défaire de mes engagements à Strasbourg.

Je vous remercie pour votre invitation à cette audition publique sur l'intelligence artificielle. Je travaille depuis deux ans au niveau européen sur le sujet de la robotique, thème qui me tient à cœur. Je suis donc ravie qu'un organe national tel que l'Office Parlementaire d'Évaluation des Choix Scientifiques et Technologiques se saisisse de la question

J'espère que mon rapport, voté le 12 janvier dernier au sein de la commission des Affaires juridiques du Parlement européen, suscitera l'intérêt et des débats dans les différents pays de l'Union européenne.

Mon rapport est le résultat de travaux d'un groupe de travail composé de députés de divers horizons politiques et d'experts. C'est une initiative législative, ce qui signifie que le Parlement européen fait des recommandations à la Commission européenne. La Commission pourra décider de donner suite ou non mais devra justifier sa décision. Pour ma part, j'ai bon espoir que notre appel à un cadre juridique en matière de robotique soit entendu : en effet, des drones aux robots industriels, la robotique et l'intelligence artificielle font de plus en plus partie de notre quotidien. La question principale est de savoir comment répondre aux nouveaux défis juridiques et éthiques liés à ces technologies.

Qu'est-ce qu'un robot ?

La première difficulté du rapport est de définir ce qu'est un robot. Le texte concerne les véhicules autonomes, les drones, les robots industriels, les robots de soins ou encore les robots de divertissement. Il ne se penche pas sur les robots pouvant être utilisés comme des armes. On entend par robot une machine physique équipée de capteurs et interconnectée à son environnement dans le but d'échanger et d'analyser des données. Il faut s'attendre à ce que la prochaine génération de robots soit de plus en plus autonome en matière d'apprentissage.



---

Quid de l'engouement médiatique pour la création de la personnalité juridique ou d'un statut légal pour les robots ?

L'émergence de robots de plus en plus autonomes nécessite une réflexion autour de nouvelles solutions. Dans ce rapport, nous demandons à la Commission européenne de se pencher sur certaines pistes.

L'une d'entre elles serait de conférer aux robots une « personnalité électronique » limitée, au moins pour les cas où une compensation est nécessaire. L'idée serait d'assurer que, dans le cas où un robot prendrait une décision autonome dont résulterait un dommage pour une victime, le robot puisse être considéré comme responsable du dommage. Cette notion permettrait de garantir l'indemnisation des victimes sans entraver la volonté d'innovation. Il s'agirait en quelque sorte du même principe que celui dont nous disposons actuellement pour les entreprises. Cette proposition n'est néanmoins envisageable que sur le long terme et, encore une fois, n'est qu'une des possibilités en matière de responsabilité.

Quelles sont les propositions en matière de responsabilité civile en cas de dommage ? Sur qui doit reposer celle-ci ? Le fabricant, le propriétaire ou l'utilisateur ?

Nous sommes face à deux possibilités. Le principe de la responsabilité stricte propose que le fabricant soit responsable car il est le mieux placé pour limiter de potentiels dommages. Si nécessaire, il peut se tourner vers ses fournisseurs.

La seconde option serait de mettre en place une évaluation des risques avec des tests au préalable et une forme de compensation à laquelle toutes les parties pourraient contribuer.

Nous proposons également la création d'un régime d'assurance obligatoire, au moins pour les « gros » robots.

Concernant l'aspect social de la robotique, notamment sur la question de l'attachement émotionnel aux robots de soins, il faut rappeler aux gens que le robot n'est pas un être humain et qu'il n'en sera jamais un. S'il peut montrer de l'empathie, il n'en ressent pas. Nous ne voulons pas de robots qui ressembleraient de plus en plus aux humains, comme c'est le cas au Japon par exemple.

Nous avons donc proposé la création d'une charte visant à empêcher les personnes de devenir émotionnellement dépendantes de leurs robots.

À quel point est-il urgent de légiférer en matière de robotique ?

Pour une fois, nous pourrions établir des principes à l'échelle européenne et un cadre légal commun avant que chaque État membre ne mette en œuvre sa propre législation. Une standardisation en la matière pourrait également profiter à l'industrie : pour rester à la pointe en matière de robotique, l'Europe doit disposer de normes communes.

Sur la question de la responsabilité, les clients doivent être certains de disposer d'une forme d'assurance en cas de dommage. L'enjeu clé est

---

celui de la sécurité, mais aussi de la protection des données : les robots ne fonctionnent pas sans échange de données, ce qui pose la question de l'utilisation de toutes ces informations.

Les robots pourraient créer des emplois dans certains domaines et en détruire dans d'autres, en remplaçant par exemple des personnes peu qualifiées. Comment résoudre ce problème ?

Je pense qu'il s'agit là du plus grand défi pour notre société et nos systèmes éducatifs. Nous ne savons pas à l'heure actuelle comment la situation va évoluer. Je suppose qu'il existera toujours des emplois peu qualifiés. Les robots ne vont pas remplacer les hommes : ils travailleront en coopération avec eux, en les aidant par exemple à transporter des marchandises lourdes.

Nous demandons à la Commission européenne de suivre cette évolution et d'analyser dans quels secteurs l'utilisation des robots détruit des emplois, pour que nous soyons préparés à tous les types de scénarios.

Notre rapport contient également un point controversé qui concerne l'éventuelle instauration d'un revenu universel et une réforme des systèmes de sécurité sociale. Si de nombreuses personnes perdent leur emploi à cause des robots, il faudra leur assurer une vie décente. De la même manière, si nous ne pouvons plus fonder notre système social sur l'emploi, comment le financer ? Nous proposons de le faire via la source de ce changement : en taxant le travail des robots.

Il faut que les citoyens soient les grands gagnants de cette nouvelle ère. Sans assurer que nos propositions soient la panacée, nous voulons étudier toutes les possibilités et ouvrir le débat. C'est pourquoi nous invitons les États membres à réfléchir et je vous remercie de répondre à cette invitation.

**Mme Dominique Gillot, rapporteure.** - J'en termine donc ici avec la lecture de l'intervention de Maddy Delvaux. Je donne la parole à Mme Falque-Pierrotin, présidente de la CNIL, qui en plus de son rôle de régulateur en matière de protection des données, s'est vu confier un rôle de réflexion et d'animation sur les règles éthiques applicables au numérique. La CNIL sera également associée étroitement à la stratégie nationale pour l'intelligence artificielle qui sera lancée demain par le Gouvernement.

## 2. Mme Isabelle Falque-Pierrotin, présidente de la CNIL

Bonjour à tous, et merci de votre invitation. Le sujet de la protection des données personnelles dans le contexte de l'émergence de nouvelles technologies d'intelligence artificielle et de robotique a clairement été identifié par la communauté des autorités de protection des données à l'occasion de leur conférence mondiale de Marrakech en octobre 2016. Lors de cette conférence, nous avons assisté à la présentation d'Helen, robot

---

japonais au travers duquel émergeaient déjà pour nous un certain nombre de questions complexes en matière de protection des données.

De prime abord, la protection des données personnelles ne semble pas aller de soi avec l'intelligence artificielle car les principes attachés à la protection des données semblent aller à l'encontre du fonctionnement de l'intelligence artificielle.

En premier lieu, plus l'intelligence artificielle collecte un nombre très important de données, plus elle est efficace. Or, aujourd'hui, la collecte des données personnelles doit répondre au principe de minimisation des données collectées.

De plus, le principe de la limitation de la durée de conservation des données personnelles se heurte au besoin de mémoire des technologies d'intelligence artificielle qui s'enrichissent au fur et à mesure.

En troisième lieu, la sécurité des données est relativement délicate et souvent imperceptible pour l'individu dans cet univers de l'intelligence artificielle qui fonctionne de manière invisible pour lui, sans couture, à partir de capteurs qui opèrent des arbitrages.

La quatrième question qui me semble enfin devoir être relevée, a trait aux droits des personnes. L'individu doit être à même de comprendre un traitement et d'exercer par rapport à lui des demandes délibérées. Or face à l'intelligence artificielle, il est très difficile d'exercer des droits car il est souvent difficile de savoir auprès de qui les exercer.

Il résulte de tout ceci, qu'en première analyse, nous sommes en présence d'un objet, qui relève pour partie de la science-fiction, mais qui remet tout de même en cause la conception humaniste européenne plaçant l'individu au centre de la politique de protection des données personnelles.

Face à l'intelligence artificielle en perpétuelle évolution, nous sentons que la protection de la personne ne va pas de soi. Il ne faut pas pour autant renoncer à l'intelligence artificielle. Par conséquent, il convient de changer de registre de réponse et de décaler celle-ci.

La question principale sous-jacente est celle de l'autonomie de la personne. Pour progresser, il convient cette fois de cerner les questions éthiques relatives à celle-ci. À cet égard, la CNIL est à la manœuvre dès lors que la loi pour une République numérique de 2016 lui a confié la mission de conduire une réflexion sur les enjeux éthiques et les questions de société soulevés par l'évolution des technologies numériques. C'est donc un domaine nouveau pour la CNIL dont l'objectif n'est pas d'émettre elle-même un avis éthique mais de susciter un processus de discussion collectif que feront vivre tous ceux qui souhaitent y prendre part. Nous avons choisi le thème des algorithmes et de l'intelligence artificielle comme premier thème de réflexion.

Il importe en effet que les organismes concernés soit les initiateurs d'un débat destiné à fixer les contours des questions éthiques. À l'automne,

---

nous restituerons les éléments les plus saillants de ce débat éthique pour, le cas échéant, faire des recommandations de politiques publiques. La réponse que nous entendons ainsi préparer à l'intelligence artificielle, pourrait provenir de l'article 1 de la CNIL - souvent oublié - qui dispose : « *L'informatique est au service des individus.* » Ce principe fondamental et très général a ensuite été décliné en dispositions plus processuelles et de méthode.

Comment pouvons-nous construire une coexistence entre des humains, des robots et des algorithmes ? Comment faire de l'« *humanity by design* » sur le modèle de la *privacy by design* ? Nous possédons des éléments permettant de bâtir un socle. À Davos, IBM a récemment proposé une première charte de bonne conduite concernant les robots. Les initiatives allant dans le même sens pourront être consolidées.

**Mme Dominique Gillot, rapporteure.** - Je passe la parole à Rand Hindi, membre du Conseil national du numérique et pilote du groupe de travail sur l'intelligence artificielle. Votre structure sera elle aussi étroitement associée à la stratégie nationale pour l'intelligence artificielle qui commencera demain. Je vous remercie par avance de votre intervention, que je vous demande assez courte.

**3. M. Rand Hindi, membre du Conseil national du numérique, pilote du groupe de travail sur l'intelligence artificielle, président de SNIPS**

Je suis également à la tête d'une *start-up* en intelligence artificielle depuis quatorze ans. En tant que chercheur et entrepreneur, j'ai donc assisté à l'évolution de ce secteur.

Nous avons beaucoup évoqué la régulation de l'intelligence artificielle, mais je pense qu'il est beaucoup plus intéressant de réfléchir à la régulation de l'accès à la donnée nécessaire à la conception et à l'apprentissage des algorithmes. Dans le domaine du nucléaire, nous avons régulé à la fois les usages et l'exploitation des matières premières radioactives. Nous pourrions imaginer procéder de façon similaire pour l'intelligence artificielle en empêchant à la source de collecter un certain type de données potentiellement négatives.

Une autre question clé dans ce domaine, concerne la *privacy*. En la matière, nous entendons qu'il faut faire un choix entre vie privée et intelligence artificielle. Je pense que c'est un faux dilemme car des technologies de chiffrement nouvelles permettraient aujourd'hui d'effectuer des calculs directement sur la donnée chiffrée. Ces aspects sont étudiés notamment dans le cadre des données de santé, beaucoup trop sensibles pour être exploitées par tout le monde. Je pense que dans les cinq prochaines années le travail sur les données de santé chiffrées sera démocratisé.

---

Par ailleurs, les intelligences artificielles ne font pas le même type d'erreurs que celles des humains. À titre d'exemple, la voiture Tesla a percuté un camion bien visible, alors qu'un humain ne serait jamais entré en collision avec lui. Même si cette erreur de la machine apparaît très grossière, en réalité elle se produit beaucoup moins souvent qu'une erreur humaine. De même dans la finance algorithmique, les erreurs commises sont très peu fréquentes mais ont des conséquences potentiellement beaucoup plus graves que les erreurs humaines.

En définitive, le taux d'erreur des intelligences artificielles est très faible et imprévisible mais les erreurs sont d'une importance bien plus grande que celles d'un humain. Finalement, nous avons vu récemment émerger de nombreuses recherches destinées à modifier le comportement des intelligences artificielles. Par exemple en altérant quelques pixels d'une image, l'intelligence artificielle ne la reconnaîtra plus alors qu'un humain ne verrait aucune différence. Potentiellement, le *hacking* d'intelligence artificielle peut avoir des conséquences très négatives. Une personne mal intentionnée pourrait par exemple introduire des biais pour modifier très légèrement des données-sources et en obtenir ainsi des bénéfices.

Tels sont, je pense, les risques plus subtils de l'intelligence artificielle, mais très difficiles à détecter aujourd'hui.

**Mme Dominique Gillot, rapporteure.** - Olivier Guilhem, vous avez développé au cours de votre expérience professionnelle une expertise précieuse sur les défis juridiques de l'intelligence artificielle et de la robotique. Merci de nous en faire part.

#### **4. M. Olivier Guilhem, directeur juridique chez SoftBank Robotics (ex Aldebaran)**

Je vous remercie de me donner la parole. Plusieurs fois, nous avons évoqué les risques et les craintes liées à l'intelligence artificielle. En ce qui me concerne, je me suis interrogé sur le régime de responsabilité applicable qui pourrait en découler. Il me semble que le régime le plus probable est celui de la responsabilité du fait des produits défectueux.

La doctrine n'est pas unanime en la matière mais une loi de 1998 transpose en droit français une directive européenne de 1985. À cette époque, l'intelligence artificielle n'avait pas la même importance qu'aujourd'hui. Dans une réponse ministérielle de 1998, il est clairement précisé que l'ensemble de la catégorie des biens meubles étaient concernés par cette responsabilité, ce qui inclut les logiciels et donc, l'intelligence artificielle. En d'autres termes, le producteur est responsable du produit qu'il met sur le marché et doit indemniser la victime. Il existe cependant des cas d'exonération lorsque la victime a elle-même concouru au dommage.

---

Si l'on s'appesantit sur les spécificités de l'intelligence artificielle, on s'aperçoit qu'il s'agit d'un système apprenant qui puise, pour apporter une solution avec une certaine autonomie, dans une base de connaissances. Cette base, tout comme l'apprentissage de l'intelligence artificielle, sont potentiellement le fruit des interactions avec l'utilisateur et donc, avec la victime. Cette situation pourrait donc aboutir à une exonération totale ou partielle du producteur de l'intelligence artificielle.

Lorsqu'un produit mécatronique (drone, tablette...) est fabriqué, le fabricant qui y intègre l'intelligence artificielle en supportera juridiquement la responsabilité. Ceci explique qu'*in fine*, le producteur de l'intelligence artificielle soit aujourd'hui très peu inquiet. Ceci pourrait nécessiter un rééquilibrage.

L'autonomie n'est en réalité que la retranscription des valeurs morales du développeur, tout comme le droit est la retranscription des valeurs d'une société à un instant donné. À ce titre, cette autonomie doit être encadrée.

Si l'on considère l'intelligence artificielle comme un bien, un produit au sens juridique du terme, il s'agit d'un produit actif. Comme tout produit actif, il nécessite de prendre des précautions, en travaillant notamment sur les notions de transparence, de neutralité, de droit de modification et de séparation des IA.

Enfin, les risques systémiques liés à l'intelligence artificielle ne sont pas négligeables et doivent être évoqués. Ces risques renvoient à des problématiques d'indépendance nationale et à des problématiques assurantielles. Si une société gérant de l'intelligence artificielle venait à disparaître, des pans entiers de l'économie disparaîtraient consécutivement. C'est pourquoi un système de continuité de service *a minima* pourrait être mis en place pour prévenir certains risques systémiques uniquement, et ce de manière à ne pas exonérer totalement la responsabilité du développeur de l'intelligence artificielle.

**Mme Dominique Gillot, rapporteure.** - Maître Bensoussan, vous êtes connu pour vos positions en faveur de la reconnaissance d'une personnalité juridique des robots. Nous avons voulu vous attendre sur ces propositions, même si elles ne sont pas consensuelles.

### **5. M<sup>e</sup> Alain Bensoussan, avocat, président de l'Association du droit des robots**

Je vous propose de rêver et de réfléchir aux « droits de l'homme de l'intelligence artificielle » qui marqueraient l'avènement d'une métamorphose. La France a exporté ses droits fondamentaux, puis ses droits numériques avec la loi informatique et libertés, elle peut aujourd'hui créer des droits de l'intelligence artificielle qui auraient vocation à se diffuser dans

---

le monde. Les droits de la personne-robot ont de l'importance car les robots vivent avec les hommes.

Je vois quatre défis juridiques essentiels à relever.

Tout d'abord, les algorithmes sont partout et les robots le seront de plus en plus, de sorte qu'il faut nécessairement pouvoir identifier les robots logiciels ou les robots physiques.

Il faut ensuite également repenser la responsabilité en biface. Le robot est toujours responsable vis-à-vis de la victime. Pour le reste, il existe une présomption de responsabilité à la condition d'une certification préalable et sous réserve de fraude.

Bien évidemment, un régime d'assurance devra être prévu lorsque les robots seront parmi nous. En effet, ils entreront bientôt de plus en plus dans les écoles et les entreprises.

Enfin la personne-robot possédant une responsabilité et un régime d'assurance propres, elle doit de ce fait recevoir une personnalité juridique singulière, dans laquelle il faudra par conséquent intégrer le principe de dignité des robots.

Bienvenue à la personne-robot reconnue en tant que telle, et qui devra vivre avec les humains. Dans ce cadre, quatre lignes directrices doivent être conduites pour une cohabitation conforme à nos valeurs humaines entre les hommes et les robots.

La première d'entre elles réside dans la nécessité de permettre d'établir un contrat. Il existe aujourd'hui des contrats entre robots mais également des contrats entre robots et humains. Il me semble temps de penser à un contrat normé où la conception, la conformité, la certification, la garantie et la propriété seraient singulières car les droits actuels sont inapplicables à l'intelligence artificielle telle qu'elle a été décrite.

Le deuxième élément de la mixité est celui de la traçabilité des robots. En traçant les robots, on trace aussi les hommes qui sont à l'origine de leur conception. Il faut donc trouver un point d'équilibre.

Le troisième élément est celui du « principe du bon samaritain », que l'on retrouve dans le droit américain et qui consisterait à revisiter en droit civil quelques règles issues des lois d'Asimov.

Enfin le dernier élément de la mixité entre l'homme et le robot réside dans la possibilité de donner la décision ultime à l'être humain. Ce serait donc un bouton rouge.

## 6. Débat

**M. Sébastien Candel, président de l'Académie des Sciences.** - J'ai été très intéressé par le débat de ce jour. J'ai beaucoup apprécié

---

l'intervention de Gilles Dowek, mais il ne faut pas oublier que nous n'avons pas résolu tous les problèmes liés à l'énergie. Les aspects relatifs à la production d'électricité, notamment, sont encore d'une actualité brûlante. La deuxième révolution industrielle continuera d'être un problème pour l'humanité à l'avenir et ne sera pas réglée par des solutions purement informatiques. De plus, nous ne sommes pas parvenus à l'ère de la fin du travail, qui ne disparaîtra pas par le biais de la troisième révolution industrielle.

**M. Gilles Dowek.** - Je plaide non coupable. Sur les questions de production d'électricité, j'irai dans votre sens. La production d'électricité à partir d'énergies nouvelles sera également cruciale. Je n'ai jamais prétendu que tous les problèmes étaient résolus.

**M. Alexei Grinbaum, CEA-Saclay, LARSIM.** - Un aspect n'a pas été assez développé, me semble-t-il. L'homme qui interagit avec une intelligence artificielle, change lui aussi. L'homme des années 1900 et celui des années 2000 est bien différent. De plus quand nous évoquons les normes, les valeurs et les droits, le sens des grands concepts qui désignent les droits de l'homme change également. La dignité en 2050 n'aura peut-être pas le même sens qu'aujourd'hui. Nous devons tenir compte de cet aspect important.

**Mme Dominique Gillot, rapporteure.** - Nous sommes en effet dans un mouvement perpétuel.

**Mme Gaëlle Marraud des Grottes, journaliste *Revue Lamy de droit civil*.** - Nous avons évoqué très rapidement le concept de transhumanisme et la logique assurantielle.

**M. Olivier Guilhem, directeur juridique chez SoftBank Robotics (ex Aldebaran).** - Effectivement, nous ne parlons pas autant de transhumanisme en France qu'aux États-Unis. Il convient d'être prudent sur ce mouvement de pensée. Le transhumanisme est un outil marketing. Un outil marketing très puissant car il cherche à donner du sens et une vision.

**Mme Laurence Devillers, professeur à l'Université Paris-Sorbonne/LMSI-CNRS.** - La façon dont vous en parlez apporte déjà beaucoup de crédit à ce mouvement philosophique non reconnu qu'est le transhumanisme.

**M. Olivier Guilhem, directeur juridique chez SoftBank Robotics (ex Aldebaran).** - Je suis entièrement d'accord.

**Mme Dominique Gillot, rapporteure.** - C'est pourquoi nous n'avons pas inscrit le transhumanisme pour le moment à nos travaux. Nous serons peut-être contraints d'y réfléchir par la suite.

**M. Olivier Guilhem, directeur juridique chez SoftBank Robotics (ex Aldebaran).** - Nous devons en discuter cependant afin de pouvoir le contester.



---

**Mme Dominique Gillot, rapporteure.** - Nous avons délibérément choisi de ne pas l'évoquer pour nous concentrer davantage sur la compréhension de l'intelligence artificielle.

**M. Gérard Sabah, directeur de recherche honoraire au CNRS.** - Nous devons faire attention à ne pas trop promettre. Nous sommes à l'aube d'une évolution encore embryonnaire, sur laquelle nous devons être prudents pour ne pas susciter de déception. Il faut faire attention à ce que les promesses actuelles n'aboutissent pas à des phénomènes de rejet comme pour la reconnaissance vocale. Nous devons aujourd'hui apaiser le *buzz* pour revenir à un débat beaucoup plus serein.

**Mme Dominique Gillot, rapporteure.** - Merci pour ces recommandations qui vont dans le sens de notre travail, essentiellement à des fins pédagogiques autour de l'intelligence artificielle.

## VIII. CONCLUSION

**Mme Dominique Gillot, rapporteure.** - À titre personnel, je tiens à remercier l'ensemble des participants d'avoir apporté leur contribution sans partir dans de grandes controverses. Nous saluons également la présence d'Axelle Lemaire et la précision de sa pensée. Les engagements qu'elle a pris pour le Gouvernement aujourd'hui serviront demain la Nation.

**M. Claude de Ganay, rapporteur.** - Je salue la qualité des intervenants. Nous espérons que l'OPECST contribuera à l'effort pédagogique autour de l'intelligence artificielle grâce à l'ensemble des auditions qu'il mène actuellement.

## ANNEXE 10 : COMPTES RENDUS DES AUDITIONS BILATÉRALES CONDUITES PAR LES RAPPORTEURS

### I. AUDITIONS DU 25 OCTOBRE 2016

#### 1. Stéphane Mallat, professeur à l'École normale supérieure (ENS), chercheur en mathématiques appliquées

Une véritable révolution est à l'œuvre sous les effets produits par le développement du *Big Data* et de l'intelligence artificielle, dont les conséquences sont encore très mal perçues dans la société civile. À travers l'analyse de données, il est possible de faire des choses qui étaient, pour la plupart des scientifiques, inimaginables il y a cinq ou dix ans. **Cette révolution résulte de la conjonction du regroupement de très nombreuses données dans des bases de données et des capacités de calcul qui ont connu une évolution algorithmique très importante.**

Cette révolution a des conséquences énormes pour l'industrie française. Il y a eu, dans un premier temps, une forme de retard en France, l'industrie numérique s'étant essentiellement concentrée aux États-Unis d'Amérique, qui a de fait attiré de nombreux talents français. Il est donc nécessaire pour la France de rattraper son retard à l'amorçage, qui est classique, mais surtout de récupérer des talents au niveau industriel.

Au-delà de l'industrie, **d'autres domaines sont concernés de manière cruciale par cette révolution.** La grande difficulté pour la médecine est que les données ne sont pas disponibles. Cela soulève un **paradoxe étrange** : **d'une part, toutes les données personnelles mises sur Internet sont librement à disposition pour des utilisations marketing et commerciales ; d'autre part, au niveau de la médecine, il est extrêmement difficile d'obtenir les données alors qu'il y aurait un impact sociétal extraordinairement bénéfique et important à les exploiter.** Ce paradoxe s'explique, au niveau de la médecine, par des problèmes fondamentaux de confidentialité et de sécurité ; cependant, tout ce qui est commercial et financier est laissé libre et évolue très vite, alors que tout ce qui est médical n'avance pas, alors même que ce serait le domaine le plus bénéfique pour la société.

Il est également crucial de prendre en considération l'enjeu de régulation et de recherche d'équilibre dans l'utilisation des données. Les enjeux de protection de la vie privée sont extrêmement importants, et ils vont le devenir de plus en plus. Il est néanmoins nécessaire d'équilibrer le rapport existant, car il ne faut pas empêcher le médical de se développer, le principe de précaution freinant grandement le domaine médical alors que tous les autres domaines évoluent très vite.

---

Les individus restent souvent sur la perception de l'intelligence artificielle des années 1980 qui était fondée sur une idée de règles et de logiques. **L'intelligence telle qu'elle est construite actuellement relève davantage de l'ordre de l'analogique, qui consiste à des mises en relation entre des éléments analysés.** Les problèmes complexes devant lesquels il est possible de se retrouver, que ce soient des problèmes de reconnaissance d'image, de son, de décisions politiques, sont des problèmes qui résultent de l'interaction d'énormément de variables. Ce ne sont pas des règles qui régissent ces variables ; c'est une forme d'agrégation progressive de ces variables qui amène vers la solution la plus probable, à l'image de signaux émis par des milliers d'indicateurs faibles dont l'agrégation permet d'obtenir une information forte. **Il est de fait impossible de savoir pourquoi un système d'information a pris cette réponse ou décision plutôt qu'une autre.** Les réponses des intelligences artificielles sont le résultat de l'agrégation d'énormément d'indicateurs ; c'est bien plus proche de l'analogie que de la logique et des règles.

L'intelligence artificielle est une opportunité pour la médecine, pour l'énergie, pour l'évaluation des politiques publiques, pour le développement d'innovations, etc. Cependant, **la protection de la vie privée devra être défendue de manière intelligente, afin de ne pas bloquer les effets positifs apportés par l'intelligence artificielle.** Néanmoins, la privauté et la liberté individuelle doivent être préservées. Par exemple, il est nécessaire d'éviter que des systèmes de prédiction totalement ouverts mettent les individus dans des cases et les empêchent d'en sortir, tout en encourageant l'ouverture des données à la médecine afin que cette dernière puisse progresser et se développer. Bien que l'introduction de nouveaux types de systèmes à intelligence artificielle implique nécessairement une prise de faible risque, les bénéfices qui en seront retirés au niveau médical sont considérables. Un équilibre juste et bénéfique doit donc être trouvé.

Du côté industriel, une difficulté tend à émerger au niveau national et européen étant donné le fait qu'il n'existe aucun acteur majeur qui stocke Internet. **Les individus et acteurs de l'économie sont conscients de l'ampleur de la révolution numérique, et l'enjeu sera d'avoir des champions industriels.** L'absence d'acteurs français et européens provoque une fuite des cerveaux et des chercheurs vers les États-Unis d'Amérique. La recherche est d'abord valorisée par le tissu de petites *start-ups* dynamiques qui commencent à émerger en France ; celles-ci se font cependant rapidement racheter par des acteurs nord-américains lorsqu'elles émergent. Il est donc essentiel qu'une politique industrielle aide à l'émergence de géants industriels dans ce domaine afin qu'ils puissent absorber la technologie. Les grandes entreprises du numérique, telles que Huawei ou Facebook, s'installant à Paris, parviennent à capter tous les talents français en particulier car il n'y a pas de plus grosse concurrence au niveau national.

---

## 2. M. Patrick Albert, entrepreneur (créateur de ILOG), chercheur et pionnier dans le domaine de l'intelligence artificielle

La France dispose d'un écosystème très favorable à l'émergence de jeunes pousses du numérique, notamment grâce au crédit d'impôt recherche, les pôles de compétitivité, les formations et écoles de l'enseignement supérieur, les sociétés d'accélération du transfert de technologies, les incubateurs d'entreprises, etc. Depuis quelques années, il s'est opéré un changement de discours, de communication et d'appréciation sur l'innovation, ainsi que des outils mis à disposition des entrepreneurs, pour favoriser le développement de l'économie numérique en France.

**L'accélération du développement de l'intelligence artificielle s'est faite en trois vagues.** La **première vague** correspond à **l'apparition des systèmes experts**, qui sont des logiciels de production des raisonnements et réponses à partir de règles et de faits. Cependant, l'informatique n'était pas prête pour ces systèmes, car ces derniers étaient très complexes.

Au cours des années 2000, **un renouveau s'est opéré avec la simplification des systèmes experts, qui sont devenus accessibles et donc commercialisables et utilisables.** Cette vague d'automatisation des décisions est arrivée à maturité à la fin des années 2000, et a donné lieu à de nombreuses acquisitions. Ce sont principalement des sociétés de *B2B*<sup>1</sup>, telles qu'IBM ou Oracle, qui ont acheté des sociétés produisant des systèmes automatisés disposant d'un niveau d'intelligence artificielle « symbolique », c'est-à-dire assez simple, ainsi que des logiciels d'optimisation, aux algorithmes plus complexes, qui permettent de prendre des décisions tirant le meilleur parti des ressources à disposition.

Les industries qui se sont « mondialisées » précocement ont eu une pression de compétition féroce, et ont donc dû baisser leurs coûts. Ces industries, notamment dans les transports, ont de fait eu recours à ce type de logiciels afin de faire face à la concurrence en utilisant des techniques de *lean management*<sup>2</sup>.

La **troisième vague** a été **l'avènement de l'apprentissage automatique** (*machine learning*). Les résultats des recherches en apprentissage automatique, suite à leur publication, ont été reçus très positivement par le grand public. Le cas du développement de la voiture autonome s'inscrit dans cette perspective : le grand public comprend son fonctionnement, son intérêt et ses enjeux notamment en matière de sécurité, mais également les dilemmes éthiques que l'utilisation de véhicules autonomes pose.

La **quatrième vague** pourrait être **celle des agents intelligents, en particulier grâce à l'Internet des objets**. Cette quatrième vague pourrait être

---

<sup>1</sup> En anglais : *business to business*. Ce terme désigne l'ensemble des activités commerciales nouées entre deux entreprises.

<sup>2</sup> Méthode de management qui vise l'amélioration des performances de l'entreprise recherchant les conditions idéales de fonctionnement de manière à ajouter de la valeur avec le moins de gaspillage possible.

---

favorisée par le développement de l'intelligence en essaim (*swarm intelligence*), dont les méthodes se fondent sur les modèles de colonie de fourmis ou d'essaims d'abeilles grâce auxquels des agents peu doués individuellement parviennent collectivement et en s'organisant à des résultats extraordinaires. **Cette quatrième vague incarne un besoin d'automatisation des actions sur Internet**, à l'image de l'apparition de nombreux agents conversationnels (*chatbot*) sur les sites Internet marchands, intervenant lors de l'acte de vente ainsi qu'en support.

**L'enjeu, pour les parlementaires, est la réglementation de l'intelligence artificielle**, en prenant en compte les problèmes sociaux et sociétaux qui risquent d'être critiques. La maîtrise de l'intelligence artificielle a la même envergure stratégique pour la souveraineté nationale que la maîtrise de l'atome l'était après-guerre. Cependant, l'intervention du politique est également essentielle pour que les bouleversements économiques et technologiques déclenchés par l'intelligence artificielle n'aboutissent pas à des crises sociales fortes, comme ce fut le cas au début de la première révolution industrielle. L'intelligence artificielle provoque une restructuration majeure de la création de valeur, profitant en premier lieu à ceux qui sauront maîtriser l'intelligence artificielle.

## II. AUDITIONS DU 8 NOVEMBRE 2016

### 1. M. Marc Mézard, directeur de l'École normale supérieure (ENS)

Une des grandes évolutions à prévoir sera l'impact des sciences informatiques, de l'intelligence artificielle et de la science des données sur le vaste champ des sciences humaines et sociales, notamment sur les aspects éthiques. De nombreux autres domaines, comme la médecine, seront également concernés par étape par l'intelligence artificielle.

**L'intelligence artificielle est un domaine dans lequel la frontière entre la recherche fondamentale et la recherche appliquée dans le monde industriel est très poreuse.** Les problèmes posés par les industriels sont des problèmes de recherche fondamentale. Il existe un besoin énorme de « bras » à fournir pour le tissu scientifique et le tissu économique, les établissements d'enseignement supérieur devant relever le défi de fournir ces « bras » à la fois du côté de la recherche et du côté de la formation.

**Le problème de l'emploi face aux enjeux de l'intelligence artificielle est à la fois économique et européen.** Il est impératif que s'enclenche une dynamique universitaire qui permette d'attirer et de stabiliser des jeunes talents pour alimenter les entreprises européennes du numérique. L'effort fourni actuellement doit s'intensifier concernant l'intelligence artificielle. **Le développement de l'intelligence artificielle n'est pas une incrémentation ; c'est un véritable changement de paradigme**

---

**d'un point de vue scientifique et industriel.** En cela, la question du développement des méthodes des algorithmes est fondamentale. Sur ce plan méthodologique, la France est très bien placée. Cependant, de nombreux autres enjeux cruciaux doivent être pris en compte dans la recherche, notamment concernant les données et l'éthique.

Bien que les champs disciplinaires puissent éventuellement être redécoupés compte tenu des changements liés à l'intelligence artificielle et aux mégadonnées (*Big data*), il demeure que l'émergence de nouveaux champs disciplinaires est incrémentale et nécessite le développement de nouveaux codes culturels. Cela implique de créer des instituts ou des centres où cohabitent des étudiants et chercheurs de différents champs disciplinaires, permettant une émulation culturelle favorisant l'interdisciplinarité et l'émergence de nouveaux champs disciplinaires.

Cependant, l'émergence d'un nouveau champ disciplinaire ne peut être décrétée ; elle est conditionnée par le temps nécessaire pour la formation. Il est important que les scientifiques restent attachés à leur discipline initiale qui possède ses questions d'intérêts et d'évaluation. Si les scientifiques sont détachés de leur communauté, l'interdisciplinarité et la création de nouveaux champs disciplinaires ne peuvent pas fonctionner. En revanche, l'interdisciplinarité et l'apparition de nouveaux champs académiques pourraient être favorisés par la mise en avant de l'importance du sujet de l'intelligence artificielle et l'élaboration des projets nécessitant des compétences interdisciplinaires, au travers d'objectifs scientifiques et politiques.

Dans cette perspective, **la participation de la puissance publique est essentielle.** Encore une fois, l'apparition de nouveaux champs ne se décrètera pas mais s'effectuera au travers d'initiatives, de projets et de programmes scientifiques proposés notamment par les pouvoirs publics qui devront accompagner ces évolutions majeures.

**L'abaissement de la barrière entre recherche fondamentale et recherche appliquée en entreprise est très perceptible dans le domaine des jeunes entreprises émergentes du numérique.** La barrière à franchir pour créer une entreprise dans le domaine de l'intelligence artificielle et, de manière plus générale, dans le domaine des sciences et technologies de l'information et de la communication, est beaucoup moins élevée que celles existant dans des processus industriels traditionnels. En cela, il n'est pas rare que des anciens élèves de l'ENS, qui n'est pourtant pas spécifiquement tournée vers le monde de l'entreprise, intègrent des incubateurs et fondent des entreprises tournées vers les sciences et les technologies numériques.

---

## 2. M. Raja Chatila, directeur de recherche au Centre national de la recherche scientifique (CNRS), directeur de l'Institut des systèmes intelligents et de robotique (ISIR)

Le fonctionnement de la robotique présente certaines différences par rapport à celui de l'intelligence artificielle. Dans la perspective de l'étude de la robotique et des systèmes intelligents, **la machine est considérée comme matérielle, et non un logiciel effectuant des opérations ou un algorithme fonctionnant à partir de données.** L'utilisation du qualificatif « matérielle » signifie que **la machine physique interagit directement et de manière automatique avec son environnement.** La différence principale existant entre la robotique et l'intelligence artificielle est issue de la complexité de l'interaction avec la réalité qui implique que les problèmes traités ne sont pas totalement abstraits dans un monde virtuel ; la réalité est porteuse de complexité, d'incertitude, de problématiques liées à la perception d'un environnement complexe qui évolue indépendamment du système et de difficultés d'interactions physiques et matérielles.

De fait, la problématique de la robotique est plus vaste que celle de l'intelligence artificielle, puisque la recherche en robotique doit tenir compte de la perception, de l'action et du mouvement. Les machines sont munies de capteurs afin de percevoir l'environnement, et sont également munies d'actionneurs pour se déplacer et agir sur cet environnement. Ces capteurs, qui extraient des données, doivent permettre de transformer ces données en connaissances pour que la machine puisse décider et agir. L'acte de décision et la capacité d'apprentissage sont permis par des logiciels, impliquant que les ordinateurs sont physiquement des parties prenantes de la machine. **Une machine peut donc disposer de capacités de perception, de mouvement, de prises de décision pour décider de ses actions, de communication et d'interaction pour évoluer dans un environnement humain, et d'apprentissage qui lui permettent d'appréhender la complexité du monde.**

Le robot est le paradigme de l'intelligence artificielle « incorporée » - je préfère ce mot à incorporée, cela correspond plus à l'anglais *embodied*. Le robot a besoin de la prise de décision et des capacités d'anticipation et d'action d'une intelligence artificielle pour construire ses raisonnements sur les moyens et les contraintes qui permettent ses mouvements. Du fait de la complexité du monde, le robot est construit de manière simple et son système peut se complexifier ensuite au fil de son apprentissage.

La recherche en robotique a débuté de deux manières différentes. D'un côté, la recherche en robotique a été motivée par l'industrie, et plus particulièrement l'industrie automobile qui avait besoin de placer sur ses chaînes de production des robots destinés à effectuer des mouvements répétitifs et rapides. De l'autre côté, les premières réflexions menées sur la notion d'intelligence artificielle et la question posée par Alan Turing sur la capacité des machines à penser ont entraîné le lancement de programmes de recherche en intelligence artificielle et, à partir des années 1960, en

« robotique intelligente ». Ce n'est que dans les années 1980 que la recherche en robotique et la recherche en intelligence artificielle se sont distinguées, la robotique s'intéressant davantage à l'interaction de la machine avec son environnement et aux représentations probabilistes. Elles se rapprochent à nouveau dans la période récente.

En matière de sécurité, les robots comprennent toujours un bouton rouge, c'est bien mais c'est déjà trop tard, il faut tout prévoir pour ne pas en arriver là. L'état du système doit être constamment observé et il faut pouvoir détecter toute déviance avant l'arrivée des problèmes.

Sur le projet transhumaniste, il faut voir que derrière ces discours, nous avons des vues de l'esprit qui n'ont rien d'opérationnelles, elles sont en réalité des idéologies, qu'on cherche à imposer pour gommer les différences entre l'humain et le non-humain. En tant que roboticien, je ne peux que m'opposer à cela.



### III. AUDITIONS DU 9 NOVEMBRE 2016

#### 1. M<sup>e</sup> Alain Bensoussan, avocat, président de l'association pour les droits des robots et M<sup>e</sup> Marie Soulez, avocate spécialisée sur les TIC dans son cabinet

Un robot peut être défini comme une machine intelligente capable de prendre des décisions de manière autonome, maîtrisant son environnement et pouvant le modifier. Un robot est un composé d'algorithmes doté de capteurs et d'actionneurs. De fait, il faut distinguer les robots logiciels des robots physiques : les robots physiques sont des robots logiciels qui peuvent modifier leur environnement grâce à leurs capteurs et leurs actionneurs. La majorité des robots actuellement en circulation sont des robots logiciels ; cependant, en 2020, ils auront tous migré dans des coques dotées de capteurs, faisant d'eux des robots physiques.

Il existe un besoin juridique fort concernant le développement de la robotique à usage non-militaire au niveau mondial. Certaines voies permettant de répondre à ce besoin furent identifiées dans le projet de rapport de la commission des affaires juridiques du Parlement européen contenant des recommandations à la Commission concernant des règles de droit civil sur la robotique rendu le 31 mai 2016. J'identifie pour ma part les points suivants, qui représentent un caractère prioritaire pour un développement juridiquement serein de l'intelligence artificielle et de la robotique.

- **Identification des robots.** Les robots doivent être facilement identifiables, « immatriculés », par un nom, un prénom, un numéro IP, un numéro de contact, etc. Ce besoin impératif d'identification correspond au fait que les robots peuvent être considérés comme une espèce à part entière. En ce sens, il est nécessaire de créer un fichier centralisant les identités de chaque robot afin de garantir le contrôle humain sur leurs activités.
- **Définition des principes de fabrication.** Il est essentiel que, comme l'a affirmé Bruno Maisonnier, P-DG d'Aldebaran et créateur du robot Nao, que les robots soient conçus « éthiques *by design* », des robots éthiques par nature. Pour cela, il faudrait notamment produire des chartes éthiques qui seraient imposées aux industries pour garantir cette fabrication éthique *by design*. Ces chartes éthiques pourraient par exemple comporter l'intégration dans le code source des robots les trois lois de la robotique<sup>1</sup> formulées par Isaac Asimov dans Cercles

---

<sup>1</sup> Un robot ne peut porter atteinte à un être humain, ni, en restant passif, permettre qu'un être humain soit exposé au danger ; un robot doit obéir aux ordres qui lui sont donnés par un être

---

*vicioux* (1942) et en les complétant, ces trois lois n'étant pas suffisantes. Si ces trois lois de la robotique comprennent de grandes limites, il est néanmoins nécessaire de s'en servir comme règles d'orientation.

- **Un droit des robots mondial.** Il serait faux de penser que, du fait de son caractère mondial, tout encadrement juridique national du développement industriel des robots et de l'intégration des robots dans la société serait inutile et inefficace. Si un droit international des robots peut annuler ce type d'effet, un droit national des robots pourrait aussi bien devenir un standard mondial, à l'image des lois sur l'informatique et des lois sur la protection des logiciels.
- **Certification des produits robotiques.** Il est possible d'envisager deux types de certifications : la certification par un contrôle de vérification des capacités et exerçant un contrôle au travers d'une commission, à l'image de Bureau Veritas par exemple ; ou une certification par déclaration. Il sera impératif de certifier les robots, et plus particulièrement les robots humanoïdes.
- **Documentation des algorithmes (*accountability of algorithms*).** Tous les algorithmes doivent être documentés et disponibles au contrôle.
- **Traçabilité des robots.** Assurer une traçabilité des robots est cruciale, afin qu'ils n'aient pas une complète autonomie. Les robots ne doivent pas pouvoir agir de manière anonyme ; il faut tracer leurs actions avant, pendant et après l'exécution de ces actions. La conservation des données tracées pourrait se limiter à une semaine, dans un souci de protection des données personnelles (car les robots ne vivent pas seuls). La traçabilité doit permettre d'identifier les erreurs et dangers (liés à la circulation de voitures autonomes notamment) tout en assurant la protection des données personnelles.
- **Responsabilité des robots.** Il est important de distinguer la responsabilité vis-à-vis de la victime et la chaîne de responsabilité. Concernant la responsabilité vis-à-vis de la victime, il faut appliquer une responsabilité sans faute aux robots. Pour une fonction donnée, le robot est supérieur ; mais le robot est inférieur sur le plan multifonctionnel. Les robots entraîneront moins d'erreurs que les humains sur une fonction donnée. De fait, le robot assume une responsabilité dommage vis-à-vis de la victime. La législation du Nevada qui, sur le plan de la régulation des voitures autonomes *NRS 482A* notamment, est la plus avancée au monde, dispose qu'est responsable la plateforme d'autonomie ; il est possible d'aller plus loin en affirmant la responsabilité du robot. L'humain est responsable car

---

*humain, sauf si de tels ordres entrent en conflit avec la première loi ; un robot doit protéger son existence tant que cette protection n'entre pas en conflit avec la première ou la deuxième loi.*

il est libre ; la responsabilité vient de la liberté. Les personnes morales, elles-mêmes libres, se sont vu attribuer une responsabilité. De fait, il faut introduire dans la chaîne de responsabilité une responsabilité singulière pour les robots.

- **Modification de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés.** La modification de la loi informatique et liberté est nécessaire, notamment concernant les compagnons personnels. La loi n° 78-17 et le règlement (UE) 2016/679 excluent les activités purement personnelles ou domestiques : ainsi, un robot domestique n'est pas soumis aux dispositions de ces deux textes. Il est donc nécessaire de modifier ces textes afin de prendre en compte la robotique personnelle.
- **Création d'un droit à l'intimité numérique.** Il existe une co-intimité entre le robot compagnon et son utilisateur qu'il faut protéger.
- **Création d'un archivage des cerveaux robotiques.** Il serait important d'archiver les cerveaux des robots, sur le modèle des archivages bibliothécaires.
- **Droit à la transparence des algorithmes.** Il est essentiel, non pas de comprendre l'algorithme, mais de comprendre ce qu'il fait faire au robot. Le robot ne doit pas pouvoir agir de manière opaque.
- **La loyauté des robots.** Le robot ne doit pas se comporter de manière déloyale vis-à-vis de son utilisateur.
- **Création d'un droit à la compréhension.** Le droit d'information et d'accès aux algorithmes est déjà prévu par la loi de 1978 ; il est cependant nécessaire de la renforcer par l'instauration d'un droit à la compréhension.
- **Personnalité juridique robot.** Il est nécessaire de donner une personnalité juridique aux robots, afin qu'ils puissent assumer leurs responsabilités devant la loi. La création d'une personnalité juridique robot est essentielle. Les robots jouissant d'une certaine autonomie, le droit des objets ne pourra pas être appliqué.
- **Création d'un commissariat aux algorithmes.** La création d'un commissariat aux algorithmes permettra la protection des données et des règles de fabrication et d'utilisation imposées aux industries.
- **Création d'un contrat-type d'assurance pour les robots.**

- **Instauration d'une politique de label et de normes.** L'instauration d'une politique de label et de normes permettrait aux industriels de développer leur activité en minimisant les contraintes.
- **Création d'un comité d'éthique.**

## **2. M. Henri Verdier, directeur interministériel du numérique, ancien entrepreneur et spécialiste du numérique**

**La plus grande valeur dans l'économie numérique est l'intelligence partagée des individus, l'intelligence de la multitude.** Les « GAFAMI »<sup>1</sup>, ainsi que les autres entreprises majeures du numérique telles que Uber, parce qu'ils maîtrisent les stratégies de plateformes, suscitent, stimulent, concentrent et captent à leur profit l'intelligence de la multitude. **L'économie numérique croît grâce à la valeur de la multitude.**

**Il n'y a pas nécessairement de frontière nette entre l'intelligence artificielle « forte »,** qui aurait un comportement quasi-humain, **et les systèmes experts** qui s'apparentent davantage à de l'intelligence artificielle « faible » et qui se nourrit de l'apprentissage profond (*deep learning*) et des mégadonnées (*Big data*). Les mutations actuelles relèvent davantage de la progression d'une intelligence systémique à tous les échelons. **Il apparaît plus important d'envisager les évolutions de l'intelligence artificielle comme un continuum.** Il y aura peut-être quelques pointes extrêmes dans l'avancée technologique de l'intelligence artificielle, comme celle d'un ordinateur battant l'intelligence humaine aux échecs, mais il est encore plus important de réaliser que de nombreux processus seront informatisés. Les petites forces seront peut-être plus importantes que les prouesses médiatiques de l'intelligence artificielle. Par analogie, le fait que l'essor prometteur de véhicules autonomes soit imminent n'empêche pas le fait que, chaque année, les constructeurs automobiles rajoutent des fonctions d'autonomie dans les voitures.

Il existe de nombreuses inquiétudes s'exprimant autour du développement de l'intelligence artificielle. Au-delà du fait qu'il y a une grande concentration des puissances des capacités, **il n'est pas certain que l'humain puisse contrôler de manière complète les dynamiques internes d'une intelligence artificielle.** Les procédures de régulation par la norme ne seront pas suffisantes pour contrôler les actions d'une intelligence artificielle. Quelques sécurités ne permettront pas non plus d'empêcher une intelligence artificielle de s'émanciper du contrôle humain. Une intelligence artificielle peut être auto-apprenante et innovante, et il faut donc être très vigilant dans la manière dont elle sera contrôlée.

---

<sup>1</sup> Acronyme désignant les entreprises de l'économie numérique nord-américaines les plus importantes en termes de chiffre d'affaire : Google, Apple, Facebook, Amazon, Microsoft et IBM.

---

Du fait de l'intelligence distribuée, des codes sources ouverts (*Open source*), et de la logique des organisations de l'écosystème numérique, **l'intervention de l'État dans le développement technologique et industriel de l'intelligence artificielle ne peut se faire au travers d'une planification d'objectifs, mais avec la stimulation de l'innovation.** La Silicon Valley ne peut être rattrapée grâce à des décrets.

**Il manque la diffusion d'une pensée éthique.** L'avènement de systèmes d'intelligence artificielle va amener des questions neuves, et les solutions proposées ne font pas nécessairement consensus. Par exemple, le développement des voitures autonomes soulève la question du comportement adopté par le système informatique pilotant la voiture lorsque celui-ci sera confronté à un danger imminent d'accident mortel soit pour son passager, soit pour un usager plus vulnérable. Pour le moment, les nombreuses solutions proposées en termes de responsabilité d'usage et d'éthique n'obtiennent pas le consensus. En cela, une pensée scientifique, et aussi une pensée sociale, sont nécessaires pour doter le développement de l'intelligence artificielle d'un cadre éthique. L'efficacité n'est pas une éthique ; si un algorithme remplace et améliore le travail humain, il sera essentiel de vérifier qu'il ne crée par des effets indirects d'exclusion. **Étant donné que l'intelligence artificielle s'invente actuellement dans un modèle économique simple, il n'y a pas de cas de conscience car il ne s'agit que de l'optimisation de techniques commerciales.** Le rôle de l'État, pour sa part, est de constamment mettre en balance l'égalité, l'équité, la justice, la croissance et le chômage ; à l'heure actuelle, les inventeurs d'algorithmes ne sont pas formés à prendre de tels facteurs en considération dans la construction de leurs algorithmes.

### **3. M. Laurent Alexandre, président de DNA Vision, fondateur de Doctissimo, chirurgien-urologue**

Le mouvement actuel en matière d'intelligence artificielle est que **l'on se dirige vers un duopole entre les géants chinois et nord-américains.** Il n'existe pas en France et en Europe d'acteurs de l'industrie du logiciel à la hauteur des géants chinois et nord-américains. **La France et l'Europe** ont des jeunes pousses prometteuses, un tissu industriel traditionnel et des chercheurs performants, mais elles **sont reléguées au rang de consommateur et non de producteur de technologies.**

**La France et l'Europe** sont des « colonisées numériques », parce qu'elles ne produisent pas de richesses à partir du numérique, privilégiant une approche appliquée à l'informatique, c'est-à-dire une approche de vente. Le véritable enjeu concernant la révolution numérique et l'émergence de l'intelligence artificielle n'est pas le code, mais la donnée. L'apprentissage d'une intelligence artificielle nécessite beaucoup de données, mais peu de logiciels. C'est en cela que **la France est en train de manquer le virage de**

**l'intelligence artificielle** : la France possède les codeurs, mais elle n'a pas les données. Ce sont les géants du numérique et de l'intelligence artificielle, tout particulièrement aux États-Unis d'Amérique, qui détiennent les données. L'ensemble des données que tout utilisateur de services numériques connectés fournit quotidiennement nourrit et éduque gratuitement les intelligences artificielles des industries numériques chinoises et nord-américaines qui se trouvent en situation oligopolistique.

L'apprentissage du codage, notamment par les enfants, est une chose positive ; cependant, dans quinze ans, l'intelligence artificielle codera mieux et plus vite que les codeurs de niveau moyen. Il est important de donner une culture générale du code, mais **le codage humain médiocre ne résistera pas à l'intelligence artificielle au-delà de 2025.**

**Il est essentiel de mettre en place des plate-formistes de l'intelligence artificielle en Europe.** L'Europe doit retrouver de l'oxygène entre les géants numériques chinois et nord-américains. Il est donc vital qu'apparaissent des plate-formistes européens, des acteurs qui produiront et recueilleront de gros volumes de données. Il serait trompeur de raisonner par branche concernant le développement de l'intelligence artificielle ; **il ne faut pas « verticaliser » les approches de l'intelligence artificielle selon les secteurs, mais les « horizontaliser ».**

En ce sens, **il est impératif d'unifier le droit de l'information et la jurisprudence des organismes de CNIL européens**, qui ne doivent pas exiger un objectif et imposer des limites au traitement des données. **Un algorithme ne peut pas être traçable car il change en continu au fil du traitement des données ; un choix doit donc être effectué : l'émergence d'une industrie européenne de l'intelligence artificielle sera toujours bridée par la volonté de traçabilité des algorithmes.** La régulation trop importante peut étouffer dans l'œuf l'émergence d'une industrie européenne de l'intelligence artificielle et laisser un pouvoir oligopolistique dans les mains des industries chinoises et nord-américaines.

**L'avènement de l'intelligence artificielle pourra entraîner des problèmes de requalification des industries dont la tâche sera remplacée par l'intelligence artificielle.** Le populisme va croître car il n'existe actuellement pas de solutions et perspectives offertes aux actifs peu ou moyennement qualifiés, menacés par l'intelligence artificielle.

Il est essentiel de réorienter l'appareil productif rapidement. Pour l'instant, les systèmes d'intelligence artificielle sont des intelligences artificielles faibles qui ne sont pas spontanément complémentaires de l'humain ; le combat social sera le combat pour la complémentarité. Cependant, la complémentarité ne sera que temporaire : l'intelligence artificielle pourra remplacer, se substituer à l'humain quand l'apprentissage profond (*deep learning*) sera puissant et produira des intelligences artificielles fortes et conscientes d'elles-mêmes. **L'intelligence artificielle pourra se substituer à l'humain lorsqu'elle aura intégré la reconnaissance de formes (*pattern recognition*), mais n'aura pas de sens commun tant qu'elle ne sera**

---

**pas forte et consciente d'elle-même.** L'intelligence artificielle représente un choc technologique récent et explosif auquel la société n'est pas préparée.

**L'Europe ne pourra percer sur le marché des « GAFAMI » et des industries numériques chinoises** dans l'état actuel des choses ; sa seule chance est d'identifier le « coup d'après » afin d'être en mesure de concurrencer les États-Unis et la Chine au niveau de l'économie numérique. Cependant, **le monde politique est immature concernant ces questions.** Il est en retard et dans le déni technologique, bien qu'il soit dépendant du numérique. Dans cette perspective, l'avènement du populisme pourrait être lié à l'essor de l'intelligence artificielle, car **l'enjeu crucial sera de réorienter le tissu corporatif et industriel alors même que l'intelligence artificielle ne concernera pas toutes les branches d'activités en même temps.**

**4. M. Pierre-Yves Oudeyer, directeur de recherche à l'INRIA, directeur du laboratoire Flowers, président du comité technique des systèmes cognitifs et développementaux de l'IEEE (Institut des ingénieurs électriciens et électroniciens)**

Je suis directeur du laboratoire *Flowers*, et je travaille sur un domaine qui s'appelle la **robotique développementale et sociale**. Nous visons les machines et les algorithmes comme des outils pour construire des modèles d'apprentissage chez l'humain. On utilise les algorithmes pour comprendre l'évolution cognitive des enfants. J'ai auparavant travaillé dans l'entreprise Sony, au sein d'un laboratoire de recherche fondamentale.

Il y a eu des avancées ciblées dans un sous-domaine de l'intelligence artificielle, qui s'appelle l'apprentissage automatique (*deep learning*). Ces avancées techniques récentes sont aussi associées à un champ original dans la recherche sur l'intelligence artificielle : ces avancées sont développées de manière massive, un peu comme un programme Apollo, à l'intérieur d'un très petit nombre de grandes entreprises du *web* (Google, Twitter, Facebook, Microsoft, IBM et Apple). Ces grandes entreprises ont construit ces cinq dernières années des laboratoires de recherche en intelligence artificielle de taille bien supérieure à tous les laboratoires académiques du monde, et dans lesquels ils aspirent la plupart des cerveaux des meilleures universités du monde. Dans ces laboratoires de recherche, ils font des avancées fondamentales très vite, mais ils sont aussi capables de les appliquer extrêmement rapidement dans des applications utilisées au quotidien par leurs consommateurs.

Les techniques de développement de l'intelligence artificielle ne sont pas si nouvelles ; ces techniques font de la reconnaissance de forme. La reconnaissance de forme peut se faire par de l'apprentissage supervisé, qui existe depuis très longtemps : étant donné une base de données avec d'un côté des images et de l'autre des étiquettes, **les algorithmes vont trouver des**

**régularités dans des bases de données pour, dans le futur, être capable de deviner une étiquette.** Récemment, les chercheurs sont parvenus à faire marcher les algorithmes bien mieux qu'avant, tellement mieux que pour un certain nombre de tailles de reconnaissances, la machine est meilleure que l'humain.

**La machine est meilleure que les humains, en ce qu'elle est plus discriminante, plus précise et plus efficace ou plus rapide que l'humain dans des cas particuliers.** Par exemple, dans le monde médical, dans lequel les médecins doivent identifier des maladies à partir d'images, ou dans le domaine de l'astronomie, dans lesquels les astronomes identifient des structures de galaxies à partir d'images de télescopes, les algorithmes vont être capables, pour certains types d'images, de reconnaître des motifs plus rapidement et avec plus de précision que les humains, y compris les experts.

Dans un certain nombre d'applications, la combinaison de l'expertise humaine et de la machine est bien plus efficace que la machine seule ou que l'humain seul. D'un point de vue scientifique, il n'y a rien de nouveau, toutes les idées qui sont utilisées aujourd'hui étaient déjà présentes il y a une trentaine d'années. Il y a cependant deux choses nouvelles. D'une part, nous avons aujourd'hui à disposition de grandes bases de données avec les étiquettes, ce qui permet aux algorithmes de repérer des régularités qu'ils ne pouvaient pas repérer avec de plus petites bases. Ces bases de données sont le résultat de l'activité commerciale de ces grandes entreprises qui les construisent, et ont les moyens de les construire et de les entretenir. D'autre part, pour que les algorithmes repèrent des régularités, **il faut faire des calculs énormes pour pouvoir traiter toutes ces bases de données.** Jusqu'il y a peu de temps, les méthodes de calcul nécessaires n'étaient pas disponibles. Ce sont les organisations privées qui ont aujourd'hui cette puissance de calcul permettant de traiter l'ensemble des données de la base. L'utilisation de leurs bases de données leur permet de construire des algorithmes de reconnaissance de forme.

Pour ces entreprises, la valeur ajoutée ne se trouve pas nécessairement dans les algorithmes qu'elles ont mis au point, mais dans les bases de données sur lesquelles elles entraînent ces algorithmes. C'est ce qui permet à ces entreprises de publier un certain nombre de leurs algorithmes de manière ouverte (*Open source*), parce que les données ont une application commerciale, et donc une valeur commerciale haute. Ces données ont beaucoup de valeur, et elles se monnaient.

Cependant, si ces systèmes sont puissants, ils ont des limites. Il est possible de faire de la reconnaissance de forme performante à partir de bases de données et des applications pertinentes, mais on est encore loin de l'intelligence humaine. **L'intelligence artificielle est encore très « stupide » par beaucoup d'aspects.**

Cette « stupidité » de l'intelligence artificielle peut être constatée notamment sur la reconnaissance d'image. En prenant deux images représentant toutes deux un camion avec quelques pixels de différence, l'œil



---

humain ne percevra pas la différence infime entre ces deux images et reconnaîtra une même forme de camion. Cependant, pour l'algorithme, la première image reconnaîtra qu'il y a 99% de chances pour que ce soit bien un camion ; mais sur la deuxième image, il sera certain à 99% de chances que c'est autre chose.

Cela signifie que ces algorithmes de reconnaissance ont des taux de réussite de 99,5% sur la reconnaissance de forme, mais que les 0,5% d'erreur sont **des erreurs de sens commun et d'intuition énormes, qui sont incompréhensibles pour les humains**. Les humains, grâce à leur intuition et à leur connaissance générale du monde acquise au cours de leur développement, sont éduqués pour ne pas faire ces erreurs stupides.

Les algorithmes générateurs des images que doivent reconnaître la machine essaient de prendre en défaut les algorithmes de *deep learning*. Ces algorithmes de *deep learning*, lorsqu'ils sont fixés dans les applications dans le monde réel, peuvent être la cible de *hackers* qui vont développer des *stimuli* qui seront présentés à ces machines afin de mal orienter la décision à prendre dans la tâche qu'elles effectuent. Ce type d'erreur peut notamment être retrouvé dans le cas des voitures autonomes, où les algorithmes de reconnaissance de forme sont utilisés afin de prévenir les accidents. Un algorithme pilotant un véhicule autonome peut en effet être amené à devoir prendre des décisions rapides après avoir reconnu la présence d'un danger sur la route, comme, par exemple, des enfants accourant subitement sur la chaussée. Cependant, la présence d'un panneau publicitaire en bord de chaussée présentant des images sera interprétée comme un piéton traversant la rue et pourra diriger de fait le véhicule contre un mur afin de sauver un maximum de vies. Ce type de méthode statistique ne permet pas de garantir que des erreurs énormes ne seront pas faites. On ne peut donc garantir que des erreurs graves ne soient pas faites du fait des **confusions de représentation et d'interprétation des formes et images** par une intelligence artificielle.

Lorsque le législateur essayera ces technologies, il devra effectuer un changement de paradigme pour essayer de se faire une idée de leur utilité ou non. Auparavant, dans le domaine des transports, les technologies étaient présentes pour aider les gens, et notamment dans le domaine de l'aviation dans lesquels cela fait bien longtemps qu'il y a des systèmes automatisés ; cependant, l'introduction de systèmes automatisés était soumise à une certification : on pouvait prouver que ce système ne ferait pas d'erreur. Aujourd'hui, il existe de nombreux autres algorithmes d'aide à la décision humaine dans **des applications qui n'ont pas une culture de la sécurité** comme c'est le cas dans le domaine de l'aviation. C'est ainsi le cas dans le domaine de la voiture, où il n'y a pas du tout cette culture de la sécurité et où ces systèmes d'information automatisés sont, *de facto*, déjà utilisés. Une réflexion similaire peut se faire dans le domaine de la médecine : en médecine, lorsqu'il est constaté qu'un médicament fonctionne sur 99 % des patients, ce même médicament peut avoir de graves effets sur les 1 % de patients restant.

---

Aujourd'hui, on ne réfléchit plus nécessairement en termes de sécurité, mais en termes de bénéfice/coût. Ainsi, pour le cas du médicament, l'interprétation du risque en fonction de la balance bénéfice/coût se base sur le calcul statistique du nombre d'années d'espérance de vies gagnées et du gain pour la sécurité sociale par rapport à ce qu'il fait perdre. **Les algorithmes ne seront donc pas dans la perspective de prévenir tout risque à l'avance, mais d'effectuer un calcul bénéfice/coût afin de garantir le meilleur résultat possible.** Cette perspective implique que, par exemple, dans le cas des voitures autonomes, un pourcentage de pertes puisse être toléré lorsque l'algorithme prend en considération la situation à laquelle il est confronté.

Dans l'état actuel des connaissances, les techniques d'apprentissage de reconnaissance de formes sont mauvaises pour apprendre en une seule fois à un algorithme. Il lui faut des séries d'exemples pour qu'un algorithme intériorise une reconnaissance de forme ou une notion.

En médecine, comme en développement de technologies, **l'incertitude occupe une grande place dans le travail du chercheur.** C'est en cela que la répétition des essais est primordiale à la vérification et à la validation du travail du chercheur.

Il existe d'autres utilisations de ces algorithmes, qui sont encore plus quotidiennes que dans les véhicules autonomes. On peut prendre l'exemple d'AlphaGo : les ingénieurs de DeepMind ont effectué un travail impressionnant avec le jeu de Go. Néanmoins, la défaite de la machine au jeu de Go montre les limites de la machine de la manière suivante : dans les deux premières parties, Lee Sedol est battu nettement, et à partir de la troisième partie, il fut capable d'adapter sa stratégie et d'équilibrer le jeu. **L'algorithme n'était cependant pas capable de s'adapter sans avoir plusieurs millions de nouveaux exemples.** C'est une différence fondamentale entre l'intelligence humaine et l'intelligence d'un système d'information. En outre, le système utilisé pour jouer contre Lee Sedol au jeu de Go ne serait pas capable de jouer aux dames contre un enfant, bien que les règles du jeu de dames soient bien plus simples à comprendre que celles du jeu de Go. Il faudra le reprogrammer entièrement pour qu'il soit capable de jouer au jeu de dames. Un algorithme pourra être très efficace pour résoudre un problème difficile, mais qui est très spécifique. **L'algorithme ne peut s'adapter aussi vite que l'intelligence humaine,** car il lui faut des millions de données et exemples pour qu'il s'adapte et comprenne de nouvelles choses.

À ma connaissance, ce type d'algorithmes est plutôt utilisé en médecine dans le cadre de l'aide au diagnostic et dans l'aide à la décision du traitement à prescrire. Ils ne sont cependant pas encore utilisés pour les actes chirurgicaux en eux-mêmes. Certains le sont aussi en génomique, afin de repérer des structures ou des *patterns* dans l'expression des gènes.

Un autre domaine dans lequel ces algorithmes sont beaucoup utilisés est celui des filtres d'information : les moteurs de recherche, les

---

systèmes de personnalisation des articles d'actualité, de plus en plus de sites *web*, des grands journaux, qui ne proposent pas à tous la même hiérarchie des articles. Ils choisissent la hiérarchie des titres et des articles de manière à correspondre aux préférences de l'utilisateur. Pour cela, ces algorithmes vont utiliser les clics des utilisateurs effectués auparavant comme des mesures des préférences de l'utilisateur, afin de détecter des régularités dans les articles consultés par chaque utilisateur.

Cela permet aux utilisateurs de trouver des informations pertinentes, mais le mauvais point est que ces algorithmes vont avoir tendance à « mettre la poussière sous le tapis » : de même qu'un enfant à qui on a demandé de ranger sa chambre fera disparaître la poussière sous le tapis, l'algorithme de filtrage d'information, en souhaitant satisfaire les préférences de l'utilisateur, va lui proposer uniquement les articles correspondant à ses opinions. La problématique de ce phénomène est que cela va contribuer à **enfermer les utilisateurs dans une bulle d'information** qui leur sera propre du fait des suggestions basées sur leurs préférences, et ainsi polariser les opinions. Cela peut engendrer des problèmes sociétaux et politiques fondamentaux, en particulier pour le développement des extrémismes.

La plupart des gens n'ont pas conscience qu'il y a des algorithmes derrière ces systèmes s'adaptant à leurs préférences. Il serait important que les gens soient éduqués à la compréhension des grands principes du fonctionnement de ces systèmes, pour leur permettre ainsi d'éduquer ces systèmes lorsqu'ils interagissent avec eux afin de ne pas s'enfermer dans une bulle d'information. Cependant, d'autres personnes connaissent très bien ces algorithmes et leurs effets, et vont les utiliser volontairement, comme ce fut par exemple le cas avec le logiciel de *Microsoft* « *Tay* » qui, à force d'indications visant à mal orienter l'apprentissage de l'algorithme par de nombreux utilisateurs, a publié sur le réseau *Twitter* des propos racistes. Cet événement est symptomatique du fait qu'il est possible de modifier le comportement de ces algorithmes qui feront à notre place des choix de filtrage d'information, ce qui soulève des questions.

Les grandes entreprises développant ces algorithmes ont pris conscience de ce problème, certaines d'entre elles ayant commencé à développer des algorithmes « *adversario* » pour contrer ces algorithmes de filtrage qui provoquent des bulles d'information. On pourrait dire que cela est une bonne chose, car cela montre que les entreprises développant ces algorithmes ont conscience du problème ; cependant, est-ce que la solution est de développer des algorithmes qui, pour votre bien, vont lutter contre d'autres algorithmes qui voulaient aussi initialement votre bien mais qui font votre mal ? Le souci est que la décision est déléguée aux algorithmes qui vont lutter les uns contre les autres, sans vraiment poser la question.

**Les systèmes d'intelligence artificielle**, présents dans les voitures autonomes et dans les systèmes de filtrage d'information, fonctionnent de manière très différente des humains, en particulier parce qu'ils **sont**

**spécialisés sur une seule tâche programmée à la main par un ingénieur.** Cependant, les enfants humains n'apprennent pas du tout de cette manière ; **l'apprentissage des enfants est d'abord dirigé par des systèmes de motivation qui sont internes à leurs cerveaux.** Les enfants ont des systèmes de motivations extrinsèques, qui vont effectuer des activités car des éléments extérieurs les y poussent, comme une récompense offerte par les parents ; et ont aussi des systèmes de motivation intrinsèques, que l'on peut associer à la curiosité, et qui vont les pousser spontanément à explorer leur environnement. Il semble que l'évolution ait doté notre cerveau de circuits spécifiques de la curiosité qui non seulement nous poussent à explorer des situations nouvelles et surprenantes, mais qui nous procurent du plaisir dans la découverte et l'apprentissage.

Dans le cadre de nos recherches, nous avons essayé de modéliser ces mécanismes, le but n'étant pas de rendre les robots plus intelligents mais de voir **comment rendre plus précis ces mécanismes d'apprentissage** en les expérimentant sur des robots, et de fournir la théorie du fonctionnement de l'enfant et de l'humain. Les mécanismes de l'apprentissage humain sont tellement complexes que l'on ne peut les comprendre par de simples matrices cérébrales ; les mécanismes d'apprentissage humain ne peuvent être totalement appliqués à l'apprentissage automatique (*machine learning*). Le robot peut tester son environnement en faisant des expériences afin d'observer des régularités, engranger de l'information en effectuant des expériences et apprendre de ses erreurs. Les expériences menées sur l'interaction du système d'apprentissage, de son système d'exploration spontanée, des propriétés de son corps et de l'environnement permettent de fournir de nouvelles hypothèses concernant les enfants. Aujourd'hui, en collaboration avec des laboratoires de psychologie et de neurosciences, il est possible de faire des expériences nouvelles avec des humains et des animaux afin d'infirmer ou de confirmer certaines hypothèses théoriques.

Cependant, les systèmes d'intelligence artificielle développés dans ce type de laboratoire visent un but très différent que les systèmes d'intelligence artificielle développés par les grandes entreprises évoquées précédemment. Il ne s'agit pas ici de résoudre des problèmes d'ingénieurs, mais d'**explorer les processus d'apprentissage et du développement.**

Il ne s'agit pas de vendre un service qui serait commercialisé, mais cela ne veut pas dire qu'il n'y a pas des applications possibles. Par exemple, le laboratoire *Flowers* travaille actuellement sur des applications dans le domaine de l'éducation. Si l'équipe de chercheurs parvient à mieux comprendre certains processus de l'apprentissage, comme ceux liés à la curiosité, cela donnera des idées d'amélioration des systèmes pédagogiques et d'éducation. Cela pourrait également permettre de **personnaliser les parcours d'apprentissage**, et fournir des outils complémentaires au travail de l'enseignant. Un enseignant, face à trente élèves, pourra ainsi se voir fournir des outils lui permettant d'organiser la classe différemment et lui donnant les informations pour personnaliser son enseignement, ce qu'il ne peut faire mécaniquement pour trente élèves.

---

Ces exemples d'application sont d'ailleurs très différents des applications faites par les grandes entreprises développant des systèmes d'intelligence artificielle.

Il existe aussi des **logiciels de réadaptation ou de remobilisation** pour des enfants souffrant d'un handicap. Ce type d'application n'est pas exploité dans le laboratoire *Flowers* en particulier, mais il y a des collègues chercheurs qui se penchent sur ce type de recherche. Cependant, leurs applications visent les enfants, mais aussi les personnes âgées. Par exemple, le laboratoire *Flowers* est actuellement en train de travailler sur un projet avec l'ENST Bretagne à Brest autour des personnes âgées ayant besoin de rééducation physique et de kinésithérapie. Si la personne doit faire des mouvements physiques particuliers, le plus grand problème sera la motivation : à quel point fera-t-elle ses exercices sérieusement quand elle sera rentrée chez elle ? Avec ce projet, nous essayons de développer des technologies et des robots qui vont jouer un rôle de coach pour encourager les personnes âgées lorsque le kinésithérapeute n'est pas à côté d'elles.

Ces robots de soutien sont mis au point avec des enseignants pour notre projet porté sur l'école, et un enseignant a même été recruté afin de le mettre au cœur du développement du projet. Le laboratoire travaille très étroitement avec les écoles, les médecins et les associations de patients dans nos différents projets.

Nos collègues de Brest travaillent avec des diplômés formés à l'université dans des formations paramédicales, mais qui ne sont ni des médecins ni des kinésithérapeutes. Ces diplômés connaissent néanmoins parfaitement bien la morphologie et la biologie du corps humain, et sont capables d'interagir sur des fonctions particulières, soit pour compenser une déficience morphologique, soit pour améliorer le fonctionnement du corps pour des personnes en réadaptation fonctionnelle. Cependant, il existe également dans le domaine de l'éducation un certain nombre de spécialistes sur le terrain qui ont énormément de connaissances pratiques et qui sont là en complément des équipes éducatives.

Il est important de se demander ce que peut faire la France pour tirer parti de ces technologies. Il y a selon moi deux défis à relever. Tout d'abord **un défi sur le long terme, le défi de l'éducation**, qui est fondamental. L'éducation doit commencer dès le plus jeune âge en enseignant les principes fondamentaux de l'informatique. C'est ce qui est actuellement en train d'évoluer, avec l'intégration dans les programmes de l'école primaire de notions relatives au monde de l'informatique. Ensuite, sur le court terme, il y a **un défi d'équilibre entre le privé et le public**. La plupart des avancées sont faites par un petit nombre de grandes entreprises disposant de laboratoires qui attirent les meilleurs chercheurs et étudiants des universités les plus prestigieuses au monde. Le défi majeur sur le court terme est donc le rééquilibrage des forces de recherche entre les laboratoires publics et les laboratoires privés.

## IV. AUDITIONS DU 24 NOVEMBRE 2016

### 1. Mme Flora Fischer, chargée de programme de recherche au CIGREF, Club informatique des grandes entreprises françaises, pilote du groupe de travail sur l'intelligence artificielle en entreprise

Le CIGREF est une association de grandes entreprises françaises créée en 1970, qui regroupe cent quarante grandes entreprises et organismes français de tous secteurs d'activité. **Le CIGREF s'est donné comme mission de développer la capacité des grandes entreprises à intégrer et à maîtriser le numérique.** La gouvernance du CIGREF est assurée par quinze administrateurs et une petite équipe de permanents qui animent des groupes de travail, avec pour sujet l'innovation, la transformation des ressources humaines, la *blockchain*, etc. Le Cercle Intelligence Artificielle s'inscrit dans cette perspective. Initiative datant de 2015, la création du Cercle Intelligence Artificielle au CIGREF est le fruit de la concertation du président du CIGREF, M. Pascal Buffard, également président de *Axa Technology Services*, et M<sup>e</sup> Alain Bensoussan. **Cette initiative Cercle Intelligence Artificielle n'incluait que des entreprises « utilisatrices »,** ce qui exclut les fournisseurs et éditeurs de solutions logicielles du Cercle. Cependant, **il fut ouvert à des acteurs extérieurs - jeunes pousses, des experts, des consultants, etc. - afin d'apporter une expertise que les entreprises utilisatrices n'ont pas forcément toutes sur ce sujet.** En septembre 2016, dans le prolongement d'un colloque, le CIGREF a publié un Livre Blanc intitulé « Gouvernance de l'intelligence artificielle dans les grandes entreprises : enjeux managériaux, juridiques et éthiques » en partenariat avec le cabinet Alain Bensoussan Avocats.

**Le sujet de l'intelligence artificielle demeure très prospectif :** les entreprises viennent au Cercle Intelligence Artificielle principalement pour se nourrir d'informations. Le Cercle Intelligence Artificielle fournit des recommandations, qui se rattachent à trois métiers du CIGREF : l'intelligence, qui consiste à rassembler tous les savoirs des entreprises ; l'influence, qui permet au CIGREF de partager les convictions de ses entreprises membres à son écosystème ; et l'appartenance, qui repose sur les directeurs des systèmes informatiques qui doivent s'approprier ces sujets et les porter au sein de leurs entreprises.

Ce Cercle a, grâce à l'appui juridique de M<sup>e</sup> Bensoussan, une orientation très sociétale. Le CIGREF doit également prendre en compte un enjeu managérial. **Certaines entreprises étaient plus matures que les autres sur le sujet de l'intelligence artificielle, et avaient déjà des propositions fortes.** Par ailleurs, l'équipe de prospective d'une entreprise a soulevé au niveau du management le problème du risque de la fracture d'intelligence dans les entreprises.

---

Dans la perspective du CIGREF, **la préoccupation majeure est celle de la transformation du management**, en premier lieu, **et ensuite celle des usages**. En ce sens, Bernard Georges<sup>1</sup> a affirmé que, **pour éviter la fracture intelligente qui suivrait la fracture numérique**, et pour que les entreprises ne se retrouvent pas en incapacité d'intégrer ces outils, **il faudrait que l'entreprise s'ouvre de manière plus fluide à tout ce qui est en mouvement autour d'elle et qui génère de la disruption**. Dans le schéma pyramidal et rigide des entreprises, le bas de la pyramide, en contact avec le terrain, peut s'ouvrir sur l'écosystème, tout comme le haut de la pyramide qui est en contact constant avec son environnement ; cependant, le milieu de la pyramide, qui gère beaucoup d'équipes centrales, a en quelque sorte pour habitude que les règles ne changent pas et ne peut donc pas s'ouvrir. Il y a un véritable enjeu de « fractalisation »<sup>2</sup> des organisations.

**L'utilisation de l'intelligence artificielle ne peut être développée et appropriée au sein des entreprises de manière isolée ; il faut fonctionner entre communautés**. Les entreprises doivent entrer soit en collaboration, soit en « co-pétition », car toutes les entreprises ont besoin et tirent un avantage des données des autres. L'intelligence artificielle repose sur l'*Open source* et l'*Open data* ; **il y a donc un enjeu d'ouverture qui est très compliqué à aborder dans les grandes entreprises**, pour des questions culturelles, de rigidité, et de pouvoir. Dans cette perspective, sur initiative de leur directeur des données (CDO – *Chief Data Director*), la MAIF tente de convoquer un maximum d'expertise et de disciplines différentes afin d'aborder le sujet de l'intelligence artificielle, dans l'optique de mettre en place un comité consultatif d'éthique sur la politique liée à l'intelligence artificielle.

Une entreprise a également intégré la dimension sociétale de l'intelligence artificielle dans les scénarios prospectifs élaborés sur ce sujet, notamment sur le type d'emplois qui seront impactés par l'intelligence artificielle, sur son acceptation sociétale, etc. Dans un premier scénario, les entreprises resteraient très méfiantes du fait de l'impact conséquent de l'intelligence artificielle sur l'emploi. Un autre scénario prévoit qu'une relation d'empathie se serait développée à l'égard des machines, en considérant que ce ne sont plus des contraintes comme cela pouvait être le cas avec l'informatique qui a contraint les relations lors de son introduction dans les entreprises. **Contrairement à l'informatique, l'intelligence artificielle nous amène de plus en plus vers des interfaces naturelles, qui visent à fluidifier son usage en limitant notamment le nombre**

---

<sup>1</sup> *Responsable maîtrise d'ouvrages stratégiques (innovation, systémique, prospective) à la Société Générale*

<sup>2</sup> *Au lieu d'une structure pyramidale, l'entreprise fractale est formée à partir d'une même structure simple et maîtrisée qui se répète à différentes échelles. Une représentation simplifiée du concept pourrait en être les poupées russes.*

---

**d'opérations à faire (par exemple, grâce à l'échange vocal avec un bot) ;  
l'intelligence artificielle se fonde sur l'interaction.**

Le CIGREF met en avant également l'éthique by design, en s'inspirant du concept de *privacy by design*. L'éthique *by design* consiste à **prendre en considération le plus en amont possible de la conception, et dans le suivi d'une innovation technologique, les questions éthiques et les questions d'acceptabilité sociales**. Cela suppose d'intégrer des valeurs éthiques dans la conception même des outils numériques (applications, code informatique...).

Il existe un sous-champ de recherche de l'éthique de la technologie, la *moral ethics*, qui s'intéresse à la théorie des agents artificiels. Cette théorie consiste à chercher la manière dont peuvent être implémentés des principes éthiques chez des agents intelligents qui agiront de manière autonome. De nombreuses expériences sont menées afin de voir si une machine pouvait se comporter de manière éthique d'un point de vue humain.

L'éthique *by design* comprend deux aspects. Le premier aspect est la **mission éthique a priori**, à effectuer en amont, car les technologies ne sont pas neutres, et **il faut donc expliciter les choix faits lors de la programmation d'un algorithme**. Le second aspect est la **mission éthique a posteriori, dans le suivi, avec l'apprentissage d'une intelligence artificielle**. Si la recherche parvient à inventer des systèmes d'apprentissage automatique (*machine learning*), ce vers quoi les chercheurs se dirigent petit à petit, il sera important de contrôler régulièrement comment une intelligence apprend et comment elle modélise à partir des données qu'elle a elle-même récoltées et analysées. Cependant, il n'est pas possible d'accéder à tous les niveaux de détails du processus d'apprentissage profond et automatique d'une intelligence artificielle.

L'éthique *by design* considère l'intelligence artificielle comme une substance active, à l'image d'un médicament, et implique donc de procéder de la même manière que pour la mise en vente d'un médicament, c'est-à-dire en appliquant de nombreuses séries de tests sur l'usage et l'interaction avec l'environnement de l'intelligence artificielle avant toute mise sur le marché.

Les entreprises demeurent très prudentes sur l'*Open data*, davantage encore que pour l'*Open source*. Les entreprises publiques ont plus de facilité à engager des initiatives d'*Open data*, car la menée de ce type d'initiatives au sein d'entreprises privées sera confrontée à des questions de protection des affaires et de confidentialité de l'information patrimoniale. **Tant que les entreprises ne seront pas matures en ce qui concerne le traitement et la valorisation des données, il sera compliqué pour elles de l'être pour l'intelligence artificielle.**

Pour conclure, dans son rapport « *Gouvernance de l'intelligence artificielle dans les entreprises* », le CIGREF effectue plusieurs recommandations visant le développement de l'utilisation de l'intelligence artificielle dans les entreprises. Le CIGREF recommande tout d'abord d'allouer des budgets et des ressources pour l'intelligence artificielle dans les



---

entreprises. Il est également recommandé de développer le travail de prospective, en développant des scénarios qui permettent d'orienter les stratégies.

Viennent ensuite deux enjeux essentiels. Tout d'abord, l'enjeu opérationnel d'anticipation de la transformation des infrastructures de technologies de l'information (IT), en passant d'architectures séquentielles à des architectures parallèles. Ensuite se pose un enjeu commercial : grâce au rôle du directeur des données, l'entreprise doit réussir à faire travailler les infrastructures technologiques de base avec les métiers afin de pouvoir répondre à des enjeux commerciaux. Enfin, les entreprises doivent attirer et valoriser les talents.

**2. M. Max Dauchet, professeur émérite à l'Université de Lille, président de la Commission de Réflexion sur l'Éthique de la Recherche en sciences et technologies du numérique (CERNA) d'Allistene, alliance des sciences et technologies du numérique**

La CERNA ambitionnerait que les débats autour du numérique aient le même statut social que les débats sur la santé, car le numérique concerne l'avenir de l'humain de la même façon. La CERNA a été créée il y a quatre ans maintenant par Allistene (alliance des sciences et technologies du numérique) : comme le CNRS et l'INRIA pensaient qu'il était nécessaire de mener une réflexion éthique spécifique au numérique, ils ont préféré former une commission unique au sein d'Allistene.

La CERNA est saisie ou s'autosaisit de sujets d'éthiques de la recherche en sciences et technologies du numérique. Comme la recherche est grandement liée aux usages, sachant à quelle vitesse les usages se propagent dans la société qui s'en empare, la CERNA a décidé de s'orienter vers la contribution du chercheur à tous les débats.

Actuellement, l'intelligence artificielle fait peur. C'est une peur qui est peut-être en partie orchestrée ; mais si vous voyez **les déclarations faites dans la presse**, notamment encore récemment celle faite par M. Stephen Hawking, elles sont **soit alarmistes ou d'une opacité rassurante**, dans le style « *ne vous inquiétez pas, on s'en occupe* ». Je souhaiterais vous faire partager quelques réflexions personnelles à partir de l'actualité.

Le 24 octobre 2016, MM. Martin Abadi et David G. Andersen, deux chercheurs de *Google Brain* [le programme de recherche en intelligence artificielle de Google], ont publié un article dans lequel ils expliquent comment deux réseaux de neurones apprennent à communiquer entre eux en secret d'un troisième réseau de neurones. L'expérience menée par ces deux chercheurs est une piste scientifique intéressante, et ne soulève rien d'inquiétant en termes d'intelligence artificielle. Cependant, **le fait que l'on ne parle de l'intelligence artificielle seulement lors d'accélération**

---

**médiatisées de la recherche peut contribuer à susciter l'inquiétude.** La première chose à faire est de former les enseignants au numérique.

Il y a, en outre, des plateformes et des outils sur Internet qui permettent à une intelligence artificielle d'apprendre à optimiser, par exemple, le débit d'arrosage de votre jardin. Il existe des détenteurs de capitaux aux États-Unis qui se présentent ouvertement comme activistes et philanthropes tout en encourageant le développement des outils d'intelligence artificielle permettant aux gens de les prendre en main. Ces plateformes restent cependant en langue anglaise.

Il est possible d'observer, dans le cadre du développement de l'intelligence artificielle, une **double fracture : une fracture numérique, et une fracture de la langue.** Il apparaît donc essentiel de développer une plateforme ayant recours à l'intelligence artificielle en français afin de lutter contre cette double fracture.

Il existe certes un problème culturel, mais aussi un **problème statutaire dans le système français de la recherche publique**. Le système français de la recherche publique est en silo : l'interdisciplinarité est prêchée, cependant, dans une carrière, pour obtenir des promotions, il faut être très pointu dans un domaine particulier. En outre, la transmission de la connaissance et le travail d'élargissement de la connaissance en interdisciplinarité ne sont pas valorisés. Cela figure dans les fiches d'évaluations des chercheurs ; cependant, les éléments d'interdisciplinarité ne sont jamais pris en compte dans la considération d'un avancement de carrière. **L'interdisciplinarité doit être promue avant tout dans les universités.**

Le débat entourant le développement des nanotechnologies peut être pris en exemple : ce débat fut confié à la commission nationale du débat public, et fut très mal instruit, car ce débat s'est résumé en une succession d'affirmations et d'oppositions non-scientifiques. Ce genre de débat doit avoir lieu dans les universités. En ce sens, si vous souhaitez engager un débat public sur l'intelligence artificielle, il faudrait le faire dans les universités, même si ces dernières n'auront peut-être pas l'enthousiasme qu'il faudrait.

Cependant, comme pour le débat sur les nanotechnologies, la gestion d'un important risque industriel est à anticiper. Il est crucial d'apprendre à **gérer le risque numérique tout comme l'est le risque industriel**, sans besoin de fantasmer. Certes, le débat est apaisé sur la question des nanotechnologies, mais cela a eu pour résultat que la France qui, en 2007, représentait 6 % de la publication scientifique sur les nanotechnologies, ne représente, en 2016, plus que 4 % de publications scientifiques sur les nanotechnologies. La Chine, pour sa part, représente 47 %, alors même qu'elle n'en représentait que 11 % il y a huit ans. De fait, il ne faut pas être paralysé par des peurs, sur le sujet des nanotechnologies comme sur le sujet de l'intelligence artificielle, car cela tournera à notre désavantage.

---

Les chiffres des parts de publication scientifique dans le monde sont absolus, mais il est possible de ramener ces chiffres à l'importance de la population. Dans cette perspective, le rapport entre la France et les États-Unis demeure correct ; et le rapport entre la France et la Chine demeure favorable à la France. La France représente le 1/8 de la production scientifique de la Chine en nanotechnologies alors que sa population est vingt fois moins nombreuse. La France n'a pas à rougir, car il y a un poids démographique à prendre en compte. La production scientifique française est excellente, mais la France ne représente que 1 % de la population planétaire, ce qui peut poser un problème en termes de relais industriels.

Il est possible d'envisager une initiative européenne concernant l'intelligence artificielle. Cependant, les premiers à pâtir de la montée en puissance en Europe de l'intelligence artificielle seraient les pays qui sont d'importants exportateurs en matière de mécanique de très haute précision, cette exportation se dirigeant majoritairement vers les États-Unis d'Amérique. Il y aura une rencontre entre l'intelligence artificielle et la mécanique de très haute précision (pour les robots et micro-robots) ; or, les États-Unis d'Amérique importent toutes leurs machines depuis les plus gros exportateurs européens de machines mécaniques de très haute précision, soit la Suisse, l'Allemagne, le Danemark, l'Italie, la Suède et l'Autriche. Il est également important de **prendre des initiatives d'information et de formation sur l'intelligence artificielle, mais plus largement sur les sujets du numérique, en langue française**. Pour les personnes à fort capital intellectuel dans le numérique, la confrontation à l'anglais dans leur découverte du savoir n'est pas problématique ; cependant, développer des initiatives d'information sur ces sujets en langue française permettrait d'impliquer plus largement les gens dans la compréhension du développement de ces technologies et ainsi de réduire leurs inquiétudes.

L'apprentissage d'un réseau de neurones sur de grandes bases de données d'images implique des millions de paramètres qui limitent la compréhension de ce processus. Cette complexité peut avoir pour effet de faire peur aux gens ; c'est pour cela qu'il est essentiel de les former.

La question de la gestion du risque a imprégné le monde industriel. Il y a, dans le numérique, des risques d'un autre type, parfois plus abstraits mais qui n'en nécessitent pas moins le développement d'une culture du risque numérique semblable à la culture du risque industriel : pas une culture qui vise à effrayer, mais qui vise à **mettre en place des garde-fous et des outils de précaution raisonnables**, pour que ces outils s'insèrent bien dans la société.

Il existe cependant un risque de se heurter à des difficultés institutionnelles, notamment pour l'introduction de ce type d'enseignements dans les programmes, encore, par rapport aux statuts des enseignants en fonction de leurs disciplines.

Concernant les missions et objectifs de la CERNA, en couvrant tout le secteur public de la recherche dans le numérique, le principal levier est la

sensibilisation et l'équipement des chercheurs pour aborder ces questions. Le but est d'intégrer davantage les chercheurs aux délibérations du pays. La CERNA pourrait disparaître au profit d'une autre structure.

La CERNA vise à se positionner comme une compétence-ressource au titre de la recherche dans le débat public, en particulier vis-à-vis de la CNIL, qui est l'ensemblier de ces questions. Actuellement, la CERNA n'est pas visible en dehors de la sphère académique. Il est souhaitable qu'une plus grande visibilité de la CERNA permette aux chercheurs de contribuer pleinement aux débats nationaux. Un certain mur de verre persiste entre les chercheurs et le reste de la société au sein du débat public. Les chercheurs sont cependant de plus en plus préparés à participer au débat public : des actions sont mises en place à destination des doctorants dans ce sens.

### **3. M. Cédric Sauviat, ingénieur, président de l'association française contre l'intelligence artificielle (AFCIA) et Mme Marie David, ingénieur, éditrice, membre du bureau de l'association**

L'AFCIA estime que l'intelligence est la manifestation d'une puissance de calcul d'un processeur, que ce soit le cerveau, un microprocesseur ou un ordinateur. À partir du moment où est affirmé que la puissance de calcul permet d'aboutir à l'intelligence, rien n'empêche l'émergence d'une conscience artificielle : si l'on considère que la conscience est un stade d'évolution de l'intelligence, elle peut très bien émerger sur des principes artificiels.

**L'accélération du progrès technique actuel rend la perspective de la création d'une intelligence artificielle douée d'une puissance de calcul comparable à celle du cerveau humain probable à l'horizon 2025-2030.** Le niveau humain n'est lui-même qu'une étape dans le développement de l'intelligence artificielle : si, actuellement, l'intelligence artificielle a le niveau d'intelligence des souris, en 2025 elle aura atteint le niveau humain, et en 2030, loi d'évolution oblige, elle sera mille fois plus puissante qu'un cerveau humain. C'est ce que certains chercheurs appellent la **super-intelligence**.

L'AFCIA part également du principe que **toute la recherche en intelligence artificielle actuelle est orientée vers l'émergence de cette super-intelligence, soit délibérément, soit inconsciemment.**

L'AFCIA constate enfin que **l'émergence d'une intelligence artificielle super-intelligente, suprahumaine, ouvre nécessairement un nouveau paradigme de civilisation.** La civilisation telle que nous la connaissons actuellement est fondée sur la limite de l'intelligence humaine ; dès lors que cette limite sera franchie, un nouveau paradigme de civilisation émergera.

**L'émergence de l'intelligence artificielle pose trois problèmes.** Le premier problème est le **risque technologique**, qui consiste à répondre à la

---

question : « l'intelligence artificielle peut-elle rester sous contrôle humain ? ». Le deuxième problème se situe au niveau des **enjeux sociaux**, des répercussions de l'intelligence artificielle sur le système socio-économique qui est le fondement de notre société. Enfin, le troisième problème relève de la **dimension anthropologique et éthique de l'intelligence artificielle**, visant à se demander quel est, dans cette évolution, le projet pour l'humanité.

Le risque technologique est largement ignoré en France, comme en attestent les déclarations régulières dans les médias de M. Jean-Gabriel Ganascia, président du Comité éthique du CNRS, qui affirme que l'émergence de l'intelligence artificielle ne pose aucun problème technologique particulier. Ce problème est néanmoins pris au sérieux par les Anglo-Saxons, en particulier par Nick Bostrom<sup>1</sup> qui est le philosophe le plus en pointe sur la question du risque technologique de la super-intelligence, ainsi que Bill Gates et Stuart Russell, l'un des auteurs du *best-seller* mondial de manuel d'intelligence artificielle, ce dernier considérant que « *l'intelligence artificielle est plus dangereuse que le nucléaire* ». **L'enjeu pour les Anglo-Saxons est de résoudre le problème du contrôle avant de mettre au point l'intelligence artificielle et que celle-ci n'atteigne un niveau de conscience humain.** Le parallèle avec le nucléaire est très éclairant, montrant que la découverte des effets du nucléaire, dont la première application remonte à 1945, s'est faite dans les années 1960, notamment sur les tissus humains. En 1970, le phénomène d'impulsion électromagnétique fut découvert, qui fait que lorsqu'une bombe atomique est déclenchée, une impulsion électromagnétique vient paralyser les systèmes électroniques environnants. Enfin, en 1980, le phénomène d'hiver nucléaire fut découvert, impliquant le fait que, en cas de conflit nucléaire mondial, l'humanité subirait une glaciation du fait de la formation de poussières dans l'atmosphère qui absorberaient les rayonnements du Soleil. Il a donc fallu cinquante ans pour maîtriser toutes les conséquences et composantes de la technologie nucléaire. Ainsi, selon Bostrom et Russell notamment, si le même processus est appliqué pour l'intelligence artificielle, l'humanité est perdue ; il faut donc réfléchir bien en amont aux conséquences de cette invention.

L'AFCIA considère, pour sa part, que ce problème est insoluble par définition : **vouloir donner de plus en plus d'autonomie à des systèmes intelligents et, dans le même temps, chercher à les garder sous contrôle est un paradoxe.** En outre, pour l'AFCIA, résoudre le problème technique - à supposer que cela soit possible - n'empêchera jamais un détournement malveillant ni une aliénation pratique, l'aliénation pratique étant la subordination consentie de l'humain à l'artificiel.

Pour les transhumanistes, pour éviter que l'Homme ne soit dominé par la machine, il faut fusionner l'Homme et la machine. La réaction de

---

<sup>1</sup> Directeur de l'Institut pour le futur de l'humanité à l'Université d'Oxford et Directeur du Centre stratégique de la recherche sur l'intelligence artificielle (initiative de l'Université d'Oxford et de l'Université de Cambridge)

l'AFCIA est d'affirmer que le **transhumanisme consistant en une fusion homme-machine** est absurde, puisqu'il représente un suicide, **une dissolution volontaire de la personnalité dans l'inconnu** : à partir du moment où vous modifiez l'intelligence de quelque chose ou de quelqu'un, vous modifiez sa personnalité. Cette solution des transhumanistes est surtout naïve : si vous attachez ensemble un cerveau humain et une super-intelligence, il est possible de deviner que la super-intelligence, si elle réagit comme un organisme hybride dans un premier temps, prendra rapidement son autonomie, ce qui ramène à une situation de domination de la machine sur l'humain.

Concernant le niveau des enjeux sociaux, **l'humanité entre dans une ère de compétition entre les cerveaux humains et les cerveaux artificiels**. Pour une tâche déterminée, le robot coûte à peu près le même prix pour une entreprise que le travailleur. L'AFCIA considère que la théorie libérale, qui affirme que de nouveaux emplois vont émerger et remplacer ceux qui disparaissent, est caduque car cette théorie n'a jamais fait l'hypothèse que le cerveau humain était lui-même en compétition. Cette théorie libérale se fonde sur l'observation de deux mille ans d'Histoire, mais ne fait en aucun cas l'hypothèse d'une situation dans laquelle une intelligence artificielle est créée. De fait, agiter comme un dogme que de nouveaux emplois seront constitués paraît pour le moins téméraire. L'AFCIA estime que les nouveaux emplois ne se constitueront pas assez vite pour remplacer ceux qui disparaissent. Certains économistes, tels que Michael Osborne, Daniel Cohen ou Nicolas Colin, constatent que, dans vingt ans, 30 % à 40 % des emplois seront robotisés. Daniel Cohen prédit, pour sa part, que la classe moyenne sera déclassée, puisque tous les emplois de classe moyenne de supervision dans les entreprises seront confiés à des algorithmes. Pour Nicolas Colin, **les emplois nouvellement créés ne remplaceront pas en nombre suffisant les anciens**, et c'est la raison pour laquelle la France doit accélérer son développement dans l'économie numérique pour éviter que ses emplois se fassent détruire par d'autres pays ; mais il n'y aura pas d'emplois suffisants pour tout le monde.

**Ce ne sont pas seulement les emplois de la classe moyenne et des ouvriers qui sont menacés, mais également des emplois d'expertise**, comme les emplois de médecins ou d'avocats. L'intelligence artificielle touche des domaines de plus en plus vastes ; de fait, toutes les couches du marché du travail seront concernées par l'intelligence artificielle.

L'économie numérique, fondée sur le monopole, prépare le règne de l'intelligence artificielle. Avec la multiplication des objets connectés, l'intelligence artificielle prépare une humanité sous surveillance (*monitoring*) permanente.

**Si les emplois ne se reconstituent plus assez vite, il va falloir organiser une société sans travail**. Cette société consisterait en l'émergence d'une ploutocratie rassemblant les détenteurs des entreprises numériques et ceux qui ont suffisamment de capital pour échapper au déclassement par le

---

travail. Cette société se caractérise également par **l'aliénation des classes moyennes par le revenu minimum universel**, ainsi que par des occupations de substitution pour les masses. Enfin, **puisque l'humain n'aura plus vocation à travailler, son éducation sera négligée car il y aura beaucoup moins d'intérêt à investir dans son avenir.**

Il est possible de voir dans les prémices de cette société sans travail un lien entre le processus économique en train de se créer au niveau mondial, qui se fonde sur la disruption<sup>1</sup> et l'exclusion de toute une partie de la population, et la montée d'une violence sociale qui se manifeste, par exemple, dans les pays arabes qui sont le parangon de ce qui arrive aux classes modestes dans les pays occidentaux. **L'intelligence artificielle a un impact extrêmement perturbateur au niveau social.**

Enfin, l'aspect anthropologique et éthique de l'intelligence artificielle, très feutré et peu souvent évoqué, est fondamental et a motivé la fondation de l'AFCIA. L'AFCIA constate que **le développement de l'intelligence artificielle suit la logique ultime et absurde d'un système technicien**, notamment décrit dans les travaux de Jacques Ellul, fondé sur la recherche perpétuelle de la performance et du « faire mieux », aboutissant au fait que **l'Homme sera exclu dans cette recherche du mieux**. Cela est absurde, dans la mesure où la technique devrait être au service de l'Homme ; mais dans cette perspective de système technicien, la technique supprimera ultimement l'humain.

L'AFCIA considère que les mobiles philanthropiques de l'intelligence artificielle et les bienfaits qu'elle apporterait sont sujets à caution. Par exemple, lorsqu'un représentant d'une grande entreprise du numérique affirme que l'intelligence artificielle apportera beaucoup de bienfaits à la planète, il refusera cependant de prendre le risque de rejoindre la masse qui subit ce phénomène ; cela ne vaut que s'ils peuvent rester eux-mêmes en haut de la pyramide.

L'AFCIA pense que **la volonté de développement d'une intelligence artificielle de niveau humain est la négation d'un besoin psychologique fondamental de l'Homme qui est la reconnaissance par ses pairs et la réalisation de soi**. Ces besoins fondamentaux se retrouvent dans la célèbre pyramide de Maslow, qui a à sa base la satisfaction des besoins élémentaires de l'Homme, puis le désir de protection, et, une fois ces besoins assouvis, le besoin de reconnaissance par ses pairs, le besoin d'amour, le besoin d'estime, et enfin l'épanouissement individuel. **À partir du moment où l'Homme est remplacé par une machine et devient inutile, lui laissant plus de temps libre pour s'adonner à ses loisirs, on croit répondre à ses besoins fondamentaux alors qu'ils sont en réalité démolis.**

---

<sup>1</sup> Le dynamisme de l'économie numérique a des effets de destruction créatrice : une innovation et sa diffusion permet de réaliser des gains de productivité, de créer de nouvelles activités et donc des nouveaux emplois, synonymes de croissance. De fait, seuls survivront sur le marché les entrepreneurs innovants capables de restructurer leur activité, d'adapter leurs procédés, d'améliorer leur processus, de manière à bénéficier des avantages compétitifs et de devenir des leaders sur un marché.

L'intelligence artificielle a également pour corollaire que l'on souhaite apporter à l'Homme une solution à tous ses problèmes. C'est l'avènement de l'autonomie absolue de l'Homme, qui n'a pas besoin des autres puisque tout lui est apporté par son assistant personnel, son robot et tous les systèmes robotisés à sa disposition. De fait, on fait perdre le sens de l'autre et le sens de la réciprocité à l'individu ; la société humaine est atomisée.

Au-delà de ce stade d'intervention de l'intelligence artificielle, il faut également prendre en compte le transhumanisme, qui consiste à modifier et augmenter l'humain afin de lui rendre service. Le transhumanisme risque, par l'augmentation de certains individus mais pas de tous, et du fait de la diversité des conditions qui seront créées par l'hybridation de l'Homme, de mettre fin à l'universalité de la condition humaine et à l'unicité de la race humaine. Cette perspective ouvre la voie à un fondement objectif du racisme, et la disparition de tous les repères éthiques qui peuvent exister dans la civilisation et l'humanité. **L'intelligence artificielle peut potentiellement signer la fin de l'éthique.**

Certains roboticiens et chercheurs affirment que l'intelligence artificielle, au même titre que l'homme, est une étape de l'évolution naturelle, qui continue son œuvre vers toujours plus d'intelligence. **L'AFCIA considère cette pensée comme une idéologie**, un système philosophique qui n'a pas de démonstration ni de scientificité, **qui est parente des idéologies totalitaires du XXe siècle** : là où les idéologies totalitaires expliquaient la suprématie de la race supérieure ou de la classe ouvrière, cette idéologie de l'intelligence artificielle explique que le développement de l'intelligence artificielle est une évolution de la nature, et que, par conséquent, il faut l'accélérer. On retrouve dans ces idéologies une explication du passé et, en conclusion, l'accélération de l'évolution. En reprenant l'analyse d'Hannah Arendt, dans le nazisme, il faut aider la nature à se rapprocher le plus possible de l'Homme parfait et de la race supérieure, et, par conséquent, éliminer les races inférieures ; et, dans le stalinisme, la domination de la classe ouvrière est le sens de l'Histoire, il faut de fait tout faire pour cheminer avec lui et éliminer tout ce qui s'oppose à son avènement. Il est possible de transposer cette analyse au cas de l'intelligence artificielle : l'évolution se faisant dans le sens d'une plus grande intelligence, il faut accélérer le sens de l'Histoire. Hannah Arendt a déclaré : « *l'objectif des régimes totalitaires n'est pas d'assurer la domination des Hommes, mais d'assurer la superfluité des Hommes. Le régime totalitaire est parvenu à ses fins lorsque l'Homme est devenu une entité négligeable superflue* ». **C'est exactement le programme, dans un tout autre domaine d'idées, de l'intelligence artificielle : rendre l'Homme superflu.** Il y a une proximité des explications idéologiques à l'œuvre dans ces trois phénomènes.

Or, cette idéologie est fautive car, dans l'évolution, la conscience humaine a donné naissance à la préoccupation morale ; à partir du moment où la préoccupation morale est arrivée dans le domaine de la civilisation, toutes les idéologies d'évolution sont caduques. C'est pour cela que



---

l'idéologie de certains chercheurs en intelligence artificielle est, aux yeux de l'AFCIA, nulle et non avenue.

Un courant de pensée post-humaniste peut être évoqué. Celui-ci est représenté notamment par Jean-Michel Besnier qui a tenté de décrire ce que pourrait être une éthique post-humaine : en constatant que les robots seraient considérés comme des consciences à respecter à part entière, il est possible d'imaginer un monde dans lequel hommes et robots vivent en harmonie. Selon l'AFCIA, cette vision est anthropocentrique, asimovienne, dépassée, et surtout pré-kantienne en cela que M. Besnier considère que, avec son cerveau humain, il est capable d'imaginer le fonctionnement d'un cerveau d'une intelligence artificielle et ce que pourrait être l'éthique d'une telle intelligence artificielle. **Le cerveau humain trouverait sa limitation dans sa non-perception de quelque chose dans lequel il n'y a pas d'espace et de temps ; de la même manière, un cerveau humain ne peut pas imaginer ce qu'il se passe dans un cerveau qui est mille fois plus intelligent que le nôtre.** Si l'intelligence artificielle ne naît pas d'une génération spontanée, même si le cerveau humain est capable d'inventer l'intelligence artificielle, cette dernière finira par s'autoalimenter, apprendre par elle-même et deviendra de plus en plus puissante par une réaction en chaîne, échappant au contrôle et à la compréhension des humains. D'autant que l'intelligence artificielle ne s'incarne pas par des robots comme ceux décrits par Isaac Asimov ; ce sont des systèmes dynamiques qui sont plongés dans un *continuum* d'informations. Il y a une dématérialisation consubstantielle à l'intelligence artificielle qui fait que ce sont des organismes qui n'ont ni commencement ni fin. **Contrairement aux individus, auxquels on peut attribuer une limite du fait de l'individualité identifiée différenciant chacun, il n'y a pas de discontinuité entre une intelligence artificielle et une autre, puisqu'elles sont en mesure d'échanger leurs informations d'une manière qui fait que l'on peut considérer qu'il s'agit d'un seul et même système.** Sur ces prémices, avec un processeur mille fois plus rapide que le cerveau humain et avec des capacités différentes de celles des êtres humains, **il est difficile de pouvoir prévoir l'éthique de ce type d'intelligence artificielle. La seule chose qui puisse être dite est que l'on ne peut pas savoir ce que peut être l'éthique d'une super-intelligence,** et peut-être qu'il est préférable de ne pas le savoir.

Le projet de l'AFCIA se fonde sur un progrès moral et non sur un progrès technique. L'AFCIA se réfère ici à l'humanisme des Lumières. L'AFCIA considère que le fait de se réaliser par son travail au service des autres est une dignité et un droit fondamental de la personne humaine, ce qui est d'ailleurs écrit dans la Déclaration Universelle des Droits de l'Homme de 1948. Il est impératif de préserver l'unité de l'espèce humaine et l'universalité de la condition humaine, car **sans universalité de la condition**

**humaine, il n’y a pas d’intercompréhension naturelle mutuelle et pas de possibilité de vivre ensemble ; c’est la possibilité de se comprendre qui est à la base de la civilisation.** S’il y a d’un côté des hommes, de l’autre des hommes augmentés, et enfin des robots, il n’y a pas d’intercompréhension mutuelle.

Selon l’AFCIA, le progrès technique, qui est en passe de révolutionner la civilisation, doit voir son rythme exponentiel subordonné à la stabilité et à l’harmonie sociale. La stabilité et l’harmonie des sociétés sont des enjeux bien plus importants que le progrès technique, qui ne devrait être au service que de cette stabilité. La société devrait être organisée de manière à permettre à chacun de trouver une place authentique par le service qu’il rend aux autres. Lorsque l’entreprise Uber met en place des systèmes capables de détruire toute une profession, d’un point de vue technicien, l’entreprise a raison ; mais elle ne rend pas service à la société. Il vaut mieux que l’objectif de la société soit que chacun ait une place, un rôle à jouer et quelque chose à offrir aux autres, plutôt que de chercher l’optimum et l’efficacité suprême.

Pour l’AFCIA, la course à l’intelligence artificielle, au même titre que la course à l’armement nucléaire, doit être stoppée. **La course à l’intelligence artificielle est fondée, au niveau des États, sur une volonté de puissance,** chacun devant se doter de l’intelligence la plus capable de rivaliser et de surpasser les rivaux. La course à l’intelligence artificielle est comparable à la course à l’armement nucléaire qui a eu lieu durant la guerre froide. Il est essentiel de désarmer cette course, afin d’éviter les conflits pouvant éclater entre des nations rivales.

Enfin, l’AFCIA considère que l’intelligence artificielle est, malgré tout, formidable, car il suffit de l’interdire pour redonner un sens au progrès technique, pour qu’il soit au service des humains. **Il suffit d’interdire l’intelligence artificielle pour que le progrès technique redevienne maîtrisable par l’Homme.**

L’AFCIA souhaite lancer **une campagne nationale pour offrir une base de réflexion aux citoyens,** afin de briser le discours dominant sur la disruption heureuse et les bienfaits de l’intelligence artificielle. L’AFCIA souhaite également **constituer un pôle d’opposition raisonné et non-violent à l’intelligence artificielle,** car, tôt ou tard, l’opposition à l’intelligence artificielle prendra des caractéristiques violentes. L’AFCIA entend anticiper ces mouvements sociaux de révolte afin d’offrir une alternative qui permette de canaliser le discours d’opposition à l’intelligence artificielle. L’AFCIA veut également **inverser la perception de l’intelligence artificielle auprès des jeunes,** qui peuvent être séduits par cette perspective, car ils n’en voient que l’intérêt à court terme sans prendre en considération le long terme. L’intelligence artificielle n’est pas une voie d’avenir ; c’est une voie qu’il faut combattre, et l’AFCIA veut instiller chez les jeunes le doute concernant les mécanismes qui sont à l’œuvre afin de les détourner de cette perspective. L’AFCIA cherche également à créer un réseau mondial, en traduisant leurs

---

arguments dans toutes les langues, en particulier en anglais, en chinois et en allemand, de manière à toucher un maximum de foyers de réflexion autour de ces problématiques. Il existe un très grand tissu d'associations et mouvements de pensée techno-critiques et critiques à l'encontre de la mainmise numérique ; mais l'AFCIA est la seule association à identifier l'intelligence artificielle comme le cœur du problème, d'où le besoin d'internationalisation de son action.

Cette volonté d'internationalisation de l'action de l'AFCIA est essentielle à la réussite de son projet de constitution d'un pôle d'opposition raisonné et non-violent à l'intelligence artificielle. Il n'y a aucun intérêt à garder cette réflexion nationale, qui n'aboutirait qu'à affaiblir un peu plus l'économie française, et n'aiderait pas à amplifier ce discours auprès des autres pays. Il faut adopter une réglementation progressive qui puisse avoir une portée mondiale. Il faut, pour cela, enclencher un mouvement à destination de l'opinion publique.

Il serait dans un premier temps possible d'imposer aux entreprises et laboratoires de déclarer quels sont leurs programmes de recherche en intelligence artificielle, afin que ces programmes puissent être contrôlés. Compte tenu des risques encourus, il est normal qu'il y ait un contrôle politique sur ces activités, au même titre qu'il y a un contrôle sur les activités atomiques. Ensuite, si un mouvement politique pouvait être rapidement constitué, il serait possible de se diriger vers l'interdiction mondiale des armes intelligentes, comme les drones, et accorder l'immunité aux objecteurs de conscience. Cependant, **le premier moteur de la recherche en intelligence artificielle est son application militaire** : aux États-Unis d'Amérique, la DARPA (*Defense Advanced Research Projects Agency*), qui est l'équivalent du ministère de la Défense, investit des sommes considérables dans des programmes de recherche en intelligence artificielle qui visent officiellement à appuyer l'action des militaires ; néanmoins, la perspective de nombreuses autres applications militaires possibles, comme le contrôle et l'utilisation de la pensée dans l'interface homme-machine, motive également ces investissements. Ainsi, interdire l'intelligence artificielle pourrait permettre de ralentir les investissements dans ce domaine.

En outre, de manière plus radicale, il faudrait **abolir la propriété intellectuelle sur les algorithmes d'intelligence artificielle**, ce qui couperait le désir de bien faire chez beaucoup d'acteurs économiques. La technique dans l'intelligence artificielle est tripartite entre les algorithmes, les données et la calibration, car c'est une science très empirique. Beaucoup de programmes et algorithmes sont en *Open source*, et simplement avoir accès à un programme ne permet pas d'en faire un système discriminant. En revanche, la connaissance dans la manière de calibrer et de faire converger les systèmes, en somme l'expertise telle que celle que l'on trouve notamment chez Google *DeepMind*, demeure cachée pour le moment. Faire tomber dans le domaine public l'ensemble du travail fourni par les entreprises privées dans la recherche en intelligence artificielle aura pour effet de tarir la volonté de ces acteurs économiques.

Il est également possible d'imaginer de limiter la puissance de calcul des systèmes à intelligence artificielle, en interdisant par exemple de faire fonctionner des intelligences artificielles qui dépasseraient le niveau d'intelligence d'une souris.

Enfin, pour l'ensemble des arguments développés ci-dessus, **l'AFCIA considère que la recherche en intelligence artificielle est illégitime, inacceptable, et par conséquent appelle l'interdiction pure et simple de la recherche en intelligence artificielle, ainsi que la mise à disposition à titre commercial ou à titre gracieux de systèmes à intelligence artificielle.**

Toutes ces réflexions auraient pu être menées en 1948, lorsque M. Norbert Wiener théorisa la cybernétique, ou dans les années 1960 lors de l'apparition des premiers calculateurs, car ces réflexions découlent de l'écriture d'algorithmes. Cependant, depuis quelques années, un point charnière a été atteint avec l'accélération des puissances de calcul et du développement des algorithmes. L'hiver de l'intelligence artificielle des années 1980, quand la discipline était décrédibilisée, est terminé ; la discipline est revenue en grâce du fait des avancées immenses qui ont été faites. L'AFCIA estime donc que maintenant est venu le moment d'engager ces réflexions et de lancer le débat, car la courbe d'évolution de l'intelligence artificielle est exponentielle, et dans cinq ou dix ans, il sera probablement déjà trop tard.

La plupart des grands scientifiques qui ont contribué à lancer l'intelligence artificielle comme discipline scientifique dans les années 1950, tels qu'Alan Turing ou Norbert Wiener, avaient compris que, à terme, l'intelligence artificielle ferait naître les risques et conséquences évoquées au cours de cet entretien ; mais cela demeurait théorique tant que la révolution numérique - que ces chercheurs n'ont d'ailleurs pas vécue - n'avait pas eu lieu.

Concernant la médecine et plus spécifiquement le transhumanisme, l'AFCIA distingue d'une part les interventions sur la physiologie de l'homme, et d'autre part les projets de modifications de ses capacités cognitives. Les premières sont justiciables, quant à leur évaluation éthique, des comités de bioéthiques. En revanche, l'AFCIA s'oppose par principe aux seconds dès lors qu'ils font appel à l'Intelligence Artificielle, car ils sont susceptibles de modifier la nature même de la conscience avec des conséquences imprévisibles.

Il y aura toujours des exemples positifs d'applications de l'intelligence artificielle dans la société, particulièrement dans le domaine de la médecine. Cependant, en prenant l'exemple de la médecine, la machine fonctionne à partir d'un apprentissage statistique, qui comporte toujours une marge d'erreur. Si la décision humaine comporte, elle aussi, une marge d'erreur, il demeure qu'appliquer l'intelligence artificielle dans le domaine médical dessaisira l'Homme de son pouvoir de contrôle et d'expertise. En

---

outre, les ingénieurs codant des machines en apprentissage automatique (*deep learning* ou *machine learning*) sont incapables de remonter aux causes de la décision d'un algorithme, alors qu'un médecin pourra toujours expliquer les raisons qui ont orienté son choix.

À partir du moment où l'homme considère que la machine est plus experte que lui, il va se dessaisir de son autonomie de jugement car il considérera que la machine a toujours raison, alors même qu'on ne comprend pas aujourd'hui pourquoi une machine prend telle ou telle décision. Les modèles de *deep learning* prenant en entrée (*input*) des milliards de données, il est impossible de savoir ce qu'il se déroule en sortie (*output*) avec un tel volume de données qui paramètre la décision de l'algorithme provoquant inévitablement un effet de boîte noire. En cela, la tragédie de l'humain dans une société à intelligence artificielle sera son inutilité sociale.

On ne sait pas quelle est la limite entre le vivant et l'artificiel. Il existe deux grandes écoles dans les neurosciences : les mécanistes, qui voient le cerveau comme un processeur se fondant sur des interactions électromagnétiques entre des molécules, et les « spiritualistes », qui pensent qu'il y aurait une composante qui ne se réduirait pas à un ensemble de calculs. Cependant, même en se positionnant du côté spiritualiste, les dangers présentés par le développement d'une intelligence dont le fonctionnement nous échapperait, qui ferait des choix semblables à ceux faits par une intelligence humaine tout en étant ontologiquement différente, sont concrets.

Avec les objets connectés, il y aura des analogues de systèmes autonomes dans toute la société et qui seront invisibles à nos yeux, et qui modifieront les rapports sociaux : un assureur qui refuse un contrat après que son algorithme eut mesuré le battement de cœur du client quelques jours plus tôt, un algorithme qui calcule la probabilité d'être un criminel à partir d'un visage, etc. **L'intelligence artificielle a une vocation totalitaire, elle cherche à englober tous les champs du réel.** Si l'individu a le choix ou non de mettre son sort entre les mains d'un algorithme en acceptant ou refusant de monter à bord d'une voiture autonome, il ne l'aura pas nécessairement dans de nombreux autres domaines où la présence de l'intelligence artificielle sera invisible. C'est pour cela que l'AFCIA est favorable à l'interdiction de l'intelligence artificielle, car les individus ne pourront pas choisir tout le temps de recourir ou non à des systèmes d'intelligence artificielle.

## V. AUDITIONS DU 28 NOVEMBRE 2016

### 1. Claude Berrou, professeur à Télécom Bretagne (Institut MINES-TELECOM), chercheur en électronique et informatique, membre de l'Académie des sciences

*M. Claude Berrou est principalement connu pour avoir inventé les premiers codes correcteurs d'erreur quasi-optimaux, au sens de la théorie de Shannon, appelés turbo-codes<sup>1</sup>. Ces turbo-codes sont massivement utilisés dans les troisième et quatrième générations de téléphonie mobile et dans de nombreux systèmes satellitaires. Cette invention a permis à M. Berrou d'être distingué en 2005 en recevant le prix Marconi, qui est considéré comme le prix Nobel des télécoms, et en 2007 en étant élu membre titulaire de l'Académie des sciences. À partir de 2009-2010, M. Berrou s'est concentré sur les neurosciences, en remarquant que le cerveau humain avait des propriétés de correction d'erreur : le cerveau humain est capable de corriger des approximations, des effacements, des inversions de lettres, de reconnaître un sens dans un discours, etc. Il a de fait soumis un programme de recherche à l'Agence Nationale de la Recherche sur les liens entre théorie de l'information, codage et cognition mentale. Cependant, l'ANR a refusé son programme ; M. Berrou a soumis donc son dossier au niveau européen auprès du Conseil européen de la recherche, qui a retenu son projet. M. Berrou a ainsi obtenu une bourse ERC Advanced Grant de l'ordre de deux millions d'euros de la part du Conseil européen de la recherche, ce qui lui a permis de composer une équipe de recherche interdisciplinaire réunissant un neuropsychologue, un psycholinguiste, des théoriciens de l'information, des électroniciens, des informaticiens, etc. **En cela, la composition d'équipes de recherche interdisciplinaire est indispensable pour avancer vers une véritable intelligence artificielle.***

L'intelligence artificielle est une expression très lapidaire. **L'intelligence naturelle est multiforme** mais il est quand même **possible de distinguer deux grandes catégories**. **La première grande catégorie d'intelligence est celle qui nous relie au monde extérieur**. Cette première catégorie d'intelligence est apparue avant la deuxième catégorie dans l'évolution, et que l'humain partage avec la plupart des espèces animales. Cette intelligence de premier type **regroupe ce qui concerne les sens, la perception, l'apprentissage, la reconnaissance et l'adaptation** - chacun est un être vivant dans un monde fluctuant, riche en détails, et cet être doit s'adapter et survivre. Aujourd'hui, dans la recherche de ce qui est appelé intelligence artificielle dans le monde de l'informatique, c'est en grande

---

<sup>1</sup> Le principe des turbo-codes est d'introduire une redondance dans le message afin de le rendre moins sensible aux bruits et perturbations subies lors de la transmission.

---

majorité l'intelligence des sens qui est recherchée. Par exemple, la voiture autonome nécessite une intelligence de la vision, du radar, etc. La voiture artificielle est une réussite en termes d'intelligence artificielle sur cette première catégorie de facultés permettant le lien au monde extérieur.

**La seconde grande catégorie d'intelligence** est propre à chaque individu. Elle englobe les facultés de l'esprit, celles qui permettent à l'humain d'imaginer, d'élaborer, d'inventer, de produire. Le seul modèle à disposition pour essayer de reproduire dans une machine des propriétés de cette intelligence créatrice est le cerveau. Le cerveau a une architecture qui est nettement différente de celle des ordinateurs classiques. Les processus de mémorisation et de traitement de l'information du cerveau s'entremêlent dans un même écheveau de connections synaptiques qui se comptent en milliers de milliards. **Cette intelligence de deuxième type peut être appelée intelligence artificielle générale.**

**Plus de 95 % de la communauté scientifique - mathématiciens, informaticiens, théoriciens de l'information - s'intéressent surtout aux sens.** Cela inclut la reconnaissance de visages, de scènes, etc. Sur ce point, la France est en retard. Cependant, sur l'intelligence de deuxième type, qui ne peut être que neuro-inspirée, la France a une carte très importante à jouer, à condition que soit mise en place une recherche pluridisciplinaire. L'association de neurosciences, d'informatique, de théories de l'information et d'électronique serait très efficace. Afin d'obtenir une intelligence artificielle neuro-inspirée, il faut s'abstraire du composant. Il est nécessaire d'avoir une vision de système.

Les GAFAs sont leaders dans le domaine de l'intelligence artificielle de premier type, dans laquelle ils investissent des milliards de dollars. Concernant l'intelligence artificielle de deuxième type, neuro-inspirée, le seul modèle est le cerveau humain et celui de quelques animaux. Il existera cependant probablement une intelligence artificielle de troisième type qui consistera à reproduire au niveau informationnel l'architecture du cerveau avec plus de modules spécialisés (essentiellement les sens) et généralistes (agrégation d'informations hétérogènes) que dans le cerveau humain. Pouvoir créer une intelligence artificielle de troisième type exigera un haut niveau de compréhension du cerveau humain et de ses modules spécifiques, ses *hubs*, ses circuits de communication ; mais il devra être possible d'en mettre encore plus que dans un cerveau humain. Il existe entre deux cents et trois cents modules dans le cerveau et trente milliards de neurones dans le cortex humain ; rien n'empêchera de créer des machines qui représenteront l'équivalent de trois cents milliards ou trois mille milliards de neurones. **Ce phénomène représentera la singularité technologique.**

**Les débats éthiques se construisant autour de l'intelligence artificielle sont nombreux.** Cependant, il semble que ces débats soient prématurés. Si les applications d'intelligence artificielle se limitent à de l'intelligence de premier type, **le débat éthique est prématuré** : le cas

typique du véhicule autonome pose certaines questions, mais pas davantage que concernant le cas d'un avion de chasse. Une machine préprogrammée, qui n'apprend pas de ses expériences, qui est non-apprenante, ne présente pas plus de dangers qu'un ordinateur de bureau. **Les algorithmes étant une succession d'opérations mathématiques, il n'y a pas de question déontologique ni éthique à se poser quant à leurs résultats.** Les questions doivent davantage se poser sur les applications.

**Le véritable enjeu pour la France est l'indépendance informatique,** qui est tout aussi stratégique que l'indépendance énergétique ou l'indépendance militaire. Cependant, concernant l'intelligence de premier type, au vu du retard accumulé, la France risque de ne pas être acteur mais simplement consommateur. En revanche, sur l'intelligence du deuxième et du troisième type, il est possible qu'un grand programme sur l'intelligence artificielle neuro-inspirée permette à la France d'acquérir son indépendance informatique. Sous une impulsion nationale, un programme de financement de recherche et développement, permettant d'aller jusqu'à l'application, est tout à fait possible. Les expertises en présence, la formation sont de qualité ; il demeure simplement des verrous à lever, en termes de nombre de postes, de financements difficiles à obtenir, etc.

## **2. M. Nicolas Cointe et Mme Fiona Berreby, chercheurs en thèse de doctorat sur l'éthique de l'intelligence artificielle**

*Les recherches de M. Nicolas Cointe et Mme Fiona Berreby sur la modélisation des raisonnements éthiques autour des agents autonomes artificiels dans le cadre du projet Ethicaa visent **trois objectifs : définir en quoi un agent autonome artificiel peut être éthique ; produire des représentations formelles de conflits éthiques et de leurs objets** – le conflit éthique peut être au sein d'un seul agent humain, entre un agent autonome artificiel et un utilisateur humain, ou entre plusieurs agents - ; et, enfin, élaborer des algorithmes d'explication et de compréhension pour les utilisateurs non-experts.*

*La question de l'éthique de l'intelligence artificielle peut être tout d'abord abordée sous la perspective de la représentation des conflits éthiques au sein d'un seul agent. La recherche développée par Mme Fiona Berreby consiste à modéliser des situations posant un dilemme éthique. L'une de ces situations posant un dilemme est appelée le « dilemme du tramway », qui est en deux parties.*

Dans la première partie, un train roule sur une voie et l'agent a le pouvoir d'aiguiller le train. Si l'agent n'aiguille pas le train et le laisse donc poursuivre sur sa lancée, cinq personnes se trouvant sur la voie seront tuées ; cependant, si l'agent décide d'aiguiller le train, ce dernier écrasera une seule personne se trouvant sur cette voie de déviation. Selon de nombreuses études exposant la première partie du dilemme du tramway à des individus interrogés sur leur choix éthique, ces derniers estiment en majorité **qu'il est**



---

**préférable d'adopter une décision utilitariste qui consiste à aiguiller le train afin de minimiser les pertes humaines.**

Dans la seconde partie, le train se dirige toujours vers cinq personnes mais l'agent, se trouvant sur un pont au-dessus de la voie, n'a plus le pouvoir d'aiguiller le train. Le seul moyen d'empêcher le train de tuer ces cinq personnes est de faire basculer depuis le pont une personne devant le train afin de sauver les cinq personnes. Cependant, les résultats de cette partie de ces études divergent de ceux de la première partie, car dans cette configuration, **les individus interrogés refusent le raisonnement utilitariste de basculement d'une personne devant le train**, solution qui privilégie la protection d'un maximum de vie humaine.

**Il existe de fait une déviation de la réflexion des individus sondés dans le dilemme du tramway.** La différence de considération éthique dans les deux parties, malgré le même rapport de cinq vies humaines contre une seule, peut s'expliquer par le fait que, dans la deuxième partie du dilemme, l'agent utilise la mort d'une personne comme une cause du sauvetage des cinq personnes, alors que dans la première partie, la mort de la personne n'est pas la cause mais la conséquence du sauvetage des cinq personnes. En outre, si dans la première partie, la mort d'une personne est une conséquence indirecte, cette mort nécessite dans la deuxième partie que l'agent commette directement un acte mauvais. **C'est ici la doctrine du double-effet qui modifie la perception éthique de ce dilemme** : bien que l'objectif soit de sauver un maximum de vies, le fait qu'une personne meure de la décision de l'agent trouble plus fortement cette décision lorsque l'agent doit directement agir - ici, provoquer la mort de cette personne - pour parvenir à son objectif.

**L'objectif d'expériences éthiques et de dilemmes appliqués au développement de l'intelligence artificielle est, à partir de règles éthiques préétablies, de permettre à l'agent autonome artificiel de calculer à partir des faits du monde et de ses règles éthiques la réaction acceptable par rapport à un dilemme.**

**La question de la représentation de l'éthique peut ensuite être abordée sous l'angle des systèmes multi-agents.** Un système multi-agents est un système qui comprend plusieurs entités intelligentes - aussi bien des humains que des agents autonomes artificiels, doués d'intelligence artificielle. **La discipline des systèmes multi-agents s'intéresse à la problématique de l'existence de systèmes qui comprennent des agents autonomes artificiels** et qui peuvent être homogènes - où les agents sont identiques - ou hétérogènes. De nombreux systèmes impliquent à la fois des humains et des agents autonomes artificiels, comme la défense, la finance, les transports et la santé, dans lesquels des dilemmes moraux sont en jeu.

**Le problème principal** abordé dans les systèmes multi-agents **est celui de la coopération d'agents autonomes artificiels avec des humains ou entre agents avec des éthiques différentes.** L'éthique des individus varie selon les individus, en fonction de la culture, de l'éducation, de la religion ;

les règles de valeurs et de morale peuvent différer mais coexister dans des systèmes multi-agents où interagissent des agents hétérogènes.

La morale et l'éthique sont distinctes. La morale décrit ce qui est bien ou mal dans l'absolu ; l'éthique est un moyen de réflexion qui permet de prendre des décisions justes dans un contexte précis, décisions prises selon une doctrine.

**Il existe trois grandes approches** formulées dans la littérature scientifique **sur la question des raisonnements éthiques des agents autonomes artificiels.**

**La première approche** fut apportée par un chercheur de l'armée nord-américaine, qui **proposait d'avoir une réflexion éthique avant la conception des robots, puis de restreindre le comportement des robots à uniquement des actions éthiques au regard de ce qui a été pensé avant sa conception.** C'est une forme de verrou qui est apposé sur la machine, ce qui suppose que la vérité en matière d'éthique appartienne entièrement au concepteur et interdit l'interaction éthique. Dans ce système, la machine n'a pas de représentation explicite de l'éthique ; elle n'est pas capable d'expliquer en quoi ses actions sont éthiques. Cette approche n'est pas générique : à chaque fois qu'un robot sera construit, le concepteur devra réfléchir à la dimension éthique de toutes ses actions, et si certaines sont omises, cela échappera au contrôle de la machine.

**La deuxième approche est celle de l'apprentissage (*learning*), qui implique que la machine n'a plus besoin d'être recodée à chaque fois.** Il suffit de lui montrer des cas différents afin qu'elle apprenne à agir éthiquement dans des contextes différents. Il n'y aurait cependant toujours pas de représentation explicite, la machine ne pouvant expliquer ses choix éthiques car elle n'a pas de représentation symbolique de ses connaissances. En outre, une machine qui va trop se fier au cas étudié lors de son apprentissage aura comme problème d'identifier des situations de manière trop précise par rapport à certaines autres et fera des biais de raisonnement. Au contraire, si une machine n'a pas rencontré lors de son apprentissage un cas auquel elle est confrontée, elle ne connaîtra pas la réaction appropriée.

**La troisième approche est l'approche de raisonnement symbolique.** C'est une approche fondée sur les règles, qui est générique, et permet donc à la machine de déduire de ses connaissances éthiques l'action appropriée. Cette approche permet aussi de fournir à la machine des connaissances explicites, en lui formulant des règles morales, en lui donnant des connaissances sur les valeurs et les situations. La machine sera de fait capable de déduire de manière exhaustive tout ce qu'elle peut déduire de manière logique à partir de cet ensemble. **L'objectif est de pouvoir expliquer à l'utilisateur humain pourquoi une action a été considérée comme juste, selon un ensemble de règles et de doctrines, en retraçant le processus de décision de l'algorithme.**

---

### 3. Mme Laurence Devillers, professeur d'informatique à l'université Paris-Sorbonne et directrice de recherche du CNRS au Laboratoire d'informatique pour la mécanique et les sciences de l'ingénieur (Limsi de Saclay)

Du fait du développement de l'intelligence artificielle et des systèmes experts fondés sur des algorithmes complexes, **il existe un besoin crucial de formation continue des individus sur les technologies du numérique**, à tout niveau et à tout âge. La révolution numérique est une évidence et afin d'éviter les spéculations et les peurs, mettre en œuvre une formation continue de l'individu sur les questions numériques permettrait de le guider vers l'évidence des bienfaits qu'il peut en tirer.

Dans la perspective du besoin essentiel d'éducation continue sur l'intelligence artificielle, la Maison Blanche a mis en place des formations en intelligence artificielle réservées aux journalistes afin que ces derniers puissent parler notamment de l'apprentissage automatique (*machine learning*) et du codage des algorithmes sans tomber dans des facilités angoissantes dues à un manque de connaissance de ces sujets.

De manière plus générale, **les personnes portant une parole publique et étant amenées à exprimer une opinion sur ces sujets doivent être au fait des connaissances**. En ce sens, **il est essentiel de simplifier et rendre plus accessible la connaissance, ce qui implique que les scientifiques prennent publiquement la parole**. Plus l'information et la connaissance seront partagées, moins il y aura d'inquiétude à propos de l'intelligence artificielle.

Cependant, un large accès à la formation continue en matière d'intelligence artificielle n'empêche pas qu'un niveau de sécurité certain doit être assuré, notamment concernant la protection des données lors de l'utilisation de logiciels « en nuage » (*cloud computing*). **Le besoin d'éducation nécessite, au-delà de la compréhension des systèmes, la compréhension des enjeux**. Le caractère virtuel du consentement concernant la protection des données personnelles amène souvent les utilisateurs à sous-estimer l'importance du consentement qu'ils donnent en un « clic ».

**Quatre leviers sont identifiés par l'Initiative pour l'éthique des systèmes autonomes de l'Institut des électriciens et électroniciens (IEEE)** pour affirmer la place de l'éthique dans le développement de systèmes d'intelligence artificielle. Tout d'abord, **le premier levier est l'éducation** : les individus doivent être conscients de ce qu'est un système d'intelligence artificielle, fondé sur des algorithmes, sous ses différentes formes. **Le deuxième levier consiste à élaborer des règles morales à intégrer aux systèmes automatiques d'intelligence artificielle**. **Le troisième levier revient à se doter d'outils de vérification, d'évaluation du traitement et de mémorisation des données**. Les algorithmes ne peuvent être complètement transparents, mais il est possible de les évaluer de différentes manières. Enfin, **le quatrième levier repose sur l'existence de règles juridiques en cas d'abus**.

---

**L'apprentissage des machines se différencie des notions technologiques fondamentales**, par rapport à la notion d'évaluation notamment. L'apprentissage d'une machine et l'enrichissement de son algorithme ne sont pas transparents. Dans le cas de l'apprentissage profond (*deep learning*), ce sont des matrices du chiffre qui structurent l'algorithme de la machine, elles ne peuvent donc être transparentes. Il faut de fait trouver de nouveaux moyens d'évaluer qui soient reproductibles. De plus, si la machine s'adapte, si elle ajoute de nouvelles données à son modèle, son modèle change. Ainsi, **si l'évaluation de la machine est reproductible malgré ses évolutions, alors un objectif de transparence est atteint**. Ce n'est pas le système qui peut être évalué de manière transparente, mais la manière dont réagit le système dans différents contextes à des temps différents d'utilisation.

En outre, **les GAFAMI ont une stratégie de récupération massive des données**. Ces données ainsi que les algorithmes et technologies utilisées par ces industries ne sont pas, pour la plupart, mises à la disposition des scientifiques, ces derniers ne constituant de fait plus des garde-fous. Il est possible d'illustrer ce point avec l'exemple de la prise en considération des enjeux éthiques. Même si les géants nord-américains de l'industrie numérique (Google, Amazon, Facebook et IBM) s'intéressent à des problématiques éthiques du développement de l'intelligence artificielle, il demeure qu'ils traitent de ces sujets entre eux, sans que la communauté scientifique n'ait de visibilité ni puisse vérifier leurs travaux.

---

## VI. AUDITIONS DU 30 NOVEMBRE 2016

### 1. M. David Sadek, directeur de la recherche de l'Institut Mines-Télécom, spécialiste en intelligence artificielle

L'intelligence artificielle n'est pas un domaine nouveau ; il a connu des périodes fastes et des hivers de déclin. De nombreuses questions d'encadrement de l'intelligence artificielle sur le plan éthique se posent.

**Dans cette période de renaissance de l'intelligence artificielle, certaines technologies ont atteint un niveau de maturité qui semble pouvoir tenir certaines promesses.** Cependant, le potentiel de progrès en l'intelligence artificielle demeure important ; et la France y a un rôle à jouer puisque le potentiel français est garanti par des équipes de recherche de niveau mondial dans les différents domaines de l'intelligence artificielle. **L'intelligence artificielle n'est pas une discipline monolithique ; c'est un ensemble de sous-disciplines et est donc, par construction, pluridisciplinaire.**

Une confusion est régulièrement commise à propos de l'intelligence artificielle résumée au simple processus d'apprentissage profond (*deep learning*) d'un système informatique. Des avancées technologiques indéniables et remarquables ont été permises par les techniques d'apprentissage profond ; néanmoins, **il est essentiel d'éviter toute confusion entre intelligence artificielle et apprentissage profond.** Il existe, *a minima*, deux grands pans de l'intelligence artificielle. Un premier pan serait le volet « stochastique », qui repose sur le traitement des données et d'apprentissage. Les avancées dans ce domaine ont été permises grâce au traitement et au croisement de données, ainsi qu'à l'augmentation de la puissance de calcul des machines et des capacités algorithmiques.

Le second volet, qualifié de « cognitiviste », relève davantage de la formalisation de modèles de comportements *a priori*. Les travaux s'inscrivant dans ce second volet se fondent sur des modèles de raisonnement, d'inférence automatique, de représentation sémantique des connaissances et d'ontologie. Ce second volet est partiellement éclipsé par l'aspect « *deep learning* » ; cependant, la France dispose également d'équipes brillantes dans ce second volet, et il est essentiel de valoriser ces équipes à leur juste niveau. **La force de frappe de la France en matière d'intelligence artificielle sera d'autant plus importante si elle se positionne sur une hybridation des deux approches.**

Force est de constater que le domaine de l'intelligence artificielle est éclairé par la présence de grands acteurs industriels mondiaux qui disposent d'une force de frappe importante. **De fait, il est nécessaire de voir comment le système industriel français peut produire de la valeur au niveau**

**économique et sociétal sur la base des activités de recherche et d'innovation.** Un véritable écosystème de jeunes pousses du numérique tend à émerger en France, la plupart de ces jeunes entreprises étant très performantes. Dans cette perspective, il est essentiel que les jeunes entreprises émergentes françaises ne se fassent pas happer par les grands groupes industriels nord-américains ou chinois, bien que bon nombre de créateurs de « *start-ups* » aient cette ambition. **Les dispositifs de soutien à la recherche et à l'innovation se doivent d'être suffisamment incitatifs pour que le fossé entre les résultats de la recherche et la production de valeur économique soit comblé de la manière la plus efficace possible.** Dans cette perspective, il est impératif de créer une boucle vertueuse entre la recherche et l'innovation, et de développer en France une culture de l'entrepreneuriat à partir du monde académique.

Lors des deux hivers de l'intelligence artificielle, les équipes de recherche françaises se sont dispersées ; néanmoins, **il est essentiel de faire en sorte que ces équipes retrouvent une synergie afin que la France retrouve sa force de frappe.** L'action de l'Association française pour l'intelligence artificielle (AFIA) pourrait participer à la redynamisation de cette synergie ; il semble cependant que cette association représente davantage un réseau qu'une véritable plateforme opérationnelle. **Les pouvoirs publics doivent favoriser l'émergence de dispositifs et de structures reflétant une dynamique, une volonté et une ambition en matière d'intelligence artificielle.** La mise en relation au sein d'un même écosystème des jeunes pousses, de certains grands groupes industriels qui font de l'intelligence artificielle un axe stratégique, et la dynamique de recherche qui se met en place sont une opportunité que la France se doit de saisir.

Les techniques d'apprentissage excellent dans le développement de la perception, incluant la vision, la reconnaissance de la parole ou le traitement des images, dans lequel il existe de nombreux résultats opérationnels et industriels remarquables et probants. Cependant, concernant la cognition, la compréhension et le traitement du sens, qui caractérisent l'intelligence humaine, ces techniques d'apprentissage n'ont pas encore démontré leur puissance. **La réalisation d'avancées concrètes de l'intelligence artificielle ne pourra émerger que grâce à l'hybridation des deux approches évoquées précédemment.**

Concernant les aspects éthiques de l'intelligence artificielle, la question de l'encadrement du développement de l'intelligence artificielle est primordiale car il existe des comportements délibérément malveillants. Dans cette perspective, certaines données peuvent être utilisées à des fins qui ne sont pas celles pour lesquelles elles ont été produites. En outre, le fait de ne pas maîtriser les décisions et actions des algorithmes et programmes d'intelligence artificielle peut être source de questionnements éthiques. En effet, ces programmes peuvent avoir des comportements non-souhaitables sans que ceux-ci ne soient délibérément malveillants.

---

À titre d'exemple est cité un cas rencontré par M. Sadek au cours de ses études qui est celui d'un programme de bataille navale utilisé par le Département de la défense des États-Unis d'Amérique pour entraîner ses amiraux. Il fut constaté que le programme, dont la consigne programmée était de maximiser la vitesse de sa flotte, coulait systématiquement les bateaux de sa flotte qui étaient touchés, car ceux-ci étaient affaiblis et donc moins rapides. La consigne d'éliminer les bateaux touchés de sa flotte n'était pas programmée dans son algorithme ; cependant, le programme a de lui-même effectué cette inférence relativement simple.

Ce cas démontre **le manque de contrôle sur ce qui est écrit dans les programmes complexes, qui comportent beaucoup d'implicite**. En cela, la fiabilité et la conformité de fonctionnement des programmes par rapport à ce qui est strictement prévu constitue un domaine de recherche en soi et implique de nombreux questionnements éthiques. Dans le cadre de l'accélération technologique qu'entraînent les progrès de la recherche en intelligence artificielle, ces travaux de recherche éthiques sur la malveillance non-délibérée de programmes ne sont pas assez abordés et valorisés, notamment en France.

## **2. M. Dominique Sciamma, directeur de l'école de design « Strate » à Sèvres**

Aux débuts de la recherche sur l'intelligence artificielle, les enjeux se concentraient sur des problématiques de programmation. **Une école de pensée, appelée le « connexionnisme », proposait une approche ascendante de l'intelligence artificielle, qui reposait sur le fait que l'intelligence est une émergence**, impliquant donc que **c'est l'interaction des composants qui fait émerger l'intelligence**. Cette école est rapidement montée en puissance dans le monde de la recherche. Cependant, dès la fin des années 1960, un chercheur du *Massachusetts Institute of Technology*, M. Seymour Papert, a démontré que de nombreux problèmes ne pouvaient être résolus par l'approche connexionniste. Les résultats de cette démonstration ont contribué à l'arrêt des programmes de financement de l'école de pensée connexionniste. Cet échec de la théorie connexionniste a favorisé **l'approche descendante de l'intelligence artificielle, celle des systèmes experts qui a permis un renouveau de la recherche dans l'intelligence artificielle** jusqu'à la découverte de ses limites qui ont plongé cette recherche dans un deuxième « hiver » dans les années 1990.

**L'intelligence artificielle se répand de manière massive à travers Internet et les objets**. À terme, les humains seront entourés d'objets qui seront capables de percevoir le monde, grâce à des capteurs. Ces objets auront une représentation du monde et ils pourront croiser leur représentation du monde avec les données qu'ils auront captées pour prendre des décisions. En raison de l'existence de ces trois éléments, ces objets peuvent être qualifiés de robots. **Les humains seront amenés à**

**partager leur monde avec des objets capables de prendre des décisions, et qui prendront dans certains domaines des décisions à leur place** – comme le montre l'émergence des voitures autonomes par exemple. De fait, le rapport des humains aux machines sera altéré : **une véritable relation entre la machine et l'humain, tous deux capables de décider et de dialoguer, se substituera au contrôle exercé par l'humain sur la machine.**

Les objets robotisés, les « robjets », s'invitent dans la vie quotidienne des utilisateurs ; c'est la première fois que l'humanité se trouverait confrontée à des objets non plus définis et caractérisés par des formes mais par des comportements. Ce ne seront plus des objets de contrôle ni des extensions du corps humains, ils seront dans la position d'un majordome et vont, sur la base de certaines informations, prendre des décisions. Les êtres humains accueilleront dans leur environnement proche des objets adoptant un comportement humain, avec une forme d'empathie.

**Un rapport de relation va se substituer à un rapport de contrôle : les robots seront capables d'observer, de prendre des décisions et de dialoguer avec les êtres humains.** Certains objets actuels, grâce à de nombreuses technologies notamment biométriques, sont capables d'observer les réactions des êtres humains ; cependant, le rapport de relations sera bouleversé par le caractère empathique qu'auront intégré les robots dans leur rapport aux êtres humains. **L'intelligence artificielle peut s'incarner et s'adapter aux êtres humains avec qui elle est en relation en fonction de leurs caractères et manières de vivre.**

**Cette situation nouvelle impliquera de façonner les nouvelles relations qu'entretiendront les êtres humains avec les robots.** Cela relève non d'une problématique de marché ou d'une question technique d'ingénierie, mais du *design* comportemental et du *design* de situation. De fait, il est nécessaire de réunir des spécialistes de la gériatrie, du personnel d'assistance, des ingénieurs, des sociologues, des spécialistes du comportement, et des *designers* pour fournir un travail d'observation, de scénarisation, d'hypothèses et de tests sur le comportement des robots dans une situation sociale.

Au-delà des enjeux comportementaux et d'acceptation sociale de l'intelligence artificielle, notamment de sa matérialisation physique, **il existe également de nombreux enjeux économiques et industriels.** La France est, depuis toujours, puissante en matière de mathématiques, d'informatique et d'intelligence artificielle ; à ce titre, l'INRIA est un lieu de production intellectuelle et d'innovation remarquable. Cependant, si les premières commercialisations de produits informatiques ne visaient qu'un public restreint et ne représentaient donc pas un enjeu économique majeur, c'est aujourd'hui plusieurs milliards d'individus qui utilisent ce type d'objets.

En ce sens, **les enjeux d'acceptation sociale et de désirabilité des objets, qui ne sont pas techniques mais économiques, sont déterminants, et reposent sur des critères d'expérience et d'imaginaire associés.** Par exemple, il est possible de citer le cas de l'entreprise Apple :



---

indépendamment des qualités techniques des produits, il existe un imaginaire associé à la marque Apple, amenant le consommateur à préférer, à qualité et à esthétique égale, un produit Apple à un produit Samsung. Ce travail profond de la relation à l'objet, de son inscription durable dans l'environnement humain et jusqu'à la désirabilité de l'objet, relève des *designers*.

De fait, il existe un enjeu industriel fort. **Il est essentiel que les industriels français se mettent en position de produire des objets désirables, achetés dans le monde entier.** L'industrie française a déjà démontré, à maintes reprises, sa capacité à produire des objets désirés dans le monde entier, à l'instar du succès que rencontrent les drones de l'entreprise Parrot et les objets connectés de la marque Withings. L'industrie française est remarquable en matière de *design* des produits ; l'expertise française n'est plus à démontrer.

Cependant, la capacité des entreprises françaises à transformer une expertise scientifique et technique en objets industriels pose question, car cette transformation ne repose pas uniquement sur l'excellence technique. Cette transformation qui ne relève pas du *marketing* mais du *design*, de la capacité à s'inscrire dans la situation de vie humaine et donc à anticiper les décisions des robots, est, sur le plan éthique, indispensable. **L'ivresse technique doit être tempérée par une réflexion sur les impacts politiques et sociaux de ces « robjets », et sur les outils de contrôle à disposition des utilisateurs et des pouvoirs publics.**

**La question du droit des robots représente également un enjeu à prendre en considération.** Il existe deux écoles d'intelligence artificielle sur cette question : l'école de l'intelligence artificielle forte (dont je suis un tenant), et l'école de l'intelligence artificielle faible. **L'école de l'intelligence artificielle forte postule que l'Homme est capable de réaliser avec l'intelligence artificielle ce que la Nature a été capable de réaliser avec lui ; de fait, à terme, l'Homme produira des machines conscientes.** Dans cette perspective, la question du droit des robots se posera en ce que ces derniers seront doués de conscience, et donc potentiellement porteurs de revendications en matière de droits fondamentaux.

L'introduction de robots doués d'intelligence artificielle forte peut amener de nombreux débats philosophiques, éthiques, et juridiques. Il est important d'envisager le type de société qui découlera de l'avènement d'une intelligence artificielle forte : les Hommes et les machines entreront-ils en compétition ? L'intelligence artificielle sera-t-elle susceptible de se substituer à l'intelligence humaine ? **L'importance du débat et les inquiétudes entourant le développement de l'intelligence artificielle se justifient par le fait que cette question touche à quelque chose de profondément humain, et nécessite une réflexion éthique, politique, juridique et sociale approfondie.**

---

**3. M. François Taddéi, directeur du Centre de recherches interdisciplinaires (Inserm, université Paris-Descartes), biologiste**

Il est important que les acteurs publics s’emparent du sujet de l’intelligence artificielle, afin de ne pas laisser les seules sociétés privées, principalement nord-américaines et chinoises, nourrir de l’intérêt pour ce thème. **L’intelligence artificielle va modifier les sociétés humaines en profondeur sans que celles-ci ne le comprennent nécessairement ni ne puissent en minimiser les dérives potentielles.** Il existe de fait un besoin important de formation. Par ailleurs, il est étonnant de constater que davantage de recherches sont menées sur l’intelligence artificielle et l’apprentissage automatique que sur l’intelligence humaine et l’apprentissage des enfants, ce que démontre l’importance des intérêts économiques dans les orientations de recherche.

Cependant, **la combinaison entre l’intelligence humaine et l’intelligence artificielle peut permettre aux humains, au-delà de la création de machines, d’accomplir des progrès qu’ils ne sauraient atteindre seuls.** Dans ce contexte, les réflexions éthiques, politiques et sociétales sont essentielles, afin que celles-ci permettent de penser l’intelligence collective entre les humains et les machines.

Face à l’accélération des progrès dans l’intelligence artificielle, les systèmes d’apprentissage profond et l’apprentissage automatique, les systèmes d’éducation, conçus au XIX<sup>e</sup> siècle, ne sont plus en mesure de s’adapter à ces évolutions toujours plus rapides, que ce soit sur le plan de la société, du marché de l’emploi, ou des défis technologiques, scientifiques et éthiques. **Si la société évolue plus vite que son système éducatif, ce système éducatif deviendra rapidement obsolète.** Il apparaît ainsi essentiel d’adapter le système éducatif à cet état de fait, en apprenant aux élèves non seulement à penser sans la machine, mais également à penser avec la machine. Le système éducatif actuel met en compétition les élèves sur les savoirs d’hier ; le défi sera de construire le savoir de demain et de créer les machines qui permettront la construction de ce savoir de demain. Il est de fait nécessaire de repenser le système éducatif, dont l’évolution ne peut se décréter de manière verticale, mais plutôt de manière plus horizontale, en faisant appel à l’intelligence collective de tous les acteurs éducatifs, le ministère de l’éducation nationale, les enseignants, les chercheurs et les parents d’élèves.

**En réponse aux défis de l’intelligence artificielle, il est important de mobiliser l’intelligence collective humaine et toutes les générations à tous les niveaux.** Selon Aristote, les trois piliers essentiels de la connaissance : *épistèmè* - qui a donné « science » -, qui repose sur la connaissance du monde ; *technè* - qui a donné « technologie » -, qui repose sur la manière d’agir sur le monde ; et *phronésis*, qui est l’éthique de l’action. **Or, à mesure que les savoirs s’accumulent et qu’ils permettent d’agir sur le monde, il est important de s’interroger sur l’éthique des actions.** Les

humains ont fait croître de manière exponentielle leurs connaissances et leur capacité à agir sur le monde, mais très peu leur capacité à penser les implications individuelles et collectives de leurs choix, tant sur le court terme et le long terme, que sur les plans locaux et globaux. L'avènement de l'intelligence artificielle et de la machine impliquent une plus grande considération de l'éthique de l'action. Les humains qui créeront ces machines, qui seront douées de *technè* mais très peu de *phronésis*, devront interroger l'éthique de leur action.

#### 4. M. Igor Carron, entrepreneur, organisateur du principal « *meet-up* » en intelligence artificielle en France intitulé « *Paris Machine Learning* »

L'évolution de l'intelligence artificielle a connu plusieurs vagues, notamment aux États-Unis d'Amérique, les périodes séparant chacune de ces vagues étant qualifiées « d'hiver de l'intelligence artificielle ». **Les hivers de l'intelligence artificielle sont les périodes où le financement public nord-américain des activités de recherche en intelligence artificielle s'est estompé drastiquement.** Les réductions conséquentes de ces financements publics qui ont entraîné ces hivers s'expliquent par le fait que les pouvoirs publics estimaient que certaines méthodes utilisées ne permettaient pas d'atteindre les objectifs promis. Si la recherche en intelligence artificielle a connu un renouveau dans les années 1980, les financements octroyés à la recherche en intelligence artificielle par la *DARPA - Defense Advanced Research Projects Agency* - se sont à nouveau asséchés au début des années 1990, les algorithmes construits présentant de nombreuses limites.

La communauté scientifique de recherche en intelligence artificielle a continué à survivre de manière dynamique. **L'avènement de Google en 1998 constitue, d'une certaine manière, un nouveau point de départ pour la recherche en intelligence artificielle.** En effet, l'entreprise, dès ses débuts, a utilisé un algorithme permettant de donner des résultats aux humains. Les moteurs de recherche existants dans les années 1990, tels que ceux de Yahoo! Portal, utilisaient des humains pour effectuer une partie des classifications. **L'émergence du moteur de recherche de Google a démontré que l'humain n'était pas indispensable pour produire des pages de résultats de recherche.** En ayant recours à des algorithmes efficaces, Google a capté les utilisateurs des autres plateformes de moteur de recherche et s'est placée en situation de quasi-monopole dont le moyen de financement principal est la publicité ciblée, qui utilise également des algorithmes. **Dès 2001, les dirigeants de Google indiquaient que les activités de moteur de recherche constituaient une forme d'intelligence artificielle.**

La majorité des algorithmes développés et utilisés actuellement n'offre pas la compréhension des décisions qu'ils prennent pour les humains. Il est de fait essentiel de construire des algorithmes permettant de comprendre et décoder les décisions prises par d'autres algorithmes, quitte à découvrir pourquoi ces algorithmes ont des biais ou pourquoi ils sont discriminants.

**5. M. Jill-Jênn Vie, chercheur en thèse de doctorat à l'École normale supérieure Paris-Saclay**

Le principe de l'intelligence artificielle consiste à inférer des faits à partir de données qui n'ont pas été explicitement programmées. Le développement d'une intelligence artificielle amène la question de l'objectif à optimiser, que ce soit le profit, l'économie d'énergie – à l'instar de l'entreprise Google *DeepMind* qui utilise son système d'apprentissage profond pour optimiser sa consommation d'énergie et réduire de 40 % la consommation d'énergie de ses centres de données –, ou encore le nombre de votes obtenus à une élection, certaines applications permettant d'optimiser l'efficacité des déplacements de campagne électorale d'un candidat.

L'utilisation de l'intelligence pour ces types d'applications a été renforcée grâce aux mécanismes d'apprentissage profond (*deep learning*). L'apprentissage profond permet entre autres à une intelligence artificielle d'effectuer de la reconnaissance d'image, de la détection d'éléments ou d'êtres vivants, de la légende automatique et de la génération d'image.

Exemple de reconnaissance d'image



Exemple de détection





Exemple de légende automatiqueExemple de génération d'image

**Le code source permet l'exécution de l'algorithme : il renferme l'ensemble des séries d'opérations que doit effectuer la machine dans un ordre prédéfini.** La machine va compiler le code source dans un langage, qui n'est certes pas compréhensible par l'humain, mais qui lui permet d'exécuter plus rapidement les tâches qui lui sont assignées. Un exemple d'algorithme permettant la résolution d'un problème est celui dit du « parcours main gauche », qui consiste à suivre le mur à sa gauche afin de sortir d'un labyrinthe. Cependant, de nombreux concepteurs d'algorithmes cherchent à trouver des cas spécifiques pour lesquels l'algorithme ne pourra pas résoudre le problème.



n'était que le revendeur de ces T-shirts, ces vêtements étant produits par une entreprise utilisant un algorithme qui créait de manière aléatoire les slogans inscrits. Ce produit n'avait pas été pensé, mais créé par une intelligence artificielle.

De même, certains livres vendus sur la plateforme de vente en ligne sont des ouvrages entièrement créés par des algorithmes, qui n'existent pas mais qui sont issus d'un procédé de « génération automatique de texte ». De fait, les acheteurs, pensant acquérir un ouvrage rédigé par un homme, reçoivent un livre dont le texte est automatiquement et aléatoirement produit depuis une base de données par un algorithme. L'étendue du nombre de cas similaires à ceux-ci s'explique par le fait que l'impression à la demande, et plus largement la fabrication à la demande, est une pratique régulièrement utilisée par les vendeurs présents sur Amazon, afin de rationaliser les coûts de production. Ainsi, les articles, aussi illicites soient-ils, n'existent pas physiquement avant qu'un acheteur ne les commande.

Les cas d'**erreurs d'implémentation** peuvent, pour leur part, également avoir des conséquences sur les conditions d'utilisation et le niveau de sécurité d'une machine. Il est possible de mobiliser le cas de la PlayStation 3, la console de jeux vidéo produite par la société japonaise Sony. Afin de sécuriser l'utilisation des jeux vidéo et lutter contre le piratage, la console ne peut exécuter que des jeux bénéficiant d'une licence. La vérification de la licence repose sur un nombre secret. Durant la phase d'implémentation, les développeurs devaient choisir un nombre au hasard pour chaque nouveau jeu ; cependant, ces développeurs ont commis l'erreur d'utiliser le même nombre secret pour chaque nouveau jeu. Cette erreur fut exploitée par des hackers qui purent résoudre une équation du premier degré et trouver ce nombre secret, leur permettant de faire lire un jeu détenteur de licence sur leur console. À la suite de la publication sur Internet de ce nombre, des poursuites ont été engagées par l'entreprise, et le nombre fut considéré comme un périphérique de contournement de sécurité, ce qui rendait illégale toute publication et transmission de ce nombre.

De nombreuses plateformes de vente utilisent des algorithmes ayant recours à des formes de discrimination volontaire avec une fixation de prix dynamique. Cette technique de fixation de prix dynamique est notamment utilisée par les plateformes de vente de billets des compagnies de transports aériens et terrestres commerciaux.

Il est également possible de recourir à des pratiques de camouflage volontaire (*obfuscation*), qui consistent à ne pas ouvrir un code source de manière délibérée. Cette technique est notamment utilisée dans des applications de prédiction de crimes (*PredPol*), dans l'affectation d'un degré de menace contre les habitants lors des patrouilles de police à partir des plateformes *Beware* ou *Twitter*, ou encore dans le diagnostic de personnes à risque de pédophilie avec la plateforme *Abel Assessment*.



Tous ces systèmes boîte noire se cachent derrière le prétexte qu'une ouverture du code pourrait influencer un comportement de contournement de la part de malfaiteurs, mais leur légitimité doit être remise en question.

Néanmoins, **l'utilisation d'algorithmes peut amener à des discriminations involontaires exercées sans que cela ne soit programmé en amont** comme, par exemple, avec l'application *Street Bump*, qui vise à détecter les nids-de-poule sur la route. L'effet discriminatoire involontaire induit par l'utilisation de cette application réside dans le fait que la présence de nids-de-poule sera détectée de manière plus rapide dans les quartiers où réside une population aisée. En effet, davantage d'habitants disposent des moyens financiers pour détenir un *smartphone capable* d'exécuter cette application dans les quartiers aisés que dans les quartiers populaires.

En connaissant les règles qui caractérisent l'algorithme, on peut repérer ce type de biais et y réagir. Sinon, on ne peut pas.