

N° 4594

ASSEMBLÉE NATIONALE

CONSTITUTION DU 4 OCTOBRE 1958
QUATORZIÈME LÉGISLATURE

Enregistré à la présidence de l'Assemblée
nationale

le 15 mars 2017

N° 464

SÉNAT

SESSION ORDINAIRE 2016 - 2017

Enregistré à la présidence du Sénat

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

Tome I : Rapport

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

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 n'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>
SYNTHÈSE DU RAPPORT	11
INTRODUCTION	15
PREMIÈRE PARTIE : ÉLÉMENTS DE CONTEXTE	19
I. LA DÉMARCHE DE VOS RAPPORTEURS	19
A. DE LA PROCÉDURE DE SAISINE À L'ADOPTION D'UN CALENDRIER DE TRAVAIL	19
1. <i>L'origine et l'instruction de la saisine</i>	19
2. <i>Un calendrier de travail très resserré</i>	20
B. LE CHAMP DES INVESTIGATIONS DE L'ÉTUDE	20
1. <i>L'étude de faisabilité du rapport conclut à la pertinence d'une étude spécifique de l'OPECST</i>	20
2. <i>Un ciblage délibéré sur un nombre limité de problématiques et de pistes d'investigation</i>	23
C. LA MÉTHODE DE TRAVAIL	26
1. <i>Une méthode de travail fondée sur des auditions bilatérales et des déplacements en France et à l'étranger</i>	26
2. <i>L'organisation d'une journée d'auditions publiques au Sénat le 19 janvier 2017</i>	27
3. <i>La consultation d'ouvrages, de rapports et d'articles parus sur le sujet</i>	28
II. L'HISTOIRE DES TECHNOLOGIES D'INTELLIGENCE ARTIFICIELLE ET DE LEURS USAGES	31
A. DES TECHNOLOGIES NÉES AU MILIEU DU XX^E SIÈCLE	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^e 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. L'ACCÉLÉRATION RÉCENTE DE L'USAGE DES TECHNOLOGIES D'INTELLIGENCE ARTIFICIELLE GRÂCE AUX PROGRÈS EN APPRENTISSAGE AUTOMATIQUE (« MACHINE LEARNING »)	54
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.....	54
2. L'apprentissage profond connaît un essor inédit dans les années 2010 avec l'émergence de la disponibilité de données massives (« big data ») et l'accélération de la vitesse de calcul des processeurs	56
3. Les technologies d'intelligence artificielle conduisent, d'ores et déjà, à des applications dans de nombreux secteurs	69
4. Par leurs combinaisons en évolution constante, ces technologies offrent un immense potentiel et ouvrent un espace d'opportunités transversal inédit	79
5. L'apprentissage automatique reste encore largement supervisé et fait face au défi de l'apprentissage non supervisé	81
III. CARACTÉRISTIQUES GÉNÉRALES DE LA RECHERCHE EN INTELLIGENCE ARTIFICIELLE ET ORGANISATION NATIONALE EN LA MATIÈRE	82
A. LES CARACTÉRISTIQUES DE LA RECHERCHE EN INTELLIGENCE ARTIFICIELLE	82
1. La place prépondérante de la recherche privée, dominée par les entreprises américaines et, potentiellement, chinoises	82
2. Une recherche essentiellement masculine	86
3. Une interdisciplinarité indispensable mais encore insuffisante	87
4. Une recherche soumise à une contrainte d'acceptabilité sociale assez forte sous l'effet de représentations catastrophistes de l'intelligence artificielle	93
5. Une recherche en intelligence artificielle qui s'accompagne de plus en plus d'interrogations et de démarches éthiques	98
B. TABLEAU DE LA RECHERCHE FRANÇAISE EN INTELLIGENCE ARTIFICIELLE.....	98
1. De nombreux organismes publics interviennent dans la recherche en intelligence artificielle	98
2. Quelques exemples de centres, de laboratoires et de projets de recherche	101
3. Une reconnaissance internationale de la recherche française et qui s'accompagne d'un phénomène de rachat de start-up et de fuite des cerveaux lié aux conditions attractives offertes à l'étranger.....	103
4. Une communauté française de l'intelligence artificielle encore insuffisamment organisée et visible.....	104
5. La sous-estimation des atouts considérables de la France et le risque de « décrochage » par rapport à la recherche internationale en intelligence artificielle	106
DEUXIÈME PARTIE : LES ENJEUX DE L'INTELLIGENCE ARTIFICIELLE.....	111
I. LES CONSÉQUENCES ÉCONOMIQUES ET SOCIALES DE L'INTELLIGENCE ARTIFICIELLE	111
A. D'IMPORTANTES TRANSFORMATIONS ÉCONOMIQUES EN COURS OU À VENIR.....	111
1. L'évolution vers une économie globalisée dominée par des « plateformes »	111
2. Un risque de redéfinition, sous l'effet de ce nouveau contexte économique, des rapports de force politiques à l'échelle mondiale	115

3. Des bouleversements annoncés dans le marché du travail : perspectives de créations, d'évolutions et de disparitions d'emplois	116
B. LA SOCIÉTÉ EN MUTATION SOUS L'EFFET DE L'INTELLIGENCE ARTIFICIELLE	122
1. Les défis lancés par l'intelligence artificielle aux politiques d'éducation et de formation continue	122
2. Une révolution potentielle de notre cadre de vie et de l'aide aux personnes	123
3. Le défi de la cohabitation progressive avec des systèmes d'intelligence artificielle dans la vie quotidienne	125
II. LES QUESTIONS ÉTHIQUES ET JURIDIQUES POSÉES PAR LES PROGRÈS EN INTELLIGENCE ARTIFICIELLE	128
A. LES ANALYSES PRÉSENTÉES PAR D'AUTRES INSTANCES POLITIQUES	128
1. Les deux rapports issus des institutions de l'Union européenne : Parlement européen et Comité économique et social européen (CESE)	128
2. Les trois rapports de la Maison Blanche	129
3. Le rapport de la Chambre des Communes du Royaume-Uni	133
4. Les initiatives chinoises et japonaises en intelligence artificielle accordent une place contrastée aux questions éthiques	135
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	140
B. DES « LOIS D'ASIMOV » À LA QUESTION CONTEMPORAINE DE LA RÉGULATION DES SYSTÈMES D'INTELLIGENCE ARTIFICIELLE	142
1. Dépasser les « lois d'Asimov » pour envisager un droit de la robotique	142
2. Les questions juridiques en matière de conception (design), de propriété intellectuelle et de protection des données personnelles et de la vie privée	144
3. Les divers régimes de responsabilité envisageables et ceux envisagés	153
4. Les différenciations du droit applicable selon le type d'agents autonomes : robots industriels, robots de service, voitures autonomes et dilemmes éthiques afférents	157
C. LA PRISE EN COMPTE GRANDISSANTE DES ENJEUX ÉTHIQUES	162
1. Le cadre national de la réflexion sur les enjeux éthiques de l'intelligence artificielle	162
2. Les nombreuses expériences anglo-saxonnes de réflexion sur les enjeux éthiques de l'intelligence artificielle	173
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)	185
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	189
III. LES QUESTIONS TECHNOLOGIQUES ET SCIENTIFIQUES QUI SE POSENT EN MATIÈRE D'INTELLIGENCE ARTIFICIELLE	191
A. LES SUJETS D'INTERROGATION LIÉS AUX ALGORITHMES UTILISÉS PAR LES TECHNOLOGIES D'INTELLIGENCE ARTIFICIELLE	191
1. Les questions de sécurité et de robustesse	191
2. Les biais et les problèmes posés par les données nécessaires aux algorithmes d'apprentissage automatique	192
3. Le phénomène de « boîtes noires » des algorithmes de deep learning appelle un effort de recherche fondamentale vers leur transparence	193
4. La question des bulles d'information dites « bulles de filtres »	195

B. LES SUJETS D'INTERROGATION LIÉS À LA « SINGULARITÉ », À LA « CONVERGENCE NBIC » ET AU « TRANSHUMANISME »	195
1. La « singularité », point de passage de l'IA faible à l'IA forte peut, à long terme, constituer un risque	195
2. Un prophétisme dystopique indémontrable scientifiquement	197
3. Les questions posées par la « convergence NBIC »	202
4. La tentation du « transhumanisme »	202
 TROISIÈME PARTIE : LES PROPOSITIONS DE VOS RAPPORTEURS	205
 I. POUR UNE INTELLIGENCE ARTIFICIELLE MAÎTRISÉE	205
<i>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</i>	205
<i>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</i>	206
<i>Proposition n° 3 : Former à l'éthique de l'intelligence artificielle et de la robotique dans certains cursus spécialisés de l'enseignement supérieur</i>	208
<i>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</i>	208
<i>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</i>	209
 II. POUR UNE INTELLIGENCE ARTIFICIELLE UTILE, AU SERVICE DE L'HOMME ET DES VALEURS HUMANISTES	210
<i>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</i>	210
<i>Proposition n° 7 : Encourager la constitution de champions européens en intelligence artificielle et en robotique, tout en poursuivant le soutien aux PME spécialisées, en particulier les start-up</i>	212
<i>Proposition n° 8 : Orienter les investissements dans la recherche en intelligence artificielle vers l'utilité sociale des découvertes</i>	213
<i>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</i>	213
<i>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</i>	215
<i>Proposition n° 11 : Assurer une meilleure prise en compte de la diversité et de la place des femmes dans la recherche en intelligence artificielle</i>	215
 III. POUR UNE INTELLIGENCE ARTIFICIELLE DÉMYSTIFIÉE	216
<i>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</i>	216
<i>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</i>	217
<i>Proposition n° 14 : Former et sensibiliser le grand public aux conséquences pratiques de l'intelligence artificielle et de la robotisation</i>	219

<i>Proposition n° 15 : Être vigilant sur les usages spectaculaires et alarmistes du concept d'intelligence artificielle et de représentation des robots</i>	219
CONCLUSION	221
SAISINE DE L'OFFICE	223
RÉUNION DE L'OPECST DU 14 MARS 2017 : ADOPTION DU RAPPORT	225
LISTE DES PERSONNES RENCONTRÉES	253
I. PERSONNES RENCONTRÉES PAR LES RAPPORTEURS EN VUE DE L'ÉTUDE DE FAISABILITÉ	253
II. PERSONNES RENCONTRÉES PAR LES RAPPORTEURS EN VUE DE L'ÉTABLISSEMENT DU RAPPORT	254
A. EN FRANCE	254
B. À L'ÉTRANGER	260
BIBLIOGRAPHIE	269

SYNTHÈSE DU RAPPORT

LES OBSERVATIONS

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^e 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 ce domaine 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 ? 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 induites par le déploiement des technologies d'intelligence artificielle**. Tout au long de l'année 2016, les initiatives en ce domaine 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 **démystifier les représentations biaisées** de ce concept et de **mettre en avant les opportunités et les risques qui lui sont liés**.

Ces représentations excessives, qui peuvent être totalement opposées, sont accentuées par la phase générale de progrès dans laquelle on se situe : 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 cette technologie invitent à **ne pas trop s'enthousiasmer et à faire preuve d'attentes réalistes**.

Les **applications sectorielles présentes ou futures, riches de la capacité prédictive de ces technologies, 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 sur le concept d'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 » et souligne 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'**imminence** d'une **superintelligence** ne fait pas partie de ces risques à court et moyen termes mais relève du fantasme. **À long terme, la réalité de cette menace n'est pas certaine.** 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.**

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 réaction au développement de cette technologie afin que **le débat public puisse s'engager sereinement.**

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 constatées 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.**

LES PROPOSITIONS

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, tout en poursuivant le soutien aux PME spécialisées, en particulier les start-up.

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

L'**intelligence artificielle n'est pas un simple terrain de jeu**, même si les victoires des systèmes *AlphaGo* au jeu de Go face au champion Lee Sedol en mars 2016 et *Libratus* au Poker face à quatre joueurs professionnels en janvier 2017, ou, auparavant celles de *Watson* au jeu télévisé *Jeopardy* en 2011 et de *Deep Blue* aux échecs face à Garry Kasparov en 1997, pourraient 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 suffisamment 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 induites par le déploiement des technologies d'intelligence artificielle.**

De plus, cette naissance du débat public sur le sujet de l'intelligence artificielle selon un mode alarmiste 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 aptes à sélectionner et combattre 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 l'attribution de la chaire d'informatique du Collège de France à Yann LeCun en février 2016¹ ou par l'événement 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. Cette victoire marque **l'histoire des progrès en intelligence artificielle et contredit 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.

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 plusieurs travaux sur lesquels le présent rapport va revenir plus loin : 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 rapports²), 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 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.

¹ 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 ».

² Le premier porte sur l'éthique du chercheur en robotique et l'autre sur l'éthique en apprentissage automatique.

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 partager la connaissance en vue 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* ».

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, souvent par leur capacité prédictive, 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 pouvoir 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 ainsi 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 ou 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 et partager la connaissance de l'état - à cet 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 robotique¹ et au-delà, parce qu'ils n'oublient pas que « *science sans conscience n'est que*

¹ 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 recherche publics concernés par le numérique, notamment par l'intelligence artificielle, comme le CNRS, le CEA, Inria, la CPU, la CDEFI et l'Institut Mines-Télécom. L'INRA, l'INRETS et l'ONERA en sont membres associés. Ce rapport est abordé au IV de la présente étude. Il peut être trouvé ici : <https://hal.inria.fr/ALLISTENE-CERNA/hal-01086579v1>

ruine de l'âme », ainsi que l'affirmait Rabelais. Ils jugent qu'il relève du devoir citoyen de diffuser et partager la connaissance scientifique et technologique.

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** ? Devant 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 un risque de futures désillusions**, il est nécessaire d'**opérer un bilan objectif** de l'état de l'art en matière scientifique et technologique s'agissant de « 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 **le point sur la recherche et les usages des technologies d'intelligence artificielle**, en souligner les enjeux multiples et les riches perspectives, car ils sont **animés d'une préoccupation pédagogique en vue 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 introduit par la loi n° 83-609 du 8 juillet 1983 qui le crée, par la commission des affaires économiques du Sénat, d'une demande d'étude sur l'intelligence artificielle. La victoire, dans le même temps, du système Alpha Go, sur le champion de Go, Lee Sedol, **confirmait, s'il en était besoin la pertinence de cette saisine**, tant cet événement marque l'ampleur des progrès en intelligence artificielle.

Vos rapporteurs étaient d'ores et déjà convaincus de l'intérêt de ce sujet, compte tenu de sa forte visibilité dans l'actualité et la vie publique.

Ainsi, votre rapporteure Dominique Gillot avait été sensibilisée aux enjeux de l'intelligence artificielle lors du colloque annuel de la conférence des présidents d'université (CPU) organisée en 2015 sur le 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. La conclusion de ce colloque évoquait même l'« Université 4.0 ».

En outre, plusieurs rapports de l'OPECST ou des tables rondes organisées par l'Office¹ avaient déjà traité partiellement ce sujet. Elle abordait donc cette problématique avec la volonté, comme d'ailleurs le rapporteur Claude de Ganay, d'éclairer et de faire partager une culture scientifique, technique et industrielle, en adéquation totale avec la démarche du CNCSTI² qu'elle préside.

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

¹ Ces travaux sont récapitulés un peu plus loin.

² Le Conseil national de la culture scientifique, technique et industrielle, placé auprès du ministre chargé de la culture et du ministre en charge 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 la recherche ».

contraints, **une étude de faisabilité a été réalisée puis présentée** le 28 juin 2016 **en application des** articles 19¹ et 20² du règlement intérieur de l'Office.

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.**

2. Un calendrier de travail très resserré

L'étude de faisabilité n'ayant pu être présentée à l'OPECST que fin juin 2016 et le calendrier des travaux parlementaires ayant été interrompu fin février 2017 en raison des échéances électorales (présidentielle et législatives puis sénatoriales), vos rapporteurs n'ont réellement pu disposer que d'un peu plus de six mois pour effectuer leurs investigations, ce qui est très court étant donné l'importance du sujet. Ce délai contraint leur interdisant d'élaborer un rapport faisant un point complet sur l'ensemble des questions posées par l'intelligence artificielle, ils ont précisé le champ des investigations de l'étude. Sans chercher à épuiser le sujet, ils ont voulu réaliser un travail exploratoire, travail qui devra être poursuivi, d'autant que la recherche, l'innovation et la communication scientifique et technologique ne cessent d'avancer.

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

L'étude de faisabilité a tout d'abord établi un **bilan des précédents travaux de l'Office** et des **autres travaux** sur l'intelligence artificielle conduits récemment **en dehors de son cadre.**

À travers cette revue des rapports de l'OPECST il s'agissait de voir si une étude analogue avait été conduite et de déterminer ce qu'il était pertinent d'analyser de manière plus spécifique concernant l'intelligence artificielle.

¹ « 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. »

² « 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. »

Ainsi, l'OPECST a, dernièrement, rendu en 2016 un **rapport sur les robots et la loi**, reprenant le compte rendu d'une audition publique organisée le 10 décembre 2015¹. Auparavant, et en allant du plus récent au plus ancien, l'Office a rendu plusieurs travaux sur le thème du numérique, démentant la thèse du livre de Laure Belot, « *La déconnexion des élites. Comment Internet déränge l'ordre établi ?* »².

Il a ainsi travaillé sur le numérique au service de la santé³, sur la sécurité numérique⁴, sur le *big data* dans l'agriculture⁵, sur les drones et la sécurité des installations nucléaires⁶, sur les nouveaux moyens de transports⁷, sur les nouvelles technologies d'exploration et de thérapie du cerveau⁸, et, précédemment, sur la gouvernance mondiale de l'Internet⁹, sur les conséquences de l'évolution scientifique et technique dans le secteur des télécommunications¹⁰, sur les enjeux de société posés par le monde virtuel¹¹, sur les techniques d'apprentissages en matière informatique¹², sur l'entrée dans la société de l'information¹³ et sur les conséquences de la révolution numérique¹⁴.

L'Office n'a cependant **jamais travaillé directement et globalement sur l'intelligence artificielle**, même si certains enjeux, tels que l'impact sur la santé, la protection des données ou les moyens de transport, ont pu être entrevus à l'occasion de précédentes études menées. Il semblait donc tout à

¹ *Les robots et la loi*, Jean-Yves Le Deaut et Bruno Sido, rapporteurs, Sénat, n° 570 (2015-2016).

² *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.*

³ *Le numérique au service de la santé*, Catherine Procaccia et Gérard Bapt, rapporteurs, Sénat n° 465 (2014-2015).

⁴ *Sécurité numérique et risques : enjeux et chances pour les entreprises*, Anne-Yvonne Le Dain et Bruno Sido, rapporteurs, Sénat n° 271 (2014-2015).

⁵ *La place du traitement massif des données (big data) dans l'agriculture : situation et perspectives*, Jean-Yves Le Deaut, Anne-Yvonne Le Dain et Bruno Sido, rapporteurs, Sénat n° 614 (2014-2015).

⁶ *Les drones et la sécurité des installations nucléaires*, Jean-Yves Le Deaut et Bruno Sido, rapporteurs, Sénat n° 267 (2014-2015).

⁷ *Les nouvelles mobilités sereines et durables*, Denis Baupin et Fabienne Keller, rapporteurs, Sénat n° 293 (2013-2014).

⁸ *Les enjeux des nouvelles technologies d'exploration et de traitement du cerveau*, Alain Claeys et Jean-Sébastien Vialatte, rapporteurs, Sénat n° 476 (2011-2012).

⁹ *La gouvernance mondiale de l'Internet*, Claude Birraux et Jean-Yves Le Deaut, rapporteurs, Sénat n° 219 (2005-2006).

¹⁰ *Les conséquences de l'évolution scientifique et technique dans le secteur des télécommunications*, Pierre Laffitte et René Tregouet, rapporteurs, Sénat n° 159 (2001-2002).

¹¹ *Images de synthèse et monde virtuel : techniques et enjeux de société*, Claude Huriet, rapporteur, Sénat n° 169 (1997-1998).

¹² *Les techniques des apprentissages essentiels pour une bonne insertion dans la société de l'information*, Franck Sérusclat, rapporteur, Sénat n° 383 (1996-1997).

¹³ *La France et la société de l'information : un cri d'alarme et une croisade nécessaire*, Pierre Laffitte, rapporteur, Sénat n° 213 (1996-1997).

¹⁴ *Les nouvelles techniques d'information et de communication : l'Homme cybernétique*, Franck Sérusclat, rapporteur, Sénat n° 232 (1994-1995).

fait nécessaire d’approfondir ce thème à travers une étude intégrée approfondie.

Les rapporteurs se sont ensuite interrogés sur l’existence d’une étude analogue concernant l’intelligence artificielle effectuée en dehors de l’OPECST. Pour répondre à cette question, différentes publications nationales et européennes ont été passées en revue, sans identifier **de rapport public faisant le point sur ces technologies, ni 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, la disponibilité de masses de données énormes et les usages de l’intelligence artificielle devaient de toute évidence être identifiés et évalués, **les pouvoirs publics paraissent ne pas avoir pris toute la mesure de l’enjeu, bien que cela fût souhaitable.**

Deux exceptions ont été relevées : une initiative émanant d’une députée européenne et le travail d’une association missionnée par la Commission européenne. En effet, au sein des institutions de l’Union européenne, Mme Mady Delvaux, présidente d’un 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**¹.

D’autres initiatives, moins en rapport avec l’intelligence artificielle, ont néanmoins été mentionnées dans l’étude de faisabilité. Au niveau national, le Parlement a débattu et adopté, en 2016, le projet de loi pour une République numérique². 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

¹ Cf. http://www.eu-robotics.net/cms/upload/PDF/euRobotics_Deliverable_D.3.2.1_Annex_Suggestion_GreenPaper_ELS_IssuesInRobotics.pdf

² 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.

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 sur le thème de l'intelligence artificielle. Les précisions apportées par cette brochure étaient cependant utiles pour éclairer la réflexion et les discussions sur le sujet. Enfin, il faut mentionner une **lettre de veille du pôle de compétitivité Cap Digital** sur l'intelligence artificielle en 2015¹ et le fait que l'**Agence nationale de la recherche** (ANR) ait consacré un de ses « cahiers »² au thème de l'intelligence artificielle et de la robotique en 2012, intitulé « *Intelligence Artificielle et Robotique : Confluences de l'Homme et des STIC* »³.

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 robotique⁴.

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, renvoyant ainsi au processus progressif dans lequel s'inscrit désormais la production de la connaissance et la diffusion des savoirs.

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, vos rapporteurs 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.

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

Vos rapporteurs **ont souhaité** préciser 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**

¹ Cf. http://www.craft.ai/images/posts/craft-ai-interviewed-by-cap-digital/craft_ai_article_cap_digital.pdf

² Cahier n° 4, mars 2012.

³ Cf. http://www.agence-nationale-recherche.fr/fileadmin/user_upload/documents/2012/Cahier-ANR-4-Intelligence-Artificielle.pdf

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

ce qu'est aujourd'hui, et ce que peut devenir demain l'intelligence artificielle.

Les enjeux de ce thème d'étude sont tout autant scientifiques et technologiques que politiques, sociétaux, économiques, philosophiques, éthiques¹, juridiques, éducatifs, médicaux, ou, encore, militaires.

Compte tenu des délais impartis, il leur a fallu ne retenir que **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, neufs 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 (pour l'aide au diagnostic, la thérapeutique, l'épidémiologie, la chirurgie...) ;
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,**

¹ L'éthique est une discipline philosophique qui porte sa réflexion sur les fondements de la morale. Elle envisage donc les normes, les limites et les devoirs de celle-ci. L'éthique se définit aussi comme un ensemble de principes moraux qui sont à la base de la conduite d'un sujet.

juridiques et éducatifs¹ de l'intelligence artificielle sont en effet **majeurs** et 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 **mis au second plan** dans la mesure où ce domaine correspond moins directement à la mission et à la plus-value apportée par l'OPECST.

Les usages de l'intelligence artificielle en matière médicale justifieraient **un rapport à part entière** et vos rapporteurs ont donc mis cette dimension de côté dans l'attente de prochains travaux. Ils évoquent toutefois ces usages dans le présent rapport mais de manière souvent illustrative.

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 souhaité s'engager dans cette voie.

Enfin, **le point 9 sur le transhumanisme est partiellement abordé, davantage sous l'aspect des enjeux éthiques de l'intelligence artificielle**. Il n'y avait pas lieu de traiter spécifiquement du transhumanisme, mouvement controversé, bien éloigné de l'intérêt général recherché par vos rapporteurs : il relève en effet davantage du projet idéologique que de la réalité scientifique et couvre de plus un champ différent de 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**, 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.

¹ 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 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.

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 **ne pas céder à 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 gêner et de ralentir l'innovation. Vos rapporteurs, au terme de leurs investigations, ne sont toujours **pas convaincus de l'urgence d'une intervention législative ou réglementaire en matière d'intelligence artificielle, en raison de son caractère 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 Paris IV Sorbonne et directrice de recherche au Laboratoire d'informatique pour la mécanique et les sciences de l'ingénieur (LIMSI-CNRS, campus Paris-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**. C'est d'ailleurs le thème du travail¹ conduit en 2016 et 2017 par l'association mondiale des ingénieurs électriciens et électroniciens (Institute of Electrical and Electronics Engineers ou IEEE).

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

À la suite de **l'adoption de l'étude de faisabilité** le 28 juin 2016, vos rapporteurs ont procédé à des auditions bilatérales et effectué des déplacements.

Ce point est développé en annexe du présent rapport, avec une présentation des auditions bilatérales et des déplacements de vos

¹ Cf. le site http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html

rapporteurs en France et à l'étranger (États-Unis, Royaume-Uni, Suisse, Belgique).

Il s'agissait pour mémoire des déplacements suivants :

- 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 d'auditions publiques au Sénat le 19 janvier 2017

Une journée d'auditions publiques a été conduite au Sénat le 19 janvier 2017¹. 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* »².

Vos rapporteurs relèvent que ces auditions ont constitué un **record d'audience** pour le Sénat³. La chaîne Public Sénat en a tiré une émission spéciale, diffusée le lendemain de la journée d'auditions⁴.

¹ Pour les auditions du 19 janvier 2017 matin, le lien est le suivant : http://videos.senat.fr/video.302142_588070cd0faea.reunion-pleniere-opecst et pour celles de l'après-midi du 19 janvier 2017 : http://videos.senat.fr/video.303157_5880bb6877f24.reunion-pleniere-opecst

² https://www.sciencesetavenir.fr/high-tech/intelligence-artificielle/les-pouvoirs-publics-s-emparent-de-la-question-de-l-intelligence-artificielle_109899

³ Selon les données disponibles à la fin du mois de janvier 2017 : 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.

⁴ <https://www.publicsenat.fr/emission/les-matins-du-senat/table-ronde-intelligence-artificielle-51796>

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*¹, 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 consultés².

La troisième édition du **manuel de Stuart Russell et Peter Norvig** « *Artificial Intelligence: A Modern Approach* » reste inégalée³. Vos rapporteurs ont rencontré l'un de ses deux auteurs, Stuart Russell, professeur à Berkeley, très engagé dans le débat public sur l'intelligence artificielle. Une traduction française de cet ouvrage est disponible, avec le concours de Laurent Miclet et Fabrice Popineau.

Écrit par deux experts de renommée mondiale, cet ouvrage constitue 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.

¹ 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.

² 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 ».

³ 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.

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.

Des **ouvrages scientifiques et/ou techniques** doivent aussi 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 : mythes, fantasmes et réalité*, **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*, **Nathalie Nevejans**, *Traité de droit et d'éthique de la robotique civile*, **Jean-Yves Girard** et **Alan Turing**, *La Machine de Turing*, **Nick Bostrom**, *Super-intelligence*, Yoshua Bengio, Aaron Courville et Ian Goodfellow, *Deep Learning*¹ **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 éclairant² qui revient sur l'histoire de l'intelligence artificielle.

¹ Ouvrage disponible en ligne : <http://www.deeplearningbook.org/>

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

Le **livre blanc d'Inria sur l'intelligence artificielle**, coordonné par Bertrand Braunschweig, a été consulté¹, ainsi que les rapports d'activités et les contrats d'objectifs de cet institut. 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 largement utilisée.

Vos rapporteurs 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 à travers différents ouvrages : **Joël Boissière**, **Simon Fau** et **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** et **Bernard Stiegler**, *L'école, le numérique et la société qui vient*, **Franck Amadiou** et **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*, **Éric 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 ?*

¹ Cf. https://www.inria.fr/content/download/103897/1529370/./AI_livre-blanc_n01.pdf

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

A. DES TECHNOLOGIES NÉES AU MILIEU DU XX^E 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 les ouvrages cités, notamment la brochure de l'Académie des technologies de « dix questions sur l'intelligence artificielle et la technologie ».

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 entiers¹. Bruce Buchanan a, quant à lui, rédigé un article sur l'histoire de l'intelligence artificielle qui revient également sur l'ensemble de ses précurseurs². Comme le relève Jean-Gabriel Ganascia, que vos rapporteurs ont pu rencontrer à plusieurs reprises, **Homère** a décrit, dans « *L'Iliade* », 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, véritables ancêtres des automates. **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^e siècle, de reproduire artificiellement les fonctions biologiques, y compris la communication et la locomotion. **Blaise**

¹ 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*.

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

Pascal réfléchit à la création d'une machine à calculer. À la fin du XVII^e siècle, **Leibniz** imagine ensuite une machine à calculer capable de raisonner. Il construit un prototype de machine à calculer en 1694.

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^e siècle, **George Boole** appelle à mathématiser la logique, **Charles Babbage** conçoit l'ancêtre mécanique des ordinateurs d'aujourd'hui¹ 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**. Laurence Devillers souligne cette réalité incontestable dans son livre *Des robots et des hommes : mythes, fantasmes et réalité*.

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.

¹ 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.

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

2. Les premières étapes de formation des technologies d'intelligence artificielle au XX^e 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, visitée 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**¹. De nombreuses applications sont possibles, à commencer par l'informatique, le fonctionnement des

¹ 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 »).

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**, 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), 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 simuler et prévoir** (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 1950¹, 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 critère permettant de qualifier une machine de « consciente »². Il fait alors le pari que « *d'ici à 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 pas encore 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* »³ dans lequel ils posent l'hypothèse que les neurones avec leurs deux états, activé ou non activé, pourraient

¹ « Computing Machinery and Intelligence », *Mind*, octobre 1950.

² 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.

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

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 artificiels¹, à partir des travaux de McCulloch et Pitts. Ces derniers publient un **article plus important que les autres en 1959**² 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 essentialiste³ 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 donc dès 1956 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

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

² « 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*.

³ 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.

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**, 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 programmation¹, 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.**

En 1958, Herbert Simon et Allen Newell déclarent ainsi que « *d'ici à dix ans un ordinateur sera le champion du monde des échecs* »² et « *d'ici à dix ans, un ordinateur découvrira et résoudra un nouveau théorème mathématique majeur* ».

En 1965, Herbert Simon assure que « *des machines seront capables, d'ici à 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* »

De même, le **premier agent conversationnel** (« chatbot » ou « bot ») est créé en 1966 par Joseph Weizenbaum et simule un psychologue grâce à sa technologie de reconnaissance des 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.

¹ 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.

² Il faudra attendre 1997 pour que le champion d'échecs Garry Kasparov s'incline devant le système Deep Blue d'IBM.

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/ashir233/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 technologies 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 connaissances¹ 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

¹ 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).

Les connaissances n'étant pas universelles, nous pouvons être conduit à des hypothèses et 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 j'ignore 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 1958¹, 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

¹ « 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.

compréhension du langage naturel, dans la vision artificielle, mais en fin de compte les résultats sont décevants, largement en raison des limitations de puissance du matériel disponible, des données à utiliser mais aussi des limites intrinsèques des technologies alors disponibles.

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'époque¹.

Après cet **âge d'or, qui court de 1956 au début des années 1970**, les **financements sont revus à la baisse** en raison de **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 sont alors en plein essor, en lien avec les sciences cognitives, 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

¹ 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, en réaction à la déception, une période noire d'une vingtaine d'années pour les réseaux de neurones artificiels.

reproduire les raisonnements d'un expert, dans un domaine particulier. La connaissance est décrite sous la forme générale de règles :

« 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 réalisées 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-connaissances¹ 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 interopérer 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

¹ Il s'agit de connaissances à propos des connaissances elles-mêmes.

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 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 scénarios 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.

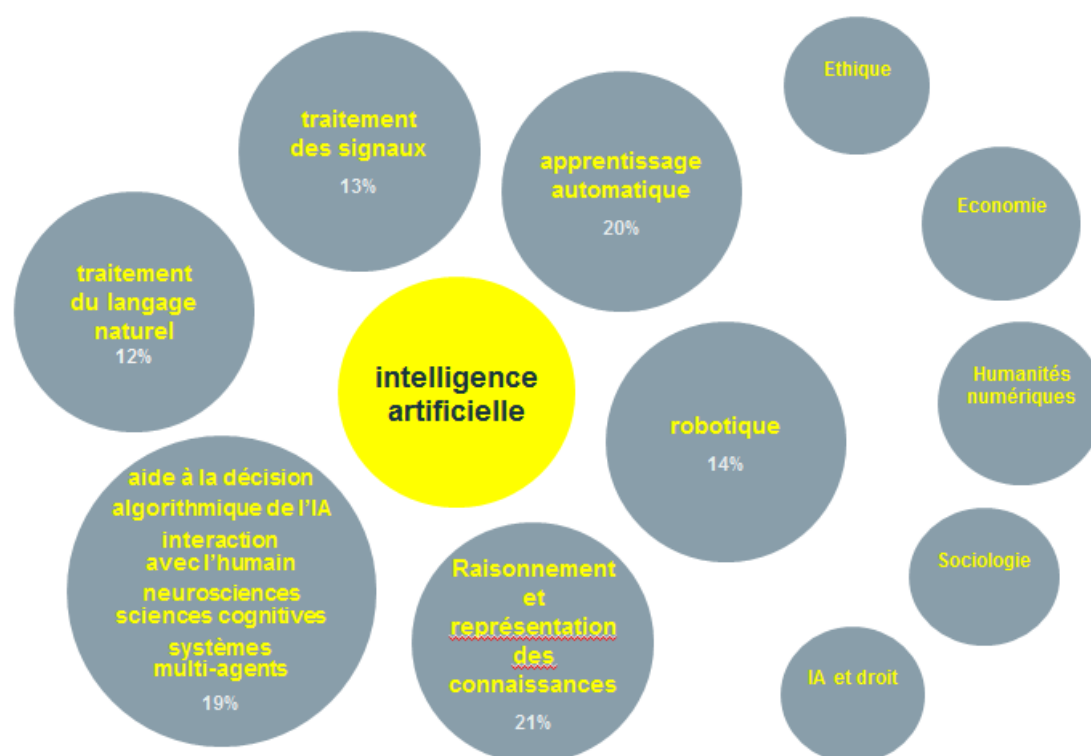
5. Les autres domaines et technologies d'intelligence artificielle : robotique, systèmes multi-agents, machines à vecteur de support (SVM), réseaux bayésiens, apprentissage machine dont apprentissage par renforcement, programmation par contraintes, raisonnements à partir de cas, ontologies, logiques 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 de manière approximative des chercheurs français entre les différents domaines de l'intelligence artificielle)

L'un des premiers domaines qui peut être pris pour exemple est celui de la **robotique**, qui a toujours entretenu des liens très étroits avec celui de l'intelligence artificielle sans se confondre avec elle, un champ de la robotique étant hors de l'IA (les simples automates, par exemple). 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

Turing¹ 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 ».

Cette manière de poser la question néglige un constat pourtant simple : au fil des ères (de la reptation à la bipédie, de la cueillette à l'agriculture, de la chasse à l'élevage par exemple), le **cerveau humain a évolué** vers ce qu'il est grâce au **développement des capacités de perception, d'interprétation, d'apprentissage et de communication** en vue d'une action plus efficace, de plus en plus déterminée par la volonté et des choix rationnels. 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 la 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 « encorporée », 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 et notamment de certaines familles d'insectes, permettant la mise en œuvre de systèmes qui s'auto-organisent. Ces comportements sociaux peuvent être

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

programmés de manière plus ou moins complexe et intégrer des croyances, des désirs et des intentions.

Vos rapporteurs ont expérimenté ces systèmes 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 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 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 coordination est un aspect important, mais des systèmes multi-agents peuvent mettre l'accent sur d'autres caractéristiques, à l'instar du modèle Voyelles développé par Yves Demazeau en 1995 et qui insiste sur les relations entre les agents, l'environnement, les interactions et l'organisation.

L'**évolution artificielle** et la **programmation génétique**¹ donnent 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 pour cartographier des espaces de paramètres et surtout répondre 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 statistiques², donnent de bons résultats parmi les 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 orienté³. Judea Pearl, prix Turing⁴ 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

¹ Il s'agit d'une technique de programmation inspirée des mécanismes d'évolution et de sélection génétique des organismes vivants.

² 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 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.

³ 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.

⁴ Ce prix est la plus haute distinction en informatique.

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 de certains raisonnements humains, et les **raisonnements à partir de cas**, qui se fondent sur la notion d'analogie mais tend à devenir plus marginal.

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, 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 patentes 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 ou privilégier 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, devant 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.

Pour comprendre l'**apprentissage automatique** ou *machine learning*, il est loisible d'introduire, comme le préconise Jean-Claude Heudin, les concepts de « **prédicteur** » et de « **classifieur** ». Ce dernier est une machine capable d'approximer un processus¹ dont on ne connaît pas *a priori* le modèle. Le calcul de la sortie en fonction des données présentées en entrée est appelé une régression. Un classifieur est une machine qui va plus loin qu'un simple prédicteur, c'est aussi une fonction linéaire, mais elle permet de différencier des éléments appartenant à des catégories différentes lorsque ces catégories sont clairement séparables.

Le mot algorithme est issu, comme il a été vu, de la latinisation du nom du mathématicien Al-Khawarizmi 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 au programme d'apprendre et d'améliorer le système d'analyse et/ou de réponse. En ce sens, on peut dire que **ces types particuliers d'algorithmes apprennent.**

Un apprentissage est dit « **supervisé** » lorsque **l'algorithme définit des règles à partir d'exemples qui sont autant de cas validés.** Ces exemples sont appelés bases de données d'apprentissage. Jean-Claude Heudin parle ainsi de méthode itérative, autrement dit un algorithme, visant à ajuster les paramètres d'un classifieur linéaire en fonction de l'erreur qu'il commet entre la sortie souhaitée et la sortie obtenue grâce aux exemples que sont les données d'apprentissage. La correction à apporter à chaque étape au paramètre du classifieur peut être calculée simplement en faisant le rapport de l'erreur sur la valeur d'entrée. Une meilleure performance est obtenue en modérant les ajustements par un taux d'apprentissage. Le classifieur peut aussi être non linéaire en vue de séparer des données qui ne sont pas

¹ Le comportement du processus est approximé grâce à un modèle qui comprend un ensemble de paramètres ajustables. Une bonne approche pour ajuster les paramètres est de les modifier progressivement de façon à minimiser l'erreur que le prédicteur produit lorsqu'on lui présente des données dont on connaît la sortie correspondante.

elles-mêmes régies par un processus linéaire¹. Les réseaux de neurones artificiels, analysés plus loin, représentent une **multiplication de classifieurs non linéaires dotés de paramètres d'ajustement** (il s'agit de coefficients qui sont appelés de manière métaphorique les « poids synaptiques »).

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

Entre ces deux formes d'apprentissage, 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 loin.

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**², les algorithmes d'apprentissage supervisés peuvent calculer automatiquement à partir de bases de données, en sélectionnant automatiquement les variables discriminantes à partir de données peu structurées et potentiellement volumineuses. Ils peuvent ainsi permettre **d'extraire des règles logiques de cause à effet** (des déterminismes) ou, au moins, **des corrélations, qui n'apparaissent pas immédiatement dans les données brutes**.

¹ L'exemple le plus connu est celui des opérateurs booléens (du nom du mathématicien George Boole) « Ou - Exclusif » (en anglais « X-OR »). Un classifieur linéaire peut traiter des opérations « Et » ou « Ou » mais bute sur le « Ou - Exclusif ». Les opérateurs booléens recourent à la solution « Non - Et » (en anglais « Nand »).

² Un arbre de décision est un outil d'aide à la décision représentant un ensemble de choix sous la forme graphique d'un arbre.

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 discrimination et de régression**. Il s'agit de classifieurs linéaires¹ 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**² et la **méthode du recuit simulé**³, 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 et d'une fonction de transfert. **Si cette somme dépasse un certain seuil, la sortie prend une valeur positive, sinon elle reste nulle**. Ainsi que l'explique Jean-Claude Heudin, ces réseaux sont des « *automates à seuil qui réalisent la somme pondérée de leurs entrées. Les coefficients synaptiques et le seuil d'activation permettent d'ajuster leur comportement* ». Le neurone formel peut être amélioré en utilisant des valeurs numériques au lieu d'un comportement binaire. Pour ce faire, la fonction à seuil est remplacée par une fonction sigmoïde⁴, les calculs restent néanmoins

¹ 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.

² 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.

³ 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 théorisée en 1985 par S. Kirkpatrick, C.D. Gelatt et M.P. Vecchi, et indépendamment par V. Černý.

⁴ La fonction sigmoïde a globalement la même forme que la fonction à seuil, mais les changements de valeur entre 0 et 1 sont plus progressifs : la courbe est plus douce et moins abrupte. L'équation de la fonction sigmoïde est la suivante : $y = 1 / (1 + e^{-x})$.

élémentaires. Un réseau de neurones artificiels comporte une **couche d'entrée** (les données), une **couche de sortie** (les résultats), et peut comporter **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, qui sont donc des coefficients capables de s'auto-ajuster, **afin que les correspondances entre les entrées et les sorties soient les meilleures possible.**

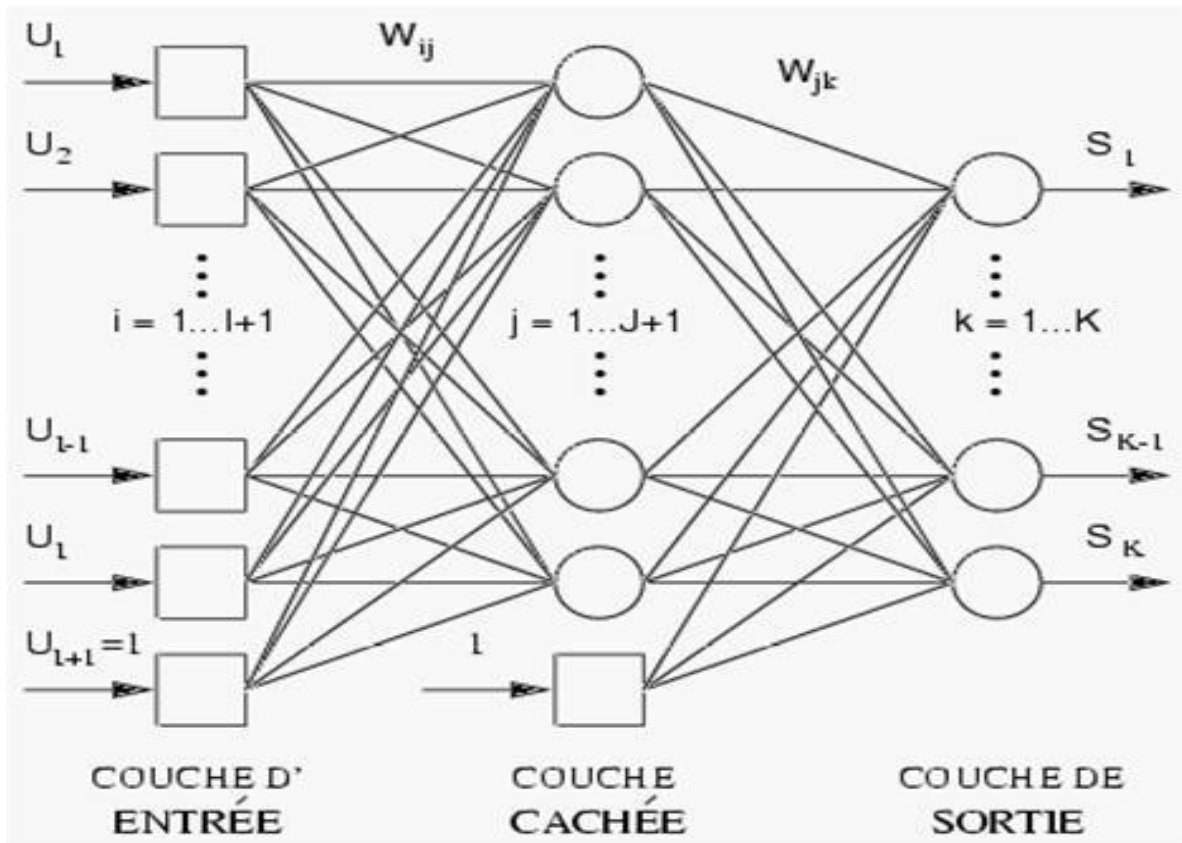
Après un nombre statistiquement pertinent d'exemples, l'apprentissage (implicite) est terminé et le réseau peut être utilisé, dans une seconde phase, pour la reconnaissance. Comme il produit toujours une sortie, même pour des entrées non rencontrées auparavant, **il a le plus souvent une bonne capacité de généralisation, qui dépend du corpus d'apprentissage.**

Il s'agit donc de **combiner 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 image. Cette métaphore biologique est sans doute malheureuse selon vos rapporteurs, car elle entretient une forme de **confusion**, en lien avec celle produite également par la notion d'**intelligence artificielle.**

Schéma d'un réseau de neurones artificiels



Source : Académie des technologies

Les applications nécessitent des structures aux couches de plus en plus conséquentes, malheureusement comme l'explique Jean-Claude Heudin, plus le nombre de couches augmente et plus les problèmes de surajustement (*overfitting*) et de disparition des gradients (*vanishing gradient*) deviennent gênants, sans même parler des temps de calcul qui explosent.

D'autres technologies peuvent encore 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 poker¹ de 120 000 mains successives sur 20 jours, intitulée « *Cerveau contre Intelligence Artificielle : on monte la mise* » (« *Brains Vs. Artificial Intelligence : Upping the*

¹ 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.

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 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 »¹ (Counterfactual Regret Minimization ou CFR), avec la méthode d'Oskari Tammelin introduite en 2014, permettant l'optimisation des résultats dans un contexte d'informations imparfaites². Un article collectif paru dans la revue *Science* en 2015 présentait déjà les évolutions théoriques nécessaires à cette victoire³.

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 cette opposition, 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.

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

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

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

Pour la DARPA, dont vos rapporteurs ont rencontré les responsables, ces deux grands types d'approches peuvent se voir donner le nom de « savoirs artisanaux » (ou faits à la main, en anglais *handcrafted knowledges*) pour la première et d'apprentissages statistiques (*statistical learning*) pour la deuxième, sachant qu'une troisième génération d'approche de l'intelligence artificielle devrait bientôt advenir : celle de **l'adaptation contextuelle** (*contextual adaptation*).

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

Les technologies d'apprentissage profond ou « *deep learning* » rencontrent un **succès particulièrement remarquable dans les années 2010**, pourtant elles sont anciennes. Leur essor doit beaucoup à l'émergence récente de la disponibilité de données massives (« *big data* ») et à l'accélération de la vitesse de calcul des processeurs, mais leur histoire remonte aux années 1940, ce que vos rapporteurs ont évoqué précédemment 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**¹ 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**². Marvin Minsky ayant pointé, comme il a été, vu les défauts de ce système, des **perceptrons multicouches** 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 multicouches, 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

¹ 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.

² Cf. « The Perceptron : A Probabilistic Model for Information Storage and Organization in the Brain ».

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*¹. 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 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 particulièrement pertinent ici (le cas échéant en complément de l'apprentissage profond) car il est un apprentissage souvent non supervisé qui permet au robot de découvrir à la fois les effets de ses actions, caractérisés par une

¹ <http://www.nature.com/nature/journal/v323/n6088/abs/323533a0.html>

« 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.

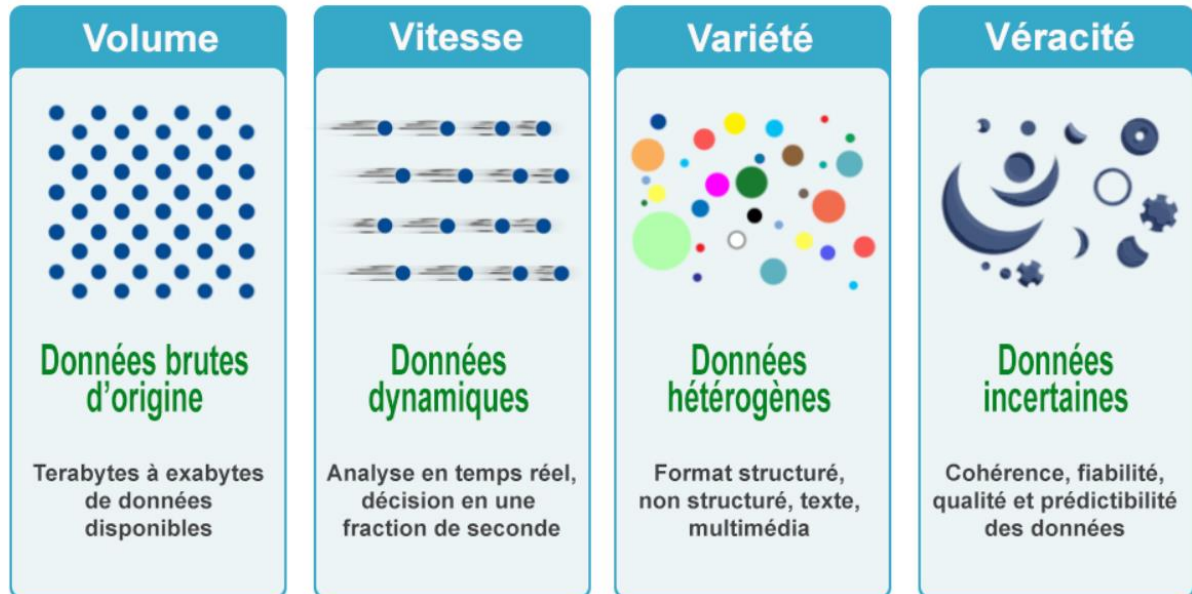
L'*affordance*, courante en ergonomie et en *design*, évoque la **potentialité ou la capacité d'un système à suggérer sa propre utilisation**, sans qu'il soit nécessaire de lui fournir un mode d'emploi. Issue de la psychologie¹, la notion renvoie d'abord à toutes les possibilités d'actions sur un objet puis a évolué vers les seules possibilités dont l'acteur est conscient. Cette utilisation intuitive est importante en robotique.

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

En *deep learning*, toute chose étant égale par ailleurs, **les algorithmes donnent des résultats d'autant plus performants que les données sont massives, variées, rapides et pertinentes** : ce sont les quatre V du *big data*, à savoir un **volume** croissant de données, issues d'une large **variété** de sources, qui circulent à une **vitesse** élevée et dont la **véracité** assure la cohérence.

¹ Elle apparaît pour la première fois en 1977 dans l'ouvrage *The Theory of Affordances* puis en 1979 dans *Approche écologique de la perception visuelle*, écrits par le psychologue James J. Gibson. Il s'agit, selon lui, de combinaisons invariantes de variables qui dépendent du contexte de l'action : les affordances ne sont donc pas des propriétés à part entière de l'objet dans ses travaux. La perception dépendant de la culture, de l'expérience et de l'apprentissage, une affordance dissimulée peut devenir perceptible par l'apprentissage. En outre, les normes peuvent contribuer à la perception des affordances. Le sens du concept a évolué et s'est élargi aux capacités d'un objet ou d'un être à suggérer sa propre utilisation.

Les « Quatre V » du *big data*



Source : Thierry Lombry, <http://www.astrosurf.com/luxorion/big-data-mining.htm>

S'agissant du volume de données, il a eu pour condition la **capacité croissante en matière de stockage de ces dernières**. Les supports de stockage des données ont rapidement évolué ces dernières décennies et sont de plus en plus divers.

Supports de stockage des données

Mémoire semi-conducteur (DRAM, SRAM, Flash)



Caractéristiques majeures :

- Contenance totale
- Densité d'info
- Vitesse d'accès
- Coût
- Pérennité/ Volatilité
- Mobilité

Papier



Bande magnétique



Disque dur magnétique



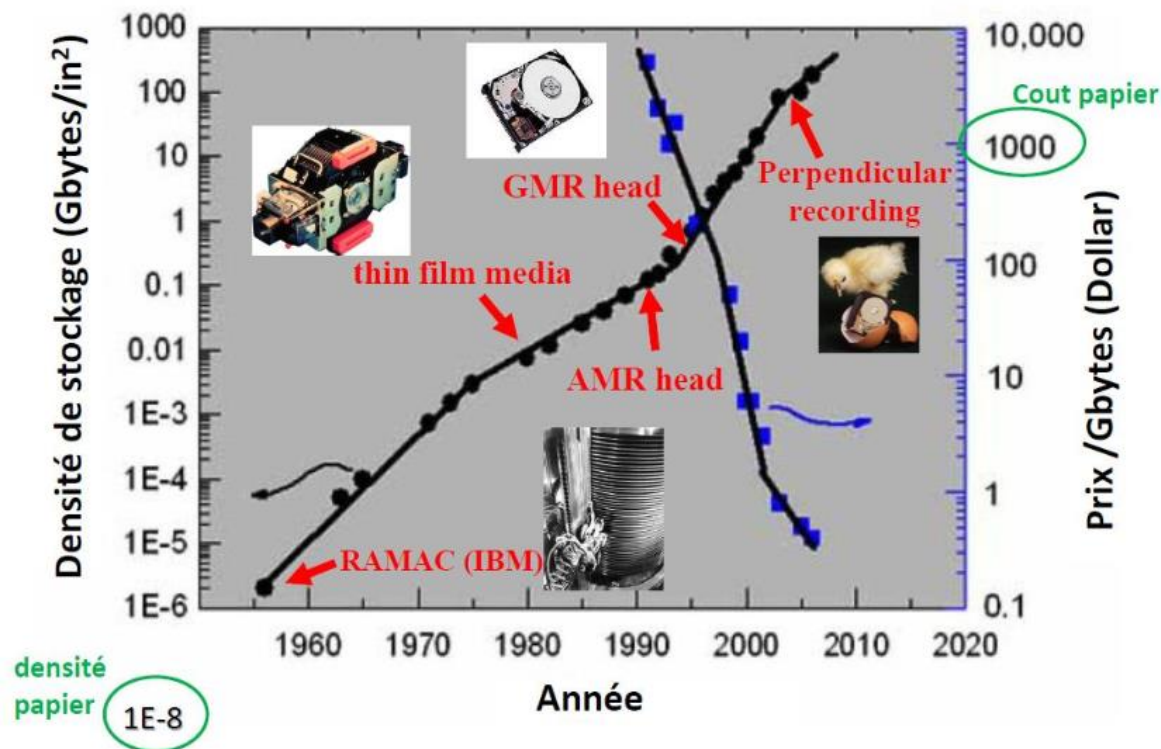
Disques gravés



Source : Thomas Hauet, maître de conférences à l'Université de Lorraine

Les disques durs des ordinateurs ont vu leur **capacité de stockage** suivre une croissance exponentielle, dans le même temps où le **coût du stockage** des données a, quant à lui, chuté très rapidement, surtout au cours des 25 dernières années. Le graphique suivant illustre la conjugaison de ces deux processus.

Progrès technologique dans la capacité de stockage des données



Source : Thomas Hauet, maître de conférences à l'Université de Lorraine

Selon Thomas Hauet, maître de conférences à l'Université de Lorraine, le stockage des données a énormément évolué depuis le premier disque dur dans les années 1950 (le RAMAC d'IBM en 1956 était une boîte volumineuse de cinquante disques qui faisaient jusqu'à 60 centimètres de large). On n'avait alors que seulement **5 mégabytes d'information** avec un coût considérable du gigabyte (10⁹ informations) de l'ordre de **10 millions de dollars**. En 1981, l'objet disque dur fait 30 à 40 centimètres de large avec des coûts moindres (100 000 dollars du gigabyte), des vitesses un peu plus grandes d'accès à l'information (20 Mbit/s). En 1994, la structure du disque rejoint celle que nous connaissons aujourd'hui, avec un coût de 100 dollars par gigabyte. En 2015, les disques durs permettaient de stocker **4 000 gigabytes** d'information avec des densités nettement plus grandes (500 Gbit/inch carré) et des coûts de **0,04 dollar le gigabyte**. Nous sommes donc passés en 50 ans d'un coût de 10 millions de dollars à 0,04 dollar le gigabyte.

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, mais la

constitution de **grandes bases de données labellisées** a souvent constitué un préalable¹.

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 utilisées pour l'apprentissage : 10 classes de données pour CIFAR-10 ou 100 classes de données pour CIFAR-100.

En 2010, le projet **ImageNet** a été lancé aux États-Unis² 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**.

Dans un article de 2014, Geoffrey Hinton indiquait que la base de données d'apprentissage de Google comprenait alors environ 100 millions d'images annotées et plus de 18 000 classes. En d'autres termes, selon Jean-Claude Heudin prenant l'exemple de la reconnaissance d'un chat, « *il vaut mieux avoir des images avec le chat dans tous ses états, plutôt que d'appliquer des prétraitements pour repositionner le chat dans une position idéale* », de même pour améliorer les performances d'une application, il vaut donc mieux augmenter le volume des données d'apprentissage que chercher à tout prix un meilleur algorithme, d'où l'aphorisme : « *ce n'est pas forcément celui qui a meilleur algorithme qui gagne, c'est celui qui a le plus de données* ».

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 maintenant 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*

¹ Une liste de ces nombreuses bases de données peut être consultée ici : https://en.wikipedia.org/wiki/List_of_datasets_for_machine_learning_research

² 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/>

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 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, vélocité et véracité sont bien les quatre 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 émergé, 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. Sans disposer d'une démonstration générale, il est cependant possible, selon Bertrand Braunschweig, de comprendre pourquoi ces technologies sont mathématiquement efficaces, et ce grâce à des approximateurs universels parcimonieux ou par la théorie de la dimension de Vapnik et Chervonenkis (dite « VC-dimension »).

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 **totalemt 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 *Bidirectionnal 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 Markov¹ 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.

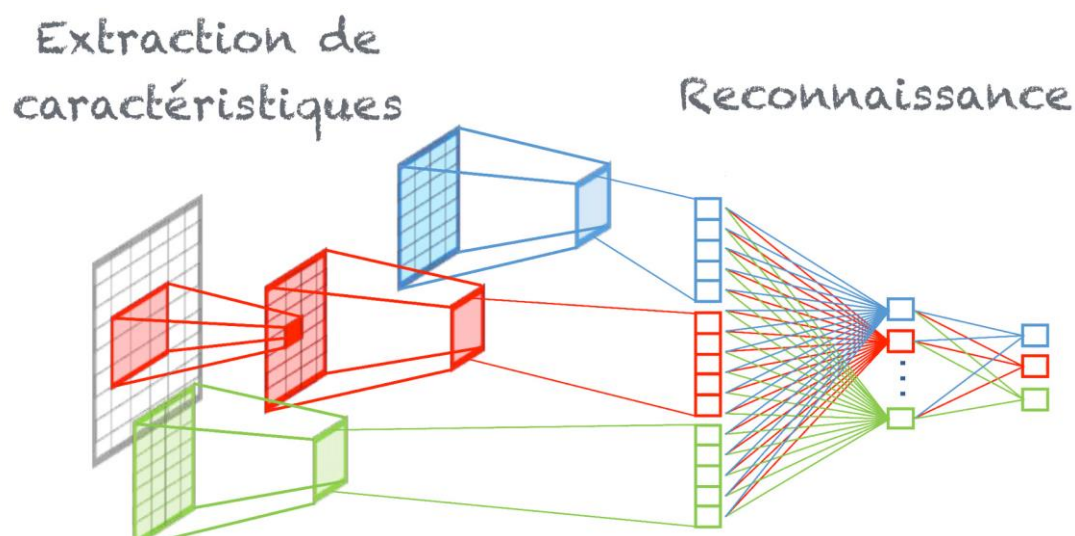
Les **réseaux neuronaux à convolution** (RNC) appelés aussi réseaux de neurones profonds convolutifs (*convolutional deep neural networks* ou CNN) sont **inspirés des processus biologiques du cortex visuel des animaux**. En effet, les neurones de cette région du cerveau sont arrangés de sorte qu'ils correspondent à des régions qui se chevauchent lors du pavage² du champ visuel et transforment un problème global à résoudre en une succession d'étapes plus petites et plus faciles à résoudre : 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'architecture des RNC comprend des couches de traitement indépendantes dédiées qui vont apprendre les prétraitements de

¹ 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.

² « Paver » l'image de départ revient à la découper en petites zones appelées tuiles. En informatique, chaque tuile est traitée individuellement par un neurone artificiel, qui effectue une opération de filtrage classique en associant un poids à chaque pixel de la tuile. Tous les neurones ont donc les mêmes paramètres, ce qui permet d'obtenir le même traitement pour tous les pixels de la tuile.

l'image au lieu de les coder¹, afin d'extraire ses caractéristiques et de les transmettre à un réseau neuronal plus classique qui effectue la phase de reconnaissance finale, comme l'illustre le graphique suivant.

Architecture d'un réseau de neurones à convolution

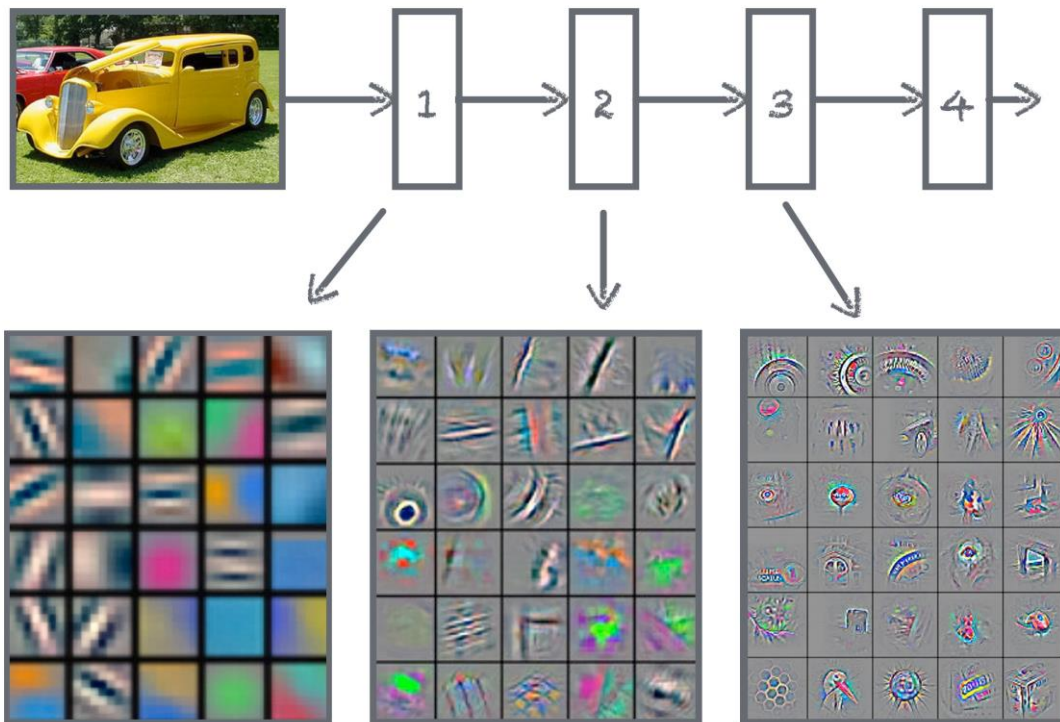


Source : Jean-Claude Heudin, *Comprendre le deep learning*.

Une autre façon de décrire l'approche par RNC est de la voir comme une **décomposition hiérarchisée du processus de reconnaissance**, où chaque couche participe à la création de représentations de plus en plus abstraites et conceptuelles.

¹ Le fait de coder les pixels est une approche plus traditionnelle de la reconnaissance d'images, dans laquelle l'ensemble des pixels d'une image (exemple 262 544 entrées pour une images de 512x512 pixels) est codé et vectorisé en une couche d'entrée du réseau de neurones.

Visualisation des caractéristiques extraites par chaque couche d'un réseau de neurones à convolution



Source : Jean-Claude Heudin, *Comprendre le deep learning*.

En outre, il convient de relever qu'au mois de janvier 2017, plusieurs chercheurs travaillant pour le projet de recherche de *deep learning* « Google Brain » ont publié un article¹ présentant les résultats de leurs travaux sur un **nouveau modèle de réseau de neurones multicouches** (appelé « MoE » pour « *Mixture of Experts Layer* »). La capacité d'un réseau neuronal monocouche à absorber les données massives étant limitée par son nombre de paramètres calculables, le modèle « MoE » représente un réseau neuronal géant, composé de plusieurs sous-réseaux neuronaux disposés en couches, permettant de traiter les quantités massives de données dont disposent les grandes entreprises du secteur. Accumuler les capacités de plusieurs systèmes experts à travers des réseaux neuronaux, permet de muscler la mémoire du modèle, réduisant ainsi le temps de formation du système et améliorant de manière considérable sa performance et sa capacité de calcul conventionnel, avec une architecture de modèle comprenant jusqu'à **137 milliards de paramètres**. Si, à ce stade, le modèle « MoE » est appliqué aux tâches de modélisation des langues et de traduction automatique, cette avancée présentée permet d'entrevoir des progrès exponentiels en matière d'intelligence artificielle, et selon les chercheurs du projet Google Brain le

¹ Geoffrey Hinton, Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc Le et Jeff Dean, « Outrageously large neural networks : the sparsely-gated mixture-of-experts layer », janvier 2017, cf. <https://arxiv.org/abs/1701.06538>

possible avènement d'une intelligence artificielle générale composée de milliers de sous-réseaux et traitant toutes sortes de données. Il s'agit aussi de réduire le nombre de processeurs (GPU) nécessaires à l'apprentissage et donc d'accélérer la capacité du système d'intelligence artificielle à processeur égal¹.

En mars 2017, la publication d'un autre article² de James Kirkpatrick basé sur l'apprentissage de plusieurs jeux Atari par un système d'IA trace la voie de futurs progrès dans la **capacité des réseaux de neurones artificiels à se souvenir de leurs tâches et de leurs expériences précédentes** et, donc, à se doter progressivement des éléments constitutifs d'une mémoire. En avril 2017, en utilisant aussi plusieurs jeux Atari, des chercheurs de Google-Deep Mind ont conceptualisé **l'apprentissage profond par renforcement**³ qui pourrait être selon eux la méthode d'apprentissage **la plus rapide**, avec l'idée de pouvoir transposer le processus d'apprentissage du système d'IA dans **l'environnement réel**. Les équipes de cette entreprise sont donc largement mobilisées pour construire des systèmes capables d'approcher l'apprentissage humain⁴.

L'essor de l'intelligence artificielle avec le *deep learning* est, par ailleurs, 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 empirique, ainsi que l'illustre le graphique suivant, démontre que **ce doublement a en fait lieu tous les dix-huit mois**.

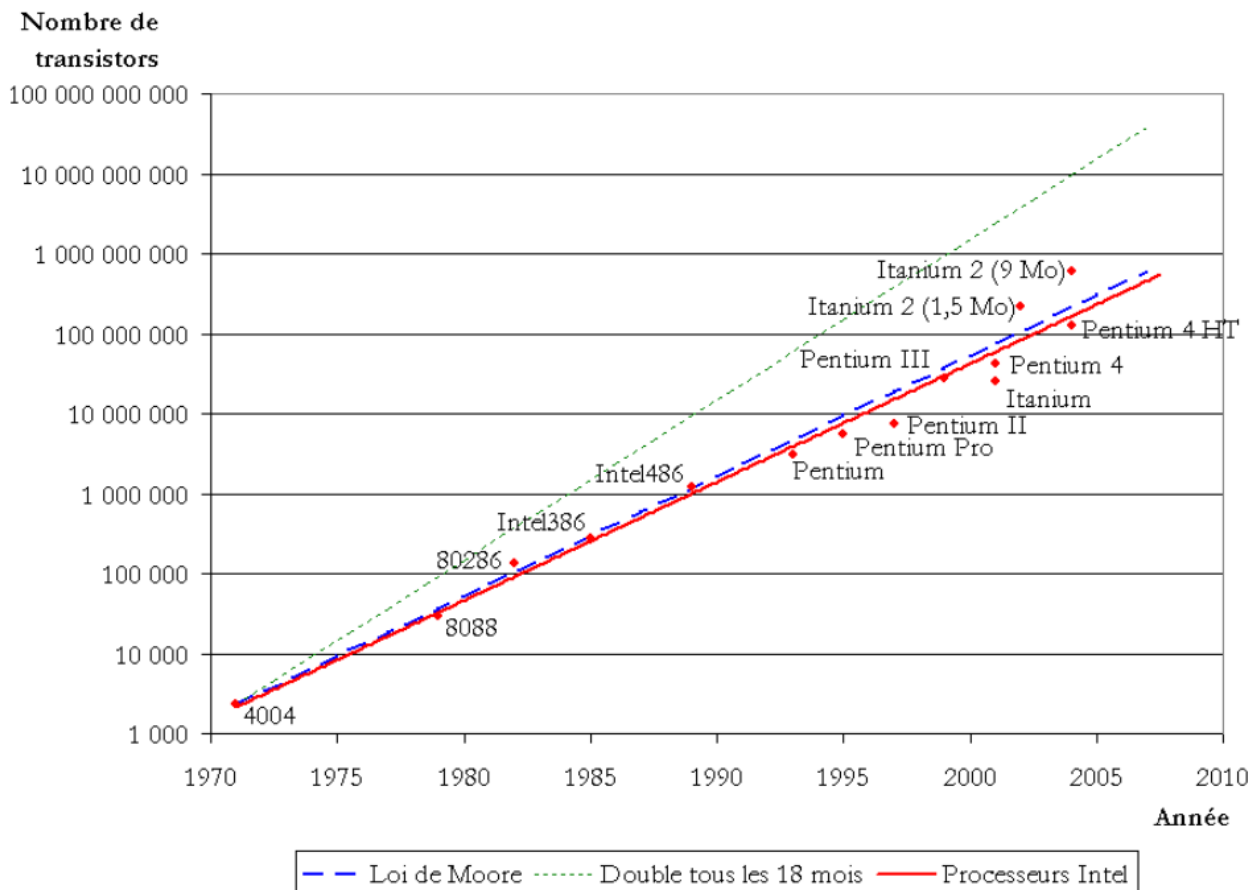
¹ Cf les explications données à ce sujet par le jeune chercheur Théo Szymkowiak, président de la société pour l'IA de l'université Mc Gill : <https://medium.com/@thoszymkowiak/google-brains-new-super-fast-and-highly-accurate-ai-the-mixture-of-experts-layer-dd3972c25663>

² Revue de l'Académie américaine des sciences, PNAS, volume 114, issue 13, march 2017 « Overcoming catastrophic forgetting in neural networks », cf. <http://m.pnas.org/content/114/13/3521.abstract>

³ Todd Hester, Matej Vecerik, Olivier Pietquin, Marc Lanctot, Tom Schaul, Bilal Piot, Andrew Sendonaris, Gabriel Dulac-Arnold, Ian Osband, John Agapiou, Joel Z. Leibo et Audrunas Gruslys « Learning from demonstrations for Real World Reinforcement Learning », cf. <https://arxiv.org/abs/1704.03732>

⁴ Un article du complément scientifique du Guardian va dans ce sens : <https://www.theguardian.com/global/2017/mar/14/googles-deepmind-makes-ai-program-that-can-learn-like-a-human>

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 se 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 capacités 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 par Charles Babbage du premier « ordinateur », sous la forme d'une « machine à différences » inspirée des machines à tisser ;

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 commerce¹, 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.** Souvent, la capacité prédictive de ces technologies se trouve mobilisée.

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, en constante évolution, qui donnent lieu à des applications spécifiques pour des tâches toujours très spécialisées.**

¹ 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
Aéronautique, Espace 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.	 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).  Cybersécurité. Circulation aérienne intelligente.  Aide à la décision  Systèmes de pilotage autonome pour avion de combat.
Banque, Assurance Recherche et analyse des clients approfondie (assistance conseillers). Systèmes anti-fraude	Crédit mutuel : partenariat IBM Watson. BNPP : investissement dans Smarty.ai (Chatbots). Partenariat avec Critéo. SG : assistants conseillers. Axa : création d'un fonds pour la recherche finance des chercheurs qui travaillent sur l'IA et sur la confiance dans le Big Data
Energie, Environnement Surveillance de sites industriels. Villes intelligentes: exploitation des données clients avec capteurs intelligents (smart grid et IoT)	eDF : développement de réseaux intelligents (smart grids). Veolia : récolte et tri de déchets intelligente (avec Huawei)
Distribution, luxe, tourisme (B2C) Aide à la recherche produit/service et à la décision des clients Placement produit multi-canal optimisés.	L'Oréal : investissement en systèmes prédictifs et relation client. Accor : intégration moteurs IA, pour CRM et marketing avancé. Publicis : système Cyc pour aide à la décision.
Santé Développement de traitements plus efficaces. Exploitation de données santé clients adaptées.	Sanofi : co-entreprise Onduo (avec Google) lutte contre la diabète. Traitement du cancer avec IBM (Watson). Accès à une base de données de 118 malades avec la FDA (US). Dassault Systèmes avec Ipsen: Optimisation processus de R&D (consortium BioIntelligence)
Transports Analyse clients. Développement de systèmes autonomes.	Alstom : investissement navettes autonomes EasyMile RenaultValéo : développement de systèmes de conduite autonome. Investissement dans des capteurs intelligents. PSA : prototype de voiture autonome Picasso C4.
Télécoms Gestion et optimisation du réseau. Développement et traitement automatique de la relation client. Interfaces client (chatbots, box en connexion avec l'environnement).	Orange : apprentissage automatique appliqué à la relation clients et centres d'appel. Bouygues : monétisation des données et systèmes anti-fraude. Nokia : 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.

Depuis 2012, IBM détient un brevet l'autorisant à procéder à une estimation de la volatilité des transactions à haute fréquence.

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 commencé à travailler¹.

¹ 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>

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 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

Caractéristiques

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

Complexité de l'IA

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

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

Concernant le secteur **automobile**, le logiciel n'a, pendant longtemps, représenté qu'une fraction très réduite de 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 à 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 et qui a signé de nombreux partenariats à ce sujet, ainsi que l'a fait son concurrent **Renault**, qui accompagne ce développement industriel 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 à cette fin 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, consacré aux véhicules sans conducteurs. Quatre prototypes y sont testés par Uber depuis septembre 2016. Les quinze employés de Geometric 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 la transformer en profondeur : **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-sons.
- 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 à 2018¹**, 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

¹ Cf. www.worldrobotics.org/uploads/tx_zeifr/Executive_Summary_WR_2015.pdf

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 au **troisième rang mondial** dans la **recherche fondamentale en robotique** derrière les États-Unis 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, 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** dits *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. Vos rapporteurs ont constaté qu'en dépit de leur absence de mémoire, les *bots* font déjà un excellent travail en matière de **prévention du suicide** en réagissant à certains mots clés, ainsi que l'a expliqué et démontré *in vivo* à vos rapporteurs Alex Acero, directeur du projet Siri chez Apple. Siri reçoit par exemple 5 000 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. Ces exemples sont un premier niveau de gestion de la dépression ne nécessitant pas de mémoire dans les systèmes d'IA.

Vos rapporteurs ont pu tester en situation ces **usages bénéfiques pour la société des agents conversationnels**. Il ne s'agit que d'un exemple, leur utilité sera bien plus diverse en pratique, surtout quand les machines amélioreront leurs capacités en termes de mémoire. Au salon professionnel *E-commerce One To One 2017*¹, Google a invité à s'extraire du modèle du moteur de recherche tel que nous le connaissons encore aujourd'hui pour

¹ Cf. <https://www.ecommerce1to1.com/>

entrer dans l'ère de l'assistance, « *The age of assistance* » était ainsi le nom de sa présentation¹.

En matière d'**éducation**, les perspectives pour l'intelligence artificielle sont riches mais les applications restent encore rares. Ce point est développé plus loin dans le rapport.

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 jeux sérieux², ou *serious games* en anglais, tout comme les visites virtuelles³ pourront de plus en plus mobiliser des technologies d'IA.

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. L'IA permet de modéliser et de simuler et est donc utilisée en météorologie, sismologie, océanographie, planétologie ou encore en urbanisme. Le GIEC a recours à ces technologies pour analyser les changements climatiques.

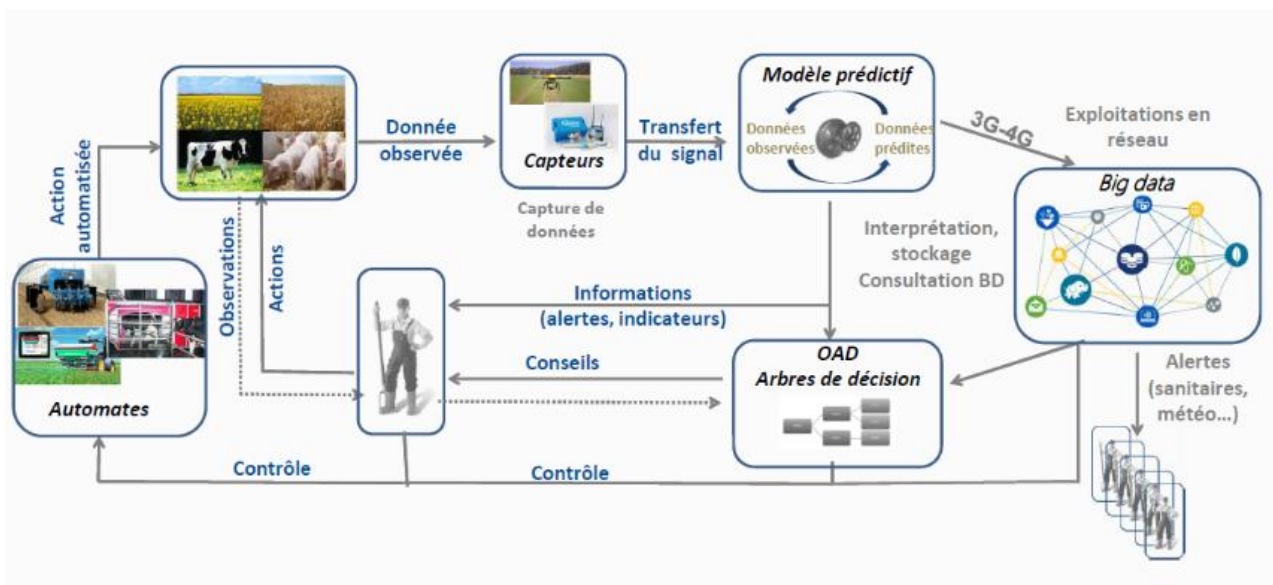
L'agriculture peut aussi être citée, car ce secteur présente des usages divers et des possibilités nombreuses. Les applications concernent la gestion des exploitations mais vont bien au-delà. Le graphique suivant décrit ainsi les cycles de traitement des données dans l'agriculture. À chaque étape, l'intelligence artificielle peut jouer un rôle de plus en plus grand.

¹ Cette économie de l'assistance repose sur trois déterminants principaux : la personnalisation de la relation avec le mobinaute (plus encore que l'internaute), le web sémantique (la question donne lieu à une réponse directement, voire la machine anticipe vos questions) et, enfin, le web vocal (20 % des requêtes formulées aujourd'hui sur Google via un mobile Android sont vocales et ce chiffre devrait dépasser les 50% en 2020). Cf. <http://www.frenchweb.fr/google-annonce-la-mort-du-moteur-de-recherche-et-lavenement-de-lage-de-lassistance/286043>

² Pour Julian Alvarez et Olivier Rampnoux, cinq types de jeux sérieux peuvent être distingués : les advergamings (jeux publicitaires), les edutainments (à vocation éducative), les edumarket games (utilisés pour la communication d'entreprise), les jeux engagés (ou détournés) et les jeux d'entraînement et de simulation.

³ Outre l'accès à la culture (visites de musées et de sites archéologiques par exemple), il pourrait s'agir de visites interactives dans des environnements historiques différents, permettant des sortes de voyages dans le temps.

Les processus de traitement des données dans l'agriculture



Source : INRA

De manière opérationnelle, de **nombreuses applications sont possibles dans le secteur agricole**, pour l'exploitant (dans le suivi ou la régulation), la filière agricole concernée (en matière de progrès génétiques par exemple) mais aussi l'ensemble de la société (avec le cas des réseaux d'épidémiologie-surveillance). Le graphique suivant l'illustre.

Exemples d'utilisations des données dans l'agriculture

Suivi des rendements et de la qualité

prévisionnel

Localisation des traitements au champ

Progrès génétique

Omic + capteurs

Conduite des troupeaux
La vache connectée

À la récolte

Régulation d'ambiance dans les bâtiments et les serres

Réseaux d'épidémiologie surveillance

Source : INRA

Il peut être donné l'exemple d'une entreprise qui a ainsi mis au point un système de conteneurs agricoles 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. La technologie permet ici d'améliorer l'efficacité du cycle de production, et ce 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.

En matière de **sécurité numérique**, 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. La DARPA a ainsi consacré son grand concours 2016 à ce sujet.

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, de même qu'elle prévoit les étapes de réparation.

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, le sujet a déjà été traité à plusieurs reprises par l'OPECST et 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* »¹. 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 cancer².

Quelques cas emblématiques peuvent être mentionnés. En **génomique**, par exemple pour la validation et la critique des thérapies, outre le *deep learning*, peuvent être utilisés 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**

¹ • Le numérique au service de la santé, *Catherine Procaccia et Gérard Bapt*, rapporteurs, Sénat n° 465 (2014-2015).

² 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>

repliage de protéines¹ ou de **segmentation des IRM du cerveau** en vue d'identifier certaines zones, 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, 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 dans **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 être utilisé pour des applications de différentes natures. Il faut rappeler 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 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**

¹ Ce processus physique, par lequel un polypeptide se replie dans sa structure tridimensionnelle caractéristique dans laquelle il est fonctionnel, est important en ce que de nombreuses maladies, en particulier les maladies « neurodégénératives », sont considérées comme résultant d'une accumulation de protéines « mal repliées ».

du langage naturel, de **reconnaissance automatique de la parole**, ou, encore de **bioinformatique**, à travers par exemple l'étude de l'ADN, ouvrent 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, disparaissent parfois 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 Prolog¹, des réseaux de neurones profonds² ou du programme AlphaGo³, **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**⁴. 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 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*

¹ Les raisonnements formels de Prolog ont été enrichis de la méthode de programmation par contraintes.

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

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

⁴ 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.

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 à l'égard de **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* »¹, 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. La complémentarité homme-machine est l'une de celles-ci, avec les potentialités d'amplification de l'action et d'amélioration de l'efficacité offertes par l'intelligence artificielle.

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 faisant absorber des milliers d'images et en les corrigeant quand ils font des erreurs d'interprétation. Or les humains découvrent le monde de façon non supervisée. Un enfant reconnaît ses proches très vite et distingue rapidement un lion d'un chat, sans apprentissage supervisé etc.

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

¹ 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 certain que l'on ainsi puisse parvenir à l'apprentissage non supervisé.

Ils relèvent que les **travaux en robotique développementale et sociale** de Pierre-Yves Oudeyer, responsable de l'équipe Flowers d'Inria, sont particulièrement féconds : il s'agit, en faisant appel à de nouvelles disciplines connexes (neurosciences et psychologie développementale), de concevoir des algorithmes et 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, comme en témoigne la liste suivante qui n'est pas exhaustive : **Yann LeCun** (Facebook), **Andrew Ng**

(Baidu, ex-Google), **Geoffrey Hinton** et **Fei Fei Li** (Google), **Vladimir Vapnik** et **Rob Fergus** (Facebook), **Nando de Freitas** (Google), **Zoubin Ghahramani** (Uber), même **Yoshua Bengio**, qui a longtemps refusé de travailler 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 »¹, ou par leurs équivalentes chinoises dites « BATX »². 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 que ce phénomène 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, continuellement enrichis par leurs clients et usagers.

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 même pour le chinois Baidu. Apple, longtemps discrète, a décidé, en 2016, d'offrir à ses chercheurs la possibilité de publier dans des revues scientifiques³. Six chercheurs de l'entreprise ont ainsi publié un article sur l'apprentissage machine à partir d'images de synthèse.

¹ 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.

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

³ 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-up »*. Ainsi, le 13^e 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 consacrés 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 »**¹.

L'entreprise **Baidu** a développé le **principal moteur de recherche chinois**, site le plus consulté en Chine et **5^e 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

¹ 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 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.

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, en conduite autonome, pilotage de drones,... Elle travaille en étroite relation avec de nombreuses universités et start-up.

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 ses 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^{er} sur cinq des huit évaluations conduites parmi quatre tâches.

Sur FDDB¹ (*Face Detection Data Set and Benchmark*) et sur la base de données de visages LFW, ce système progresse 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* KITTI² 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 :

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

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

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 ».

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 présente 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 (télécom, 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 chinois 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 essentiellement conduite par**

des hommes. Cette situation d'extrême masculinisation est critiquable et n'est pas souhaitable.

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

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 et souvent recherchée par les équipes de chercheurs mais, elle-même sous-domaine de l'informatique, l'intelligence artificielle demeure éclatée en une **cinquantaine de sous-domaines de recherche** identifiés, qui parfois s'ignorent les uns les autres ou ne coopèrent pas suffisamment.

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 Web sémantique	Optimisation Méthodes bayésiennes	Reconnaissance de la parole et traduction automatique	Recherche dans des banques d'images et de vidéos	Flottes de robots		Théories SAT		Apprentissage multi-agents	
Sociologie	Ontologies	Réseaux de neurones ou neuronaux		Reconstruction 3D et spatio-temporelle	Apprentissage des robots		Raisonnement causal, temporel, incertain		Ingénierie Multi-Agents (Langages, plateformes, méthodologies)	
Humanités numériques		Méthodes à noyau Apprentissage profond Fouille de données Analyse de données massives		Suivi d'objets et analyse des mouvements Localisation d'objets Asservissement visuel	Cognition pour la robotique et les systèmes		Programmation par contraintes Recherche heuristique Planification et ordonnancement		Simulation Multi-Agents (intérêts aussi les SHS)	

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 démarches et les connaissances.** Elles

peuvent même aller jusqu'à s'inscrire dans d'autres espaces disciplinaires (mathématiques, physique, biologie...).

Issue des mathématiques, de la logique et de l'informatique, l'intelligence artificielle fait appel depuis des décennies, et de plus en plus, 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 de recherche 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. Aujourd'hui, ce courant scientifique situé à la frontière de l'informatique et des « sciences de la culture » est désigné 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^e 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*), le terme étant plus pertinent que celui « d'incorporées » car elles s'incarnent dans des corps physiques plus qu'elles n'y seraient simplement placées. Selon lui, au-delà du traitement du langage naturel, les problèmes d'interaction et d'action conjointe homme-robot posent des questions d'interdisciplinarité : 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 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) 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, une pluricom pétence dans les équipes, indispensable au progrès dans ce domaine où se télescopent les disciplines. Selon Bertrand Braunschweig, directeur du centre de d'Inria de Saclay, « *nous avons besoin de doubles, voire de triples-compétences* ».

Comprendre l'intelligence sous toutes ses formes 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 scientifique¹, notamment en matière d'intelligence artificielle. Pour Jean-Gabriel Ganascia, cela résulte notamment des modes de financement : les logiques de court terme des projets n'offrent pas la pérennité et la vision de long terme souvent souhaitables et rendent difficiles le développement de grandes ambitions. Il note toutefois 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 Gérard Sabah, membre de l'Académie des technologies, c'est la structure de la recherche française qui génère la recherche en intelligence artificielle parce que, selon lui, à l'exception du cas des chercheurs du CNRS

¹ Pierre-Gilles de Gènes alerte ainsi : « 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 termes, 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 ».

(eux-mêmes 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 (en France). 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 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 constitue une appréciation controversée : souvent reproduite, elle devrait être évaluée et approfondie.

Par ailleurs, avec un point de vue plus interdisciplinaire il est possible de supposer que l'intelligence artificielle permettra aussi de **mieux modéliser le fonctionnement du cerveau** et les relations entre le cerveau et la conscience, ainsi que la **mémoire** ou la récupération et la stimulation des **facultés cognitives et/ou motrices**.

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 sur place en Suisse devraient aller dans ce sens. Jean-Pierre Changeux, partie prenante des projets, leur en a fait partager l'ambition.

L'équipe du BBP de l'École polytechnique fédérale de Lausanne (EPFL) se donne pour objectif depuis mai 2005 de créer un cerveau synthétique par rétroingénierie.

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 européenne¹ 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 plateforme informatique 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 2015², ce qui a conduit à le réorienter en partie, 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 deux ans et demi. Un premier bilan est donc à dresser.

1 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).

2 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.

Pour simuler le fonctionnement d'un cerveau humain, la puissance calculatoire nécessaire est estimée à un exaflops¹. 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.

Visant la création de modèles pour étudier les maladies neurodégénératives, une équipe française du laboratoire CellTechs de l'école SupBiotech, en collaboration avec le CEA (Commissariat à l'énergie atomique et aux énergies alternatives) travaille sur la « reprogrammation » de cellules souches sans aucune spécialisation, dites alors « pluripotentes induites » (ou IPS), afin de les différencier pour devenir, par exemple, des **cellules cérébrales** : Frank Yates, enseignant-chercheur à SupBiotech et responsable de cette équipe explique que leurs cultures d'organoïdes neuroectodermiques « comptent entre 500.000 et un million de cellules, et mesurent environ 2 mm de diamètre »².

Les perspectives d'application de ces projets et de BBP et de HPB en particulier sont grandes, mais la perspective d'un **homme non seulement réparé mais augmenté qui pourrait à terme 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 (même si les cas du HBP de l'EPFL peut s'apparenter à cette famille de grands projets pluridisciplinaires), le caractère pluridisciplinaire est évident : au-delà du matériel informatique et des logiciels utilisés, cœur historique de l'intelligence artificielle, d'autres savoir-faire et domaines de connaissances sont largement mobilisés : la robotique, au sein de laquelle 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. Il faudrait aussi revoir les modalités de carrière des chercheurs en encourageant les parcours

¹ 10¹⁸ soit un milliard de milliards, ou un trillion de flops (un flop est une unité de mesure de la vitesse d'un système informatique).

² Cf. <http://www.futura-sciences.com/sante/actualites/cerveau-faire-pousser-mini-cerveau-cest-possible-66821/>

transgressant les frontières disciplinaires et en valorisant dans les évaluations les expériences de partenariats, notamment public-privé.

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 ainsi que ses applications sont soumises 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 qu'un sondage par échantillon représentatif, 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 devant une **logique d'appréciation contrastée** qui n'est sans doute pas sans lien avec l'accent mis sur les risques de l'IA dans la culture populaire (littérature et surtout 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 cette analyse selon laquelle les inconvénients de l'intelligence artificielle seraient, à ce stade, supérieurs à ses avantages 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 autre **sondage, mené par l'IFOP pour la CNIL** en janvier 2017, peut être cité¹. 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 %).

¹ Cf. https://www.cnil.fr/sites/default/files/atoms/files/presentation_ifop_-_presentation.pdf

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

Enfin, deux études parues en avril 2017 au sujet de l'utilisation de l'intelligence artificielle à des fins commerciales vont dans le même sens¹ : la première montre que **72 %** des 6 000 internautes sondés dans six pays **déclarent en avoir peur**² et la seconde montre que **42 %** des 2 000 personnes interrogées **ne font pas confiance** à l'intelligence artificielle³.

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**⁴, des **pannes**

¹ Cf. une synthèse dans l'article suivant : <http://www.zdnet.fr/blogs/watch-it/intelligence-artificielle-les-consommateurs-restent-craintifs-39851386.htm>

² <https://www.pega.com/ai-survey>

³ <https://uk.insidesales.com/research-paper/state-artificial-intelligence-uk-2017-public-perceptions-disruptive-technology/>

⁴ 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

informatiques¹, du piratage, de l'escroquerie commerciale, d'un nivellement intellectuel par le bas² 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 n'entonnent pas le refrain d'un scientisme optimiste et naïf devant les 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, alimentant la peur des véhicules autonomes. 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 sept 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**, mais sans victime, 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

l'appauvrissement. 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 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 ! ».

¹ La peur infondée du « bug de l'an 2000 » dans les années 1990 avec son cortège de commentaires médiatiques excessifs devrait être un exemple à méditer.

² « 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*).

raciste en miroir de ses interlocuteurs¹ quelques heures après son activation le 23 mars 2016 ;

- une association française contre l'intelligence artificielle (AFCIA) a déposé ses statuts le 18 juillet 2015². Son président Cédric Sauviat a été rencontré par vos rapporteurs, qui ont jugé à la fois son discours bien construit et ses analyses excessives voire infondées, révélatrices d'un certain climat d'angoisse puisque la France serait le seul pays où une telle association existerait³. L'AFCIA juge « *illégitime et dangereuse la recherche scientifique visant à créer des organismes à intelligence artificielle suprahumaine* » 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 artificielle⁴.

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 superintelligence : « *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 votre rapporteur Claude de Ganay.

¹ 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.

² 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/>

³ Ce constat est renforcé par l'existence d'autres associations technophobes, inspirées 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/>

⁴ 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 dans l'introduction du présent rapport.

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 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, particulièrement en France où les préoccupations en matière de sécurité et de données personnelles sont fortes et font régulièrement débat.

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 interviennent dans la recherche en intelligence artificielle

Notre pays dispose, en matière de recherche en intelligence artificielle, d'**importants atouts à faire valoir**, riches 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 (Polytechnique, ENS, Mines-Télécom...) 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, Inria, le CNRS, le CEA, différentes universités et grandes écoles sont les **principaux organismes de recherche publique en intelligence artificielle et produisent des travaux à visibilité internationale**.

Cette excellence est reconnue à l'international comme ont pu le constater vos rapporteurs lors de leurs déplacements. Le bon niveau des étudiants et des enseignants est également souvent cité. Il leur a même été demandé à San Francisco de former en France plus d'étudiants pour alimenter en ressources la *Silicon Valley*.

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 en

1989 l'Institut Blaise Pascal (IBP) 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 France¹. Beaucoup sont rappelés sous leur nom actuel au paragraphe suivant.

Un inventaire des douze principaux laboratoires du CNRS en matière d'intelligence artificielle ayant été dressé par Gérard Sabah², 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) basé à l'Université Paris-Dauphine, 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), établi à Paris, 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 ;

¹ 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 de Gérard Huet) à Rocquencourt, le GIA (autour d'Alain Colmerauer) à Marseille, le CRIN à Nancy ou, encore, le CERFIA à Toulouse.

² Cet inventaire est présenté dans la brochure de l'Académie des Sciences déjà citée.

- le laboratoire d'informatique de Paris-Nord (LIPN) 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) spécialisé dans l'apprentissage, les contraintes, la représentation des connaissances, les systèmes multi-agents, le traitement 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.

En plus de ces laboratoires publics, de nombreuses entreprises françaises ont établi des laboratoires en partenariat avec des organismes publics de recherche. Il peut être cité les cas des entreprises Thalès, EDF, Engie, PSA, Total ou encore Solvay (entreprise belge mais cotée à la bourse de Paris et faisant partie de l'indice CAC 40). Les entreprises étrangères font aussi de la recherche fondamentale en IA sur le territoire national, à l'instar de Facebook, IBM, Microsoft ou encore Huawei.

Votre rapporteure Dominique Gillot souligne la volonté des organismes de recherche de développer des partenariats avec les entreprises pour soutenir la fécondité des initiatives de type start-up (dont il sera question aux pages suivantes) et surtout d'offrir un cadre de recherche stabilisé au-delà du rythme contractuel souvent dénoncé, soit trois ans.

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

Les **start-up 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 vidéos¹, 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. Bon exemple, la start-up **Angus.AI**, créée par d'anciens ingénieurs de l'entreprise **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 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 son 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 Orpailleur² mené à Nancy en lien avec

¹ 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.

² 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

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-up.

3. Une reconnaissance internationale de la recherche française et qui s'accompagne d'un phénomène de rachat de start-up 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 ont constaté un phénomène de **rachat de start-up 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.

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.

Votre rapporteure Dominique Gillot relève que les salaires sont beaucoup plus élevés à San Francisco qu'à Paris, mais les conditions de vie n'y sont pas toujours meilleures même avec un salaire bien supérieur (vie plus chère, infrastructures de transports, de santé, présence d'écoles pour les enfants etc.) : l'équilibre avantages/inconvénients est donc plus complexe qu'il n'y paraît à première vue, surtout que la France - et Paris en particulier - représente un écosystème de qualité, avec une vie sociale et culturelle très riche, un environnement intellectuel stimulant, des lieux de formation performants et des chercheurs de très bon niveau.

Il faut permettre à ces jeunes génies, qui sont autant d'entrepreneurs en devenir, de **disposer d'opportunités en France et permettre aux start-up 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** - de **l'association française pour l'intelligence artificielle (AFIA)** en particulier - **à travers les *meetups***. Le principal d'entre eux, le « *Paris Machine learning Meetup* »¹, auquel ont été invités vos rapporteurs, représentent 5 205 membres au 1^{er} 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*.

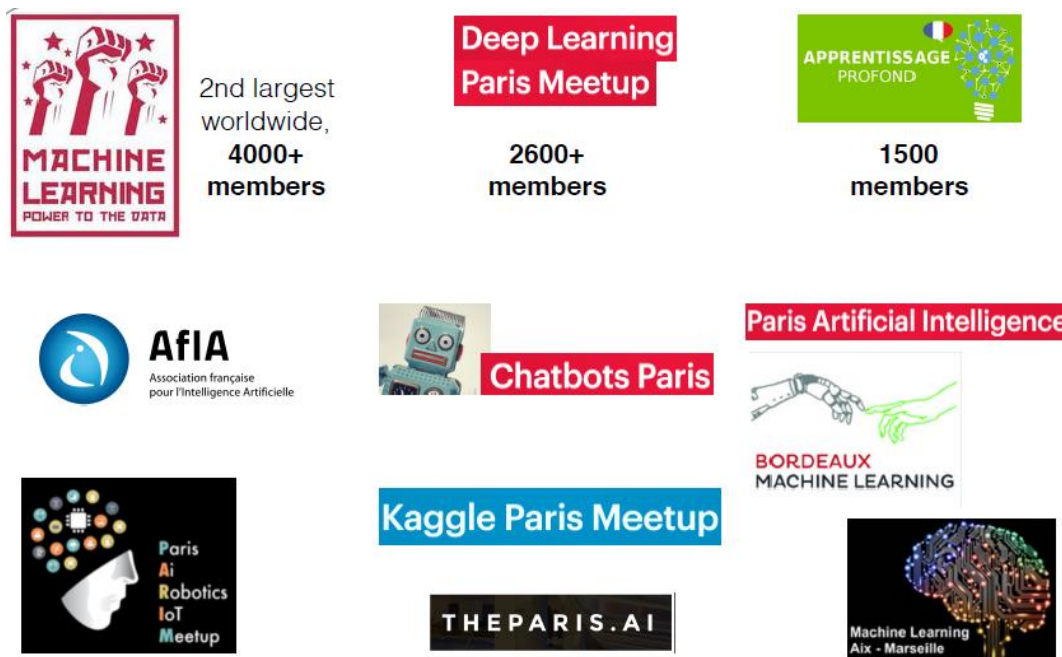
D'autres *meetups* peuvent être cités : « *Paris.AI Meetup* », « *Big Data Paris Meetup* », « *Deep Learning Meetup Paris* », « *Big Data et Machine learning Paris Meetup* ».

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, ce qui était une nécessité 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. Elle aurait tout intérêt à **transcender ses propres limites pour relever le défi d'une intelligence artificielle française ouverte, visible et conquérante**.

¹ 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/>

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 fait état 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 postdoctorales. 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 tout en disposant d'un **réseau de chercheurs très compétents et d'un tissu de start-up très dynamique**. Les atouts considérables du système d'enseignement supérieur et de recherche français sont connus : excellence de l'école mathématique française, qualité de la recherche et des diplômés...

On dénombre un total de **240 start-up spécialisées en intelligence artificielle** (comme, en France, Partnering Robotics, 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-up, que l'initiative French Tech 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. Depuis, cette connaissance s'est améliorée, la prise en compte de cet atout est notamment visible du côté des pouvoirs publics dans la stratégie « France IA ».

Les start-up 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 consacrée 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* ».

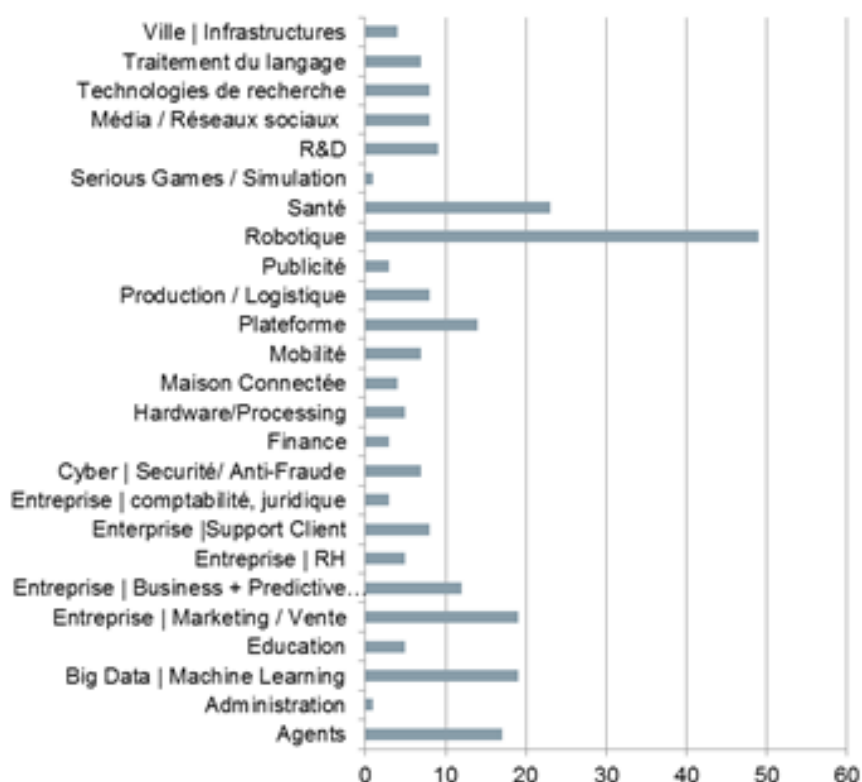
Mais, comme le précise un article de *Venture Beat*¹, utilisant des données collectées par Paul Strachman, entre janvier 2014 et mi-octobre 2016, une trentaine de start-up françaises spécialisées en intelligence artificielle levaient 108 millions de dollars (98 millions d'euros), quand dans le même temps, outre-Manche, huit start-up 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.

Selon votre rapporteure Dominique Gillot, la disponibilité des données et des articles auxquels se référer ne rend pas réellement compte de la réalité nationale dans une démarche où les choses évoluent très vite : chaque jour des start-up naissent de la rencontre et de la volonté de jeunes (ou de moins jeunes) diplômés qui découvrent les opportunités de lier l'emploi et la valeur. Même si tous ces créateurs de start-up ne rêvent pas d'être cotés en bourse et d'être rachetés par de grands groupes, l'intérêt de ces derniers existe : ils voient dans cette effervescence un moyen d'expérimenter de nouvelles pratiques, des découvertes qui pourront se révéler fructueuses, dans un cadre beaucoup plus souple et réactif que celui contraint par des règles du droit du travail et de la législation vécue comme un carcan. Les structures qui investissent dans des secteurs peu identifiés garantissent la plasticité et la réactivité de ces initiatives dont il faudra analyser la longévité.

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

Les start-up françaises en intelligence artificielle par domaine d'application



Source : Gouvernement

L'exemple de Snips

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 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). Elle travaille aussi sur des applications verticales : pour les véhicules connectés, dans l'hôtellerie, la maison connectée et les loisirs numériques. Elle s'appuie sur du *deep learning*, des modèles probabilistiques, du traitement du langage, de la gestion de graphes et du cryptage de données pour garantir le respect de la vie privée.

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 actes des cardiologues mais bien 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 à l'étranger d'une **importante diaspora des chercheurs français en intelligence artificielle**. Elle pourrait sans doute être davantage connue, animée, considérée et mise à contribution.

Au final, vos rapporteurs déplorent la **sous-estimation des atouts considérables de la France** mais ils appellent l'attention sur la nécessité 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 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 « **Gafa** », parfois des « **Gafami** », mais il serait plus juste de parler des « **Gafamitis** »¹, « **Natu** »² et « **Batx** »³. 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*

¹ 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.

² Netflix, Airbnb, Tesla et Uber.

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

vendeurs de pioches et de pelles. C'est exactement ce que nous sommes : nous fournissons des technologies ». Grâce à leurs efforts en R&D et aux rachats de start-up, 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-up en intelligence artificielle achetées de 2011 à 2016, on a vu, en 2016, 40 start-up 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

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

Assistants
Santé
Vidéo
Prédictif 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*. IBM, 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-up 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**, dont 800 en France. Selon l'entreprise, aujourd'hui, « *à l'ère des systèmes cognitifs, les systèmes peuvent apprendre après 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 « Geometric 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. 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 d'évoluer 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*). • À 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'attendent.

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-up 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 de haut niveau ; il serait dommage que la France se fasse piller ses talents. Il est donc indispensable d'aider les start-up 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 avec des partenaires tels que l'Allemagne.

Source : 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, GAFAMITIS¹, NATU² et BATX³ fait courir d'importants risques au principe traditionnel de souveraineté ainsi qu'aux systèmes démocratiques que nous connaissons.

¹ Google, Apple, Facebook, Amazon, Microsoft, IBM, Twitter, Intel et Salesforce.

² Netflix, Airbnb, Tesla et Uber.

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

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

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 »*¹.

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 »*.

¹ 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>

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 s'agit d'apprécier globalement l'incidence future 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* »¹ 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 création d'emploi**. Elles soulignent également l'importance des effets indirects positifs sur l'emploi des innovations de produits, qui viennent plus que contrebalancer 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,

¹ 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

T. Gregory et U. Zierahn sur « *Les risques de l'automatisation pour l'emploi* »¹, 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 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. Osborne², 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 Berger³ 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 numériques.

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 » devant 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

¹ 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

² Pour consulter l'étude de la *Oxford Martin School* : http://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf

³ 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

si leur caractère indirect les rend plus difficile à quantifier. À 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 délicates, 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'Eurostat¹, 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.

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

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 macro-économiques, 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 ;

- **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 détruise 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 multimodales, enrichies et transparentes. Cette coopération devrait être heureuse.

Votre rapporteure Dominique Gillot propose de **réfléchir à un nouveau mode de financement du système de protection sociale, qui serait un complément des cotisations existantes et qui pourrait consister en un prélèvement additionnel aux cotisations sociales, assis sur les agents autonomes et les robots, 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 Delvaux¹. 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.

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

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.**

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 en effet 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 Merriaux, 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 Merriau 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**, car venant en soutien de l'effort pédagogique. Jean-Marc Merriau 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. Ces ressources en ligne **permettront aux apprenants (en formation initiale ou continue) 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 **l'imminence de la possibilité 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 de fourniture d'énergie permettant une consommation optimisée grâce à l'IA) et les *smart cities* (villes intelligentes) seront les expressions des bénéfices que nous pouvons tirer de l'intelligence artificielle. Et cela se traduira en matière de transports, de sécurité, de santé, de dépassement de la dépendance et du 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 de 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 *data-centers* et de la réduire drastiquement grâce à l'intelligence artificielle, tout en encourageant la production d'énergies renouvelables. En partenariat avec EDF, Kawartech développe à Toulouse un contrôle intelligent et autonome de l'éclairage public.

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'acceptabilité sociale, notamment en France, que le passage aux voitures autonomes risque de se voir retarder. 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...). Il peut être constaté que les véhicules collectifs autonomes, tels que les bus et navettes, sont mieux acceptés et sont d'ailleurs déjà en service.

Votre rapporteur Claude de Ganay a été étonné par des propos tenus par le 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 au profit du monde rural et des seniors.

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.

Au-delà de l'automobile et de la santé, d'autres secteurs amélioreront leurs offres au profit de toute la société : le contrôle de la qualité de l'air, avec les robots d'analyse de l'atmosphère, la domotique et l'électroménager etc.

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

Les agents conversationnels, *chatbots* ou *bots*, les robots de service, les agents d'assistance, d'accompagnement, d'aide à la mobilité vont progressivement cohabiter avec les humains. Cela nécessitera une grande **vigilance**. Une adaptation réelle sera aussi requise.

Au quotidien, nos *smartphones* avec des logiciels tels que « Siri », « Cortana » ou « Google Now », font d'ores et déjà cohabiter leurs utilisateurs avec des algorithmes puissants, qui connaissent beaucoup d'aspects de la vie de chacun d'entre 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 - souvent pour plus de sécurité et d'autonomie - 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 encore assez peu souples dans leur fonctionnement, ce qui peut contribuer à susciter de la méfiance.

Alors que la robotique industrielle connaît des applications selon le respect de la règle d'évitement des 4 D : « *dangerous, dull, dirty, dumb* » (dangereux, ennuyeux, sale, idiot), la robotique de service 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 de consacrer une grande attention aux logiques d'empathie et aux aspects émotionnels qui ne manqueront de se développer.

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 de s'y attarder 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* ».

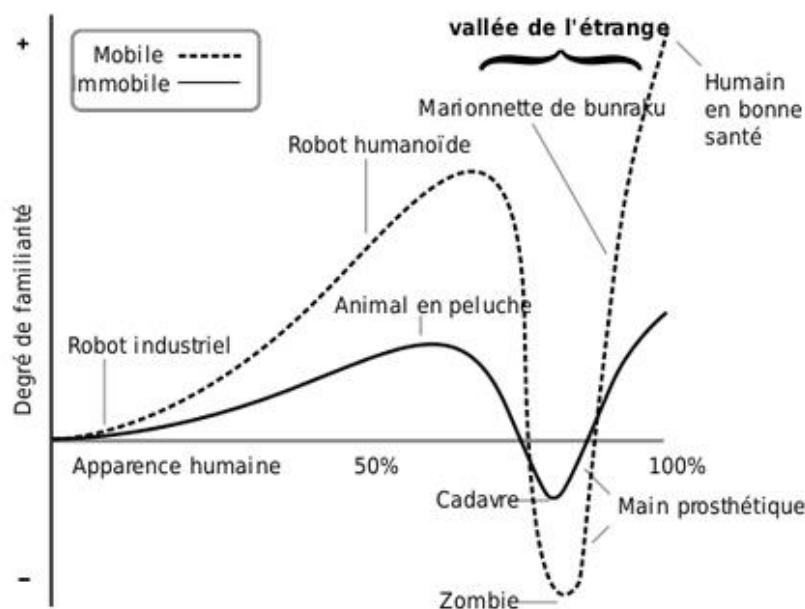
Pour le psychanalyste Serge Tisseron, spécialisé en intelligence artificielle et en robotique¹, **utiliser ces techniques pour faire croire que les robots seront capables de sentiments serait malhonnête** car la reconnaissance des émotions présente avant tout pour eux l'intérêt de permettre de s'adapter 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 ou tromper sur les intentions de la machine.

¹ « *Le jour où mon robot m'aimera* », 2015.

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.

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



Source : creativecommons.org/licenses/by-sa/3.0/

Votre rapporteur Claude de Ganay rappelle que dans la nouvelle de science-fiction « *L'histoire du robot et du bébé* » de 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, il est imaginé qu'après l'émergence de l'intelligence artificielle forte et avec la vaste pénétration des robots dans la vie quotidienne, des 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 ».

Au-delà de l'exemple tiré de cette fiction¹, les **enjeux de la cohabitation quotidienne avec des intelligences artificielles, de leur acceptation et de leur régulation doivent et devront être appréhendés.**

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 d'un 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 2017², 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.

¹ L'évocation de cette nouvelle science-fiction permet aussi de replacer la réflexion dans les objectifs réellement poursuivis, qui ne sont ni l'IA forte, ni les robots d'assistance maternelle remplaçant les mères.

² La résolution a été adoptée par 451 voix pour, 138 voix contre et 20 abstentions.

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 ainsi qu'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.

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. Ils pensent que 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 sa qualité de 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***¹ sous-titré « *seizing opportunities, preserving values* », piloté par John Podesta, au sein du bureau exécutif du

¹ 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>

président des États-Unis¹ (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 2016², 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 l'avenir de l'intelligence artificielle** »³ (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** »⁴ (*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**⁵ (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 monde⁶. 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

¹ 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.

² 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. :

³ <https://www.whitehouse.gov/blog/2016/05/03/preparing-future-artificial-intelligence>
⁴ https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

⁵ https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/national_ai_rd_strategic_plan.pdf

⁶ <https://www.whitehouse.gov/sites/whitehouse.gov/files/images/EMBARGOED%20AI%20Economy%20Report.pdf>

⁶ '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/>

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 publics co-animés tenus à travers le pays, ainsi qu'une demande d'information¹, 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.

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 fouillée². Ainsi, les

¹ Cette procédure « Request for information » correspond à une consultation publique formalisée.

² 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.

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 lois¹.

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

¹ 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.

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* »¹. 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*

¹ Pour consulter le rapport :
<https://www.publications.parliament.uk/pa/cm201617/cmselect/cmsctech/145/145.pdf>

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 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 ses auteurs, 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 œuvre.

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 réponse¹ 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 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 recherche² 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 domaine³. 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

¹ La réponse du Gouvernement britannique est disponible ici :

<https://www.publications.parliament.uk/pa/cm201617/cmselect/cmsctech/896/89602.htm>

² Pour consulter le projet de la Royal Society : <https://royalsociety.org/topics-policy/projects/machine-learning/>

³ Une note complète sur le sujet de la place de l'IA en Chine est annexée au présent rapport.

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 et les start-up* ».

Le 13^e 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 consacrés 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 ».

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 à 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 Suining¹. D'autres dispositifs d'évaluation individuelle des citoyens ou des consommateurs sont étudiés, comme le *Sesame Credit* du distributeur en ligne Alibaba².

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

¹ 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

² Cf. <http://www.bbc.com/news/world-asia-china-34592186>

sont l'une des dimensions majeures des progrès des algorithmes de *deep learning* et d'apprentissage statistique de façon plus générale.

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 Japon¹. Le gouvernement japonais a fait le choix d'annoncer dès mai 2015 un **grand plan quinquennal pour la robotique et l'intelligence artificielle**², entré en vigueur le 1^{er} 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^e révolution industrielle », une « société 5.0 », société superintelligente 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 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 artificielle³. 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 consacrée aux Sciences et technologies de l'information et de la communication**, format qui n'avait pas été mis en œuvre depuis 20 ans.

¹ Une note complète sur le sujet de la place de l'IA au Japon est annexée au présent rapport.

² Soutenu par 200 entreprises et universités, le 5^e Plan-cadre pour la Science et la technologie prévoit le triplement des investissements dans ce domaine d'ici à 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.

³ Plusieurs exemples de projets sont détaillés dans la note sur le sujet de la place de l'IA au Japon annexée au présent rapport.

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* »¹ (*Advanced Integrated Intelligence Platform Project Center*), l'AIRC (*Artificial Intelligence Research Center*), hébergé par l'AIIST (*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 consacré 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

¹ Les activités s'articulent autour de cinq 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 génie civil, résilience aux catastrophes naturelles...) ; prise en compte des aspects éthiques, légaux et sociaux ; développement des ressources humaines.

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 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 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'OCDE¹ 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

¹ <http://www.oecd.org/sti/ieconomy/Yuko%20Harayama%20-%20AI%20Foresight%20Forum%20Nov%202016.pdf>

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 ?

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 République¹. 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, 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 à cinq ans d'investir dans dix start-up 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 à la fin 2017 ;

- le lancement d'un appel à projets pour des plateformes sectorielles de partage de données pour trois à six secteurs, d'ici à 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 à l'été 2017 d'une concertation de France Stratégie sur la question des effets de l'intelligence artificielle sur l'emploi.

¹ Cf. 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 et
http://www.economie.gouv.fr/files/files/PDF/2017/Dossier_presse_France_IA.pdf

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 dix ans** et le **livre blanc de l'enseignement supérieur et de la recherche**¹.

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 au comportement des robots. Ces « trois lois », qui s'apparentent à des règles éthiques, 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* » ;

- 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* ».

¹ 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.

Vos rapporteurs rappellent que **les « lois d'Asimov » sont des règles issues de la fiction et qu'elles posent des problèmes réels 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 prémices 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 ressemblants) 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 dignité¹. 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 libertés², **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**

¹ 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 ».

² 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>

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**¹.

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 commercialisation**². 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**

¹ Ernst Kantorowicz, « *Les Deux Corps du roi. Essai sur la théologie politique au Moyen Âge* », Gallimard, 1957.

² 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.

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 a été 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, à la suite de la construction jurisprudentielle et doctrinale reconnaissant la notion de savoir-faire. Enfin, conformément à la « Classification de Nice »¹, 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.

¹ Les 45 classes établies par la « Classification de Nice » sont disponibles ici : <http://web2.wipo.int/classifications/nice/nicepub/en/fr/edition-20170101/taxonomy/>

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, le plus souvent avec notre complicité, **sans que nous ne connaissions l'usage qui peut être fait de ces millions d'informations à caractère personnel**. Les agents conversationnels étant appelés 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**¹, 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* »². 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 communes et des sanctions renforcées dans les cas de traitements de données transnationaux* »³.

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

¹ Le texte intégral du règlement est disponible ici : <http://eur-lex.europa.eu/legal-content/FR/ALL/?uri=CELEX%3A32016R0679>

² 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>

³ Ibid.

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^{er} 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 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* »¹. 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 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'État* »².

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.

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.

¹La 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>

² Le droit applicable à la communication des documents administratifs est disponible ici : https://www.legifrance.gouv.fr/affichCode.do?jsessionid=9D32BEC5092D1A4CC3D228A1256F8AB.A.tpdila23v_2?idSectionTA=LEGISCTA000031367696&cidTexte=LEGITEXT000031366350&dateTexte=20170312

Afin de répondre plus spécifiquement aux préoccupations d'explicabilité, de régulation et de responsabilité des algorithmes, le Conseil général de l'économie (CGE) a remis au Gouvernement un rapport intitulé « **Modalités de régulation des algorithmes de traitement des contenus** »¹. Cinq préconisations ont été formulées en vue de vérifier la conformité aux lois et règlements dont la détection de discrimination illicite. Suite à ce rapport, le secrétariat d'État à l'économie numérique a confié à l'Institut national de recherche en informatique et en automatique (Inria) le rôle d'opérateur de la **plateforme scientifique d'évaluation de la responsabilité et de la transparence des algorithmes** avec le soutien de l'IMT et du CNNum. Ce projet a été placé sous la direction de Nozha Boujemaa, directrice de recherche chez Inria. Les travaux de la plateforme « TransAlgo »² ont ainsi été lancés en janvier 2017 et rassemblent des chercheurs de plusieurs établissements (SciencePo, UVSQ, CEA, CNRS, Inria, IMT etc).

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. L'encadré suivant rappelle l'échange que notre collègue a eu à ce sujet en séance publique avec le Gouvernement, représenté par 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.

¹ <http://www.economie.gouv.fr/cge/modalites-regulation-des-algorithmes-traitement-des-contenus>

² http://www.economie.gouv.fr/files/files/PDF/Inria_Plateforme_TransAlgo2016-12vf.pdf

Inégalités devant l'orientation après le bac (extrait des débats¹)

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.

¹ L'intégralité des débats conduits lors de la séance du 15 novembre 2016 après 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

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**¹. Dans ce 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 défectueux² 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 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

¹ Cf. http://www.eu-robotics.net/cms/upload/PDF/euRobotics_Deliverable_D.3.2.1_Annex_Suggestion_GreenPaper_ELS_IssuesInRobotics.pdf

² 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.

robots : le régime de **responsabilité du fait des produits défectueux**¹, celui de la **responsabilité du fait des animaux**², celui de la **responsabilité du fait d'autrui**³, ou, encore, celui, traditionnel⁴, 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 Cohen⁵, **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 physiquement⁶.

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 de façon autonome, qui est son gardien ? Le concepteur de son intelligence artificielle ou le propriétaire qui a réalisé son apprentissage ?

¹ 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.

² Pour les cas où un individu est déclaré responsable des actes dommageables commis par un animal dont il a la garde ou la propriété.

³ Pour les cas où un individu est déclaré responsable des actes dommageables commis par un tiers.

⁴ Issu du fameux ancien article 1384 du code civil.

⁵ Article « Le droit à l'épreuve de l'intelligence artificielle » du 28 novembre 2016 paru dans la revue *Village de la Justice*.

⁶ 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.

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 responsabilités** », 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-up *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 (FFA)** interrogée par vos rapporteurs, l'intelligence artificielle est un sujet naissant sur lequel **les impacts et les solutions 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 et le fait d'apporter suffisamment de protection aux consommateurs**. De nouvelles questions vont émerger avec ces nouvelles technologies, et notamment certaines concernant l'assurance et ses régimes.

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 cassation¹, 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*

¹ 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

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 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**¹. 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/CE², 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 ;

¹ 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>

² 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/LexUriSero/LexUriServ.do?uri=OJ:L:2006:157:0024:0086:fr:PDF>

- 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 ;
- 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'emballage¹ ou d'un employeur ayant fait travailler un salarié sur une ligne de fabrication robotisée sans prendre les mesures de sécurité nécessaires².

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 examiné 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

¹ Arrêt n°02-87666 du 30 septembre 2003 de la chambre criminelle de la Cour de cassation.

² Arrêt n°01-21192 du 16 septembre 2003 de la deuxième chambre civile de la Cour de cassation.

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.

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 et de ses gains en termes de sécurité ou d'environnement 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 y a donc lieu pour vos rapporteurs, et, plus largement, l'OPECST, de suivre cette mission sans interférer avec elle.

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 à 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* »¹ 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 auto-protectrice** 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, **devant 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.

¹ 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>

Le laboratoire du MIT « *moral machine* », visité par vos rapporteurs, travaille notamment sur les **dilemmes éthiques concernant les voitures autonomes**. Le dilemme du tramway inventé par Philippa Foot a été réaffiné et testé. Les résultats provisoires des tests conduisent cette équipe 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 plutôt qu'un être humain, le fait de sacrifier en priorité une personne plus âgée plutôt qu'une personne plus jeune et *a fortiori* un enfant, le fait de sacrifier en priorité un homme plutôt qu'une femme... Ces hypothèses soulèvent néanmoins la question du cadre éthique dans lequel situer les voitures autonomes.

Le 6 février 2017 a été donné le coup d'envoi d'un **programme européen de trois ans baptisé « Autopilot »**. Pour la France, **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 apparaît nécessaire de se doter à terme, sur un plan global dans tous les États membres de l'UE (voire au-delà), 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 et d'**accompagner ses usages d'une réflexion éthique**.

Comme il a été vu, la science-fiction, avec Isaac Asimov, a envisagé 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 soutiennent 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 d'Inria. 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 CERNA¹. Il convient de relever que le COMETS s'est, à plusieurs reprises, penché sur les problèmes éthiques posés par les STIC².

Le rôle de la CERNA en matière de réflexion éthique sur le numérique est remis en perspective³ avec les **nouvelles missions dévolues à**

¹ Rapport du COMETS sur l'éthique des STIC : <http://www.cnrs.fr/comets/IMG/pdf/08-rapportcomets091112-2.pdf>

² 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), 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).

³ Selon Max Dauchet, président de la CERNA, cette dernière a « pour mission de mener une réflexion éthique au titre du monde de la recherche. La proximité dans le numérique étant grande entre la recherche et les usages, la CERNA a milité activement pour la création par la loi pour une République numérique d'un dispositif traitant plus largement des questions

la CNIL, après la loi du 7 octobre 2016 pour une République numérique¹. Aux termes de cette loi, la CNIL, qui se définit comme l'autorité française de contrôle en matière de protection des données personnelles², 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 créée³. 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 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 plateforme⁴ 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 posées⁵.

éthiques et sociétales, comme le CCNE le fait dans le secteur de la vie et de la santé. La création d'un tel dispositif que la CERNA appelait de ses vœux, ayant été confié à la CNIL, elle doit conduire à conforter la CERNA comme coordinateur du monde de la recherche au sein du débat de société sur ces questions ». Il ajoute que « le rôle de la CERNA est à positionner suite aux nouvelles missions dévolues à la CNIL, sous l'angle des rapports entre la recherche dans le numérique et la société ».

¹ Loi n° 2016-1321 du 7 octobre 2016 pour une République numérique.

² 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.

³ Cf. <https://www.cnil.fr/fr/ethique-et-numerique>

⁴ Cf. <https://www.cnil.fr/fr/webform/contacter-lequipe-de-la-mission-ethique-et-numerique>

⁵ 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 » ?

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 qui représente le monde de la recherche, et celui de la CNIL, chargée de la dimension sociétale de la réflexion éthique, sont **nécessaires** et pourraient être formalisées. À ce stade, vos rapporteurs relèvent que les manifestations envisagées¹ 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 décembre 2016 un rapport au ministre de l'économie et des finances portant sur les « Modalités de régulation des algorithmes de traitement des contenus »². 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 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 appellent aussi au **développement de bonnes pratiques** dans les différents secteurs économiques. Ils soulignent par ailleurs 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 la **plateforme scientifique collaborative** « **TransAlgo** »³, lancée en janvier 2017, portée par Inria et placée sous la direction de Nozha Boujemaa, travaille de manière utile au **développement de pratiques transparentes et responsables dans le traitement algorithmique des données**⁴. Dans la notion de « responsabilité », sont considérés à la fois le respect des règles juridiques et le

¹ Cf. <https://www.cnil.fr/fr/les-partenaires-et-evenements>

² <http://www.economie.gouv.fr/cge/modalites-regulation-des-algorithmes-traitement-des-contenus>

³ http://www.economie.gouv.fr/files/files/PDF/Inria_Plateforme_TransAlgo2016-12of.pdf

⁴ <https://www.inria.fr/actualite/actualites-inria/transalgo>

respect des règles éthiques. Il s'agit de répondre aux préoccupations exprimées d'**explicabilité, de loyauté, de neutralité, de non-discrimination des algorithmes mais aussi des biais des données et des algorithmes**. Plus récemment, un **Institut « Convergence i2-DRIVE »** (son nom vient de « Interdisciplinary Institute for Data Research : Intelligence, Value and Ethics ») porté par l'Université Paris-Saclay et lui aussi piloté par Nozha Boujemaa - incluant 14 partenaires académiques et bénéficiant de soutiens industriels - a été accepté en avril 2017 dans le cadre de l'appel à projet du PIA 2 et comporte un programme de travail sur dix ans. Il vise la conjonction des expertises en sciences du numérique et en sciences humaines et sociales pour l'innovation en matière de recherche et de formation dans le domaine des sciences des données et de l'intelligence artificielle avec leurs enjeux socio-économiques et éthiques. Cette conjonction annonce aussi des perspectives fécondes en matière de transfert et d'innovation par la vision systémique que proposera i2-DRIVE pour adresser les défis technologiques et éthiques de la transformation numérique à l'ère des Big Data et des systèmes cognitifs.

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 déplacent les frontières d'utilisation, d'exploitation et d'usage, ce qui pose de nouvelles difficultés¹. 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 le chercheur Hiroshi Ishiguro au Japon. **Les interrogations éthiques sur les finalités d'un tel projet d'intelligence artificielle humanoïde et sur ses effets s'imposent**. La ressemblance avec l'humain renvoie à l'hypothèse de la « Vallée de l'étrange », introduite par Masahito Mori en 1970. Cette sorte de malaise ressenti par les êtres humains face à des entités presque semblables à eux, mais pas au point de s'y tromper a été présentée précédemment de manière plus détaillée par vos rapporteurs.

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 :

¹ 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 avril 2017, ses préconisations sur la *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 »¹ 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 Avocats².

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-up, 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 pouvoir 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*³, et l'éthique *by design*⁴, ce qui sera plus facile pour des jeunes et petites

¹<http://cigref.fr/Publication/2014-CIGREF-Ethique-et-Numerique-une-ethique-a-reinventer-Rapport-mission-F-FISCHER.pdf>

²<http://www.cigref.fr/wp/wp-content/uploads/2016/09/Gouvernance-IA-CIGREF-LEXING-2016.pdf>

³ Respect de la vie privée dès le stade de la conception du produit.

⁴ Respect de règles éthiques dès le stade de la conception du produit.

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*.

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 »** à venir. 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 paraître 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 disponibles¹. Coordonné par Grégory Bonnet (GREYC - Normandie Université), il associe différents partenaires² 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

¹ <https://ethicaa.org>

² 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).

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 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 résultats intermédiaires des 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 a 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 propriétés éthiques ayant pour but de vérifier *a posteriori* si la spécification d'un système multi-agents 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 que 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 d'aboutir à une preuve de concept (« *proof of 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 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 rôles 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)¹, 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, après un appel à projets lancé en 2015, pas moins de 37 projets de recherche destinés à prévenir les risques liés à l'intelligence artificielle². 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 expert³ à s'exprimer devant quelques

¹ L'Institut a été fondé en mars 2014 par Max Tegmark cosmologiste au MIT, Jaan Tallinn cofondateur 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.

² 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.

³ 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.

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, 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 Russel¹, 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**². La responsable des affaires publiques d'Apple a alors expliqué à vos rapporteurs **que les valeurs spécifiques à**

¹ 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).

² 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/>

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. 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*¹ 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.

¹ L'intégralité du rapport publié par l'Université Stanford est disponible ici : https://ai100.stanford.edu/sites/default/files/ai_100_report_0831fnl.pdf

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.

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 qu'ils puissent 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.**

À la suite de ce rapport de septembre 2016 et puisqu'il n'existait **aucun guide commun encadrant le domaine de l'intelligence artificielle,** édictant notamment 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 2017¹, 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* » constituent 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* »² 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 intelligences artificielles 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.

¹ Son programme figure ici : <https://futureoflife.org/bai-2017/>

² <https://futureoflife.org/ai-principles/>

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.

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 utilisation 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 consacré 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**.

En mars 2017, la fondation *Future Society* de la **Harvard Kennedy School** a rendu un rapport « **La Révolution de l'intelligence artificielle au service de tous** » dans le cadre de son initiative pour l'intelligence artificielle¹. Coordonné par Cyrus Hodes et Nicolas Mialhe, il décrit l'émergence de ces technologies, en souligne les opportunités et les défis en matière de politiques publiques puis formule quelques recommandations.

Vos rapporteurs résument ci-après le contenu des trois dernières parties de ce rapport.

1- Pour ce qui concerne tout d'abord les **opportunités économiques et sociales offertes par l'intelligence artificielle**, il faut observer que le développement de l'IA pourrait se traduire par une révolution de la gouvernance, à travers **l'efficacité des processus décisionnels** pour les acteurs publics et privés. La capacité des algorithmes d'apprentissage machine à exploiter les stocks et les flux croissants de données est susceptible de déclencher une **vague d'optimisation** dans de nombreux domaines, à l'instar de l'énergie et des transports. L'intelligence artificielle peut être essentiellement utilisée comme une « **technologie de prédiction** », dont la diffusion pourrait réduire considérablement le coût du traitement des données historiques et donc faire des prédictions pour un large éventail de tâches telles que le profil des risques, la gestion des stocks ou la prévision de la demande. Le développement de l'intelligence artificielle annonce des formes de **collaboration et de complémentarité** nouvelles et économiquement plus efficaces **entre les humains et les machines**.

¹ Cf. le site : <http://ai-initiative.org/>

L'intelligence artificielle peut potentiellement être considérée comme un **nouveau facteur de production**, en améliorant l'efficacité des facteurs traditionnels du travail et du capital et en créant un hybride capable de créer des effectifs entièrement nouveaux. Dans de nombreux cas, l'intelligence artificielle sera capable de s'auto-améliorer, et de surpasser les humains en termes d'échelle et de vitesse.

Dans le domaine de la santé, les progrès réalisés dans l'intelligence artificielle créent ce que beaucoup considèrent comme des **technologies véritablement transformatrices et évolutives** pour aborder et traiter les maladies et les médicaments. Tous les principaux acteurs de l'intelligence artificielle entrent sur le marché de la santé, et l'intelligence artificielle est perçue par les praticiens comme un facilitateur de la **révolution médicale de précision et de prévention**, ainsi que dans le séquençage génomique. Le recours à des technologies d'intelligence artificielle dans le domaine de la santé peut également inclure la découverte et la création de molécules et de médicaments, la personnalisation des soins, l'utilisation de « bots » médicaux, ou encore les soins apportés aux personnes âgées.

L'intelligence artificielle a déjà un impact conséquent sur le transport avec l'introduction de capacités de conduite autonome. Le développement de réseaux neuronaux profonds est l'un des principaux moteurs des progrès impressionnants réalisés dans les véhicules autonomes au cours de la dernière décennie. **La révolution des formes de mobilité transformera en profondeur l'industrie automobile**, qui pourra muter en industrie de services, et avoir un impact sur les paysages urbains et suburbains, avec la possibilité que cela puisse traiter la ségrégation spatiale qui a souvent été accompagnée de la marginalisation sociale.

Bien que l'éducation soit, à court terme, un secteur moins impacté par la substitution de l'emploi à l'intelligence artificielle, avec un potentiel d'automatisation estimé à 27%, les outils que l'intelligence artificielle apporte aux éducateurs et aux étudiants se révéleront très précieux en termes d'efficacité. À mesure que les MOOC (*Massive Open Online Courses*) et les SPOC (*Small Private Online Courses*) gagnent en popularité en donnant accès au **meilleur contenu de cours pour tous**, alors que le *crowdsourcing* et l'apprentissage machine sont développés pour assurer notamment les tâches d'enseignement et d'évaluation.

En analysant et comprenant les normes et les variations des comportements des utilisateurs, l'intelligence artificielle s'est avérée être un outil efficace adoptant des **mesures proactives** (contrer les cyberattaques, le vol d'identité, la lutte contre le terrorisme...). Au-delà de la cybersécurité, l'intelligence artificielle prend de plus en plus de poids dans la défense et la sécurité. Elle permet au personnel de sécurité nationale d'être **mieux informé** grâce à des systèmes de gestion et des outils de simulation plus intelligents.

Dans le domaine de la robotique, les progrès de l'intelligence artificielle influent radicalement sur **les tactiques et les capacités de mission des plates-formes autonomes** qui sont en mesure de façonner et d'influencer la manière dont les conflits et les opérations de sécurité se déroulent à l'avenir.

Dans la police, l'intelligence artificielle est utilisée comme un **outil d'identification puissante**, et de plus en plus comme un instrument prédictif indiquant où et quand les crimes peuvent se produire. Comme dans la défense, ces applications devront être très sérieusement régies pour éviter les abus, ce qui pourrait avoir des effets dévastateurs.

2- Les défis pour les politiques publiques sont grands. En raison de sa dynamique d'entreprise, la montée en puissance de l'intelligence artificielle a été caractérisée par la formation de ce que l'on pourrait appeler un «**oligopole global dissymétrique**» qui soulève des problèmes importants en termes de répartition de la richesse et de la puissance, mais pourrait également étouffer l'innovation en empêchant l'apparition de nouveaux acteurs. Le marché de l'intelligence artificielle est dominé par quelques multinationales américaines (GAFAMITIS) et chinoises (BATX), dont la capitalisation boursière représente aujourd'hui plus de 3,3 mille milliards de dollars. Le marché de l'intelligence artificielle présente de fortes caractéristiques «*the winner takes most*» en raison de la prévalence particulière d'effets de réseau et d'effets d'échelle. Cet oligopole est dissymétrique en cela qu'il présente de **sérieux déséquilibres** entre, d'une part, les multinationales numériques hautement innovantes dont les modèles économiques perturbateurs se déroulent à l'échelle transnationale et, d'autre part, les États-nations. Cette dissymétrie qui concerne la révolution numérique en général est encore aggravée par le développement de l'intelligence artificielle puisque les multinationales sont maintenant en mesure de capturer et de fournir des fonctions clés de gouvernance. L'«**oligopole global dissymétrique**» de l'intelligence artificielle pourrait être une source de tensions sévères au cours des prochaines décennies dans les pays et entre eux. Sans régulation et redistribution, **l'accumulation déséquilibrée de richesse et de pouvoir** entre les mains de quelques acteurs privés pourrait **favoriser la montée en puissance des populismes** et des extrémismes à l'échelle mondiale.

Un autre élément crucial des défis politiques associés au développement de l'intelligence artificielle se rapporte à la délégation croissante de compétences aux agents autonomes. Le potentiel de l'intelligence artificielle pour la croissance, le développement et le bien public exigera l'adoption de normes techniques et de mécanismes de gouvernance qui maximisent la libre circulation des données et des investissements dans des services à forte intensité de données.

En outre, il pourrait y avoir des tensions entre la **nécessité d'une réglementation rigoureuse des données** afin de garantir la protection de la vie privée, le consentement éclairé et la lutte contre la discrimination, d'une

part, et les **conditions de marché nécessaires** pour **favoriser l'innovation et développer des bases technoindustrielles** puissantes d'autre part.

Certains chercheurs ont exprimé de graves inquiétudes quant au risque d'amplification des biais sociaux présenté par l'apprentissage machine, les algorithmes pouvant devenir ainsi source de discrimination, comme l'a montré l'expérience du bot conversationnel « Tay » développé par Microsoft.

L'ensemble de défis politiques majeurs associés au développement de l'intelligence artificielle se réfère à l'impact de l'automatisation sur les emplois et les inégalités, certains chercheurs craignant un **potentiel «décrochage» des classes moyennes**. De nombreux experts estiment que la vague d'automatisation alimentée par **l'intelligence artificielle influencera profondément les profils d'emplois**. Les résultats des études sur l'impact de l'automatisation du travail menées au cours des cinq dernières années diffèrent cependant radicalement dans leur évaluation et leurs projections (comme il a été vu le rapport de l'OCDE publié en juin 2016 a conclu qu'une moyenne modeste de 9 % des tâches est automatisable, alors qu'une étude de 2013 de Frey et Osborne se voulait plus alarmiste avec sa conclusion qu'environ 47 % des emplois américains seraient sensibles à l'automatisation au cours des deux prochaines décennies).

La capacité des sociétés à façonner la révolution de l'intelligence artificielle dans une **destruction créatrice** et à **diffuser son bénéfice à tous** dépend principalement de la façon dont elle réagit collectivement. Des réponses politiques systémiques seront nécessaires, y compris la réforme et la réinvention éventuelle de la sécurité sociale et de la taxe de redistribution. Les systèmes d'éducation et de développement des compétences devront également être réformés pour permettre des **transitions professionnelles viables**. Compte tenu de la difficulté à prédire les zones les plus impactées et de désagréger l'automatisation axée sur l'intelligence artificielle d'autres facteurs, les réponses politiques devront d'abord viser l'ensemble de l'économie, jusqu'à ce que les stratégies ciblées deviennent plus efficaces et que les pratiques de suivi et d'évaluation aient été conçues.

3. En matière de recommandations, le rapport « La Révolution de l'intelligence artificielle au service de tous » formule huit propositions :

1 : Les institutions privées et publiques devraient être encouragées à examiner si et comment elles peuvent **influencer de manière responsable l'intelligence artificielle dans un sens bénéfique pour la société**.

2 : Les États et les organisations internationales devraient travailler sur le développement de **mécanismes de coordination mondiale** pour surveiller collectivement l'état de l'intelligence artificielle.

3 : Les États et les organisations internationales devraient soutenir davantage **la recherche et la prévoyance** sur la dynamique (vitesse et magnitude) du développement de l'intelligence artificielle.

4 : Les États et les organisations internationales devraient soutenir le développement de **modèles prédictifs analysant l'impact des machines autonomes sur les marchés du travail**, les emplois ou tâches futurs, les mécanismes économiques, le bien-être collectif et les structures de risque et les compétences associées.

5 : Les États et les organisations internationales devraient soutenir davantage de **recherche en sciences sociales** sur l'interaction entre la science, la technologie et la société à l'époque de l'intelligence artificielle.

6 : Les États et les organisations internationales devraient soutenir davantage de **recherches sur le renforcement des protections juridiques contre les biais algorithmiques**.

7 : Les États et les organisations internationales devraient s'efforcer de créer un **cadre politique mondial** et des normes harmonisées régissant les **flux et les stocks de données**, en particulier les données personnelles, ainsi que **l'utilisation d'algorithmes**.

8 : Les institutions privées et publiques ainsi que les organisations professionnelles devraient être encouragées à **diriger de manière proactive les débats publics** en engageant les experts de l'IA, les praticiens, les parties prenantes et la société en général sur le thème de la gouvernance de l'intelligence artificielle.

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 à 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 ci-dessous.

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 d'activité. Les pratiques industrielles doivent être davantage marquées par une culture éthique et la communauté concernée doit s'emparer des sujets idoines et assumer sa responsabilité éthique. Du fait du mode de fonctionnement et de prise de décision des algorithmes, il est nécessaire d'inclure des composants de type « boîtes noires » décryptables *a posteriori* 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 relatif au 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 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 de garantir 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 pourraient 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 critiqués, et la légitimation de leur développement pourrait potentiellement créer des précédents, qui du point de vue géopolitique pourraient être 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 aujourd'hui d'adopter des règles éthiques clairement définies.

8. Les problèmes économiques et humanitaires

L'objectif de ce rapport, en matière économique et social, est d'identifier les principaux moteurs de l'écosystème mondial des technologies dans ce domaine et de prendre en compte les ramifications économiques et humaines, voire humanitaires, afin de suggérer des opportunités clés de solutions. Ces dernières pourraient être mises en œuvre de manière à

débloquer 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 », qui n'est pas encore advenu et n'est pas une réalité envisageable à court terme.

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, de nombreuses craintes émergent, notamment concernant l'emploi, les changements technologiques évoluant trop rapidement pour permettre aux méthodes de formation de la main-d'œuvre de s'adapter. Par ailleurs, l'accès aux technologies d'intelligence artificielle n'est pas équitablement réparti, entraînant un manque de compréhension des informations par une 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 actuellement 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 du temps, 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 **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.**

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 2015¹, 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.**

L'**Alliance sciences sociétés (ALLISS)** se donne justement pour première mission d'animer le débat public², mais aussi de renforcer les capacités d'initiatives des acteurs de la société civile, de soutenir les initiatives des institutions de l'enseignement supérieur et de recherche, de

¹ 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.

² Cf. la charte de l'Alliance : https://uploads.strikinglycdn.com/files/5d463c19-d192-418c-9987-9d3a62b95a22/CHARTE_ALLISS.pdf

favoriser le croisement des savoirs académiques, d'action et d'expérience et, enfin, d'accompagner les collectivités territoriales et leurs politiques publiques. En collaboration avec l'OPECST, l'IHESST et le ministère de l'Enseignement supérieur et de la recherche, ALLISS a organisé en novembre 2016 la session inaugurale de lancement de son livre blanc « Prendre au sérieux la société de la connaissance »¹. Fruit de quatre années de travail, riche de la participation de plus de 1 500 personnes et de 450 organisations et de la contributions de plus de 150 organisations de la société civile, ce **livre blanc a été rendu public en mars 2017**, avec l'idée de mettre en lumière des expériences exemplaires, de contribuer à un diagnostic partagé et d'effectuer des recommandations pour les acteurs publics et les acteurs impliqués dans l'enseignement supérieur et la recherche : établissements d'enseignement supérieur et de recherche mais aussi acteurs de la société civile, associations, syndicats, entreprises... Les recommandations du document portent sur les interactions entre société civile, enseignement supérieur et recherche, dans le but d'aboutir entre autres à une meilleure intégration de la société civile dans les choix des politiques publiques d'innovation.

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.

Le rapport « **Vers une société apprenante** »² remis en mars 2017 par François Taddéi, directeur du Centre de recherches interdisciplinaires (CRI), à la ministre de l'Éducation nationale, de l'enseignement supérieur et de la recherche va dans ce sens et propose une approche systémique interdisciplinaire marquée par « *une culture de la confiance, de la liberté et du mentorat bienveillant* », avec le numérique comme catalyseur des évolutions afin de mobiliser les moyens matériels et humains pour répondre aux défis éducatifs.

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 scientifiques³, peut être cité de manière significative. Les réseaux sociaux de chercheurs⁴ sont une autre illustration.

¹ Le livre blanc est téléchargeable sur le site d'ALLISS : <http://www.alliss.org/>

² Cf. http://cache.media.education.gouv.fr/file/03_-mars/19/0/2017_rapport_taddei_740190.pdf

³ Cf. <https://pubpeer.com/>

⁴ 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 numérique¹, 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 termes 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**

¹ Rapport « Sécurité numérique et risques : enjeux et chances pour les entreprises », n° 271 (2014-2015).

militaire en raison d'une dérive de systèmes d'intelligence artificielle, dans le cas d'une hypothèse « faible » de dérive malveillante dans le sens où 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 le soit. Un article devrait paraître en 2017 dans la revue de la Défense Nationale à la suite de cette étude.

En robotique il est **nécessaire de toujours pouvoir arrêter un système, la question peut se poser plus généralement 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 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**¹.

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 Aron², **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

¹ Morgane Tual, « Pourquoi Google a conçu un bouton rouge pour désactiver des intelligences artificielles », *Le Monde*, 7 juin 2016.

² Il utilisait l'expression de « Paix impossible, guerre improbable ».

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.**

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 tant qu'elles ne sont pas corrigées. 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 le précise 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 travers des réseaux de neurones artificiels en couches multiples, fonctionnent, ou, pour être plus précis, donnent, dans un certain nombre de domaines, d'excellents résultats et très rapidement.

Ce problème d'opacité reste entièrement à résoudre. On parle à ce sujet 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 décembre 2016 un rapport au ministre de l'économie et des finances portant sur les « Modalités de régulation des algorithmes de traitement des

contenus »¹. Les algorithmes de traitement des contenus sont inséparables des données qu'ils traitent et des plateformes qui les utilisent pour proposer un service. Alors qu'il y a de **nombreux travaux sur la protection des données et sur la loyauté des plateformes, il en existe encore peu sur les algorithmes eux-mêmes**. Ce rapport souligne donc qu'il faut **développer la capacité à tester et contrôler les algorithmes**².

Vos rapporteurs rappellent à nouveau ici qu'Inria avec sa **plateforme scientifique collaborative « Transalgo »**, lancée en 2017 et 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**³, afin de répondre aux préoccupations exprimées de **régulation et d'explicabilité des algorithmes** par l'élaboration d'outils et de pratiques transparentes et responsables dans le traitement algorithmique des données. Une telle démarche, en lien avec le *Data Transparency Lab* du MIT, Mozilla et Telefonica, va dans le bon sens mais gagnerait à voir sa force de frappe **démultipliée par une** dynamique de **mobilisation de plusieurs équipes de recherche** autour de challenges scientifiques nouveaux. Inria ne doit pas être la seule structure en France à porter des projets 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 transversales. Inria explique ainsi : « *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* ».

¹ <http://www.economie.gouv.fr/cge/modalites-regulation-des-algorithmes-traitement-des-contenus>

² 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 préconisent **la création d'une plateforme collaborative scientifique**, destinée à favoriser le développement d'outils logiciels et de méthodes de test d'algorithmes, ainsi que de promouvoir l'utilisation de ces outils et méthodes. 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.

³ Cf. http://www.economie.gouv.fr/files/files/PDF/Inria_Plateforme_TransAlgo2016-12vf.pdf

Comme il a été mentionné précédemment, l'Institut Convergence i2-DRIVE (Interdisciplinary Institute for Data Research : Intelligence, Value and Ethics) s'est donné parmi ses objectifs de se consacrer aux défis scientifiques des algorithmes d'apprentissage et de traitement des données mais aussi à leurs impacts socio-économiques.

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 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 aussi 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 et Google.

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**¹. La

¹ 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

singularité représente un tournant hypothétique supérieur dans l'évolution technologique, dont l'intelligence artificielle serait le ressort principal. Vernor Vinge a rédigé un essai remarqué à ce sujet en 1993¹.

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* », « *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 à 2020** et qu'elles le **dépasseront en 2045**². 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 à 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,*

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.

¹ Cf. Vernor Vinge, *The Coming Technological Singularity*.

² 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.

l'Odyssée de l'espace »¹. 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 dès la fondation de l'intelligence artificielle, mais elle correspond surtout à une invention conceptuelle des années 1980. L'explication donnée par Marvin Minsky, l'un des pères de l'intelligence artificielle, 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. »²

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 Bostrom³ « 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

¹ « From Socrates to Expert Systems : The Limits and Dangers of Calculative Rationality », 2004.

² Marvin Minsky, « It's 2001. Where Is HAL? », *Dr. Dobb's Technetcast*, 2001.

³ Vos rapporteurs les ont rencontrés tous les deux, le second est auteur de « Superintelligence : Paths, Dangers, Strategy », 2014.

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, qui est à l'origine d'AlphaGo, il est impossible 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 eux-mêmes. 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 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 Searle¹. Cette expérience vise à montrer qu'une **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 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 calculette 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*

¹ Il pose l'hypothèse d'une personne qui n'a aucune connaissance du chinois, enfermée dans une chambre, et disposant d'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 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.

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.

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*¹ »), 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.

¹ En ergonomie ou en informatique, l'*affordance* est la capacité d'un objet à suggérer sa propre utilisation.

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 homme-machine 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 technologies Gérard Sabah, **il est difficile et risqué d'établir des prévisions détaillées sur l'avenir à court, moyen et long termes 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*¹. 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 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 devant 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.**

¹ http://www.wtec.org/ConvergingTechnologies/Report/NBIC_report.pdf

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 1951¹. Julian Huxley² 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 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. En 2013, Max More et Natasha Vita-More ont publié un essai particulièrement riche à ce sujet³.

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.**

¹ 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.

² 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.

³ Cf. Max More et Natasha Vita-More, *The Transhumanist Reader : Classical and Contemporary Essays on the Science, Technology, and Philosophy of the Human Future*.

TROISIÈME PARTIE : LES PROPOSITIONS DE VOS RAPPORTEURS

Les propositions présentées ici n'ont **pas vocation à couvrir l'ampleur des questions posées par la recherche en intelligence artificielle**, ainsi que l'affirmait Jean de la Fontaine, « *Loin d'épuiser une matière, On n'en doit prendre que la fleur* ». Il s'agit plus de recommandations sous forme de contributions à la réflexion collective et vos rapporteurs n'ont pas entendu répondre en totalité aux défis posés pour la France et l'Europe par les **enjeux industriels** de l'intelligence artificielle, par **l'automatisation de certaines tâches** et par **l'économie globalisée de plateformes** qui se dessine avec l'émergence d'entreprises privées puissantes basées aux États-Unis ou en Chine. Ces dimensions, évoquées dans le rapport, ne sont pas le cœur de métier de l'OPECST et ne se traduisent donc que partiellement dans les propositions formulées ci-après.

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

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 et promeuvent. 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.

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**.

L'**approche européenne**, voire **internationale** est importante à ce niveau. L'initiative IEEE présentée précédemment va dans ce sens, en particulier en ce qui concerne les problèmes liés aux **standards, normes et certifications**.

Les considérations de **sécurité** et de robustesse évoquées dans le présent rapport amènent à conclure qu'il est pertinent de poursuivre, en général, la réflexion sur la sécurité numérique et, en particulier, de prévoir des mécanismes d'arrêt d'urgence. **L'interruption d'un système d'intelligence artificielle ou d'un robot doit 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, surtout que l'arrêt brutal d'un système peut avoir des conséquences graves et imprévisibles¹.

Il faut rechercher la **transparence des algorithmes** contre les boîtes noires, les biais et les discriminations.

¹ Dans ce cas, l'opérateur est conduit à faire un choix rapide entre deux types de conséquences négatives, celles liées aux dysfonctionnements du système d'IA ou celles liées aux conséquences de l'arrêt du système.

Il convient de prévoir aussi des mécanismes de traçabilité, du type **journaux de bord** ou *loggings*¹, 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 être en mesure de **remonter aux derniers circuits de décision et à 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 d'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 d'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 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. En face de systèmes complexes, des « **systèmes de systèmes** » représentent des perspectives fondamentales en vue de la maîtrise des machines.

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**, quoi qu'il en soit.

¹ *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).*

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)**.

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 des algorithmes, 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 formation 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.

La **diffusion de ces connaissances** sur l'éthique de l'intelligence artificielle et de la robotique **vers les autres étudiants, voire le grand public**, pourrait être recherchée.

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 pourra se préoccuper de **l'acceptabilité individuelle et sociale** de ces technologies et étudier les **effets secondaires** imprévus et/ou indésirables des différents outils de l'intelligence artificielle.

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 de 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 gagnerait à être élargi au-delà de l'Inria, comme il a été vu dans le rapport. 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**.

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 entreprises devront envoyer plus systématiquement leurs salariés en formation sur leur temps de travail, de manière à actualiser leurs connaissances et leurs compétences, par exemple tous les cinq ans. Le

Compte personnel d'activité (CPA) peut être un outil mobilisable à cette fin et le service « **Formation Tout au Long de la Vie** » (FTLV), guichet unique pour l'usager comme pour l'entreprise, permet de concrétiser le droit à la formation continue, tant pour les démarches d'orientation, de bilan, d'accompagnement vers l'emploi, que pour les actions de formation et de validation des acquis de l'expérience.

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

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 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**.

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'hyperspécialisation 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 pourraient orienter le futur PIA dans le sens de cette hyperspécialisation, ce qui est de nature à les inquiéter.

Le mode de financement des projets pourrait de plus 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 ainsi le cas aux États-Unis pour la plupart des projets soutenus par la Fondation nationale pour la science, visitée par vos rapporteurs à Washington (National Science Foundation ou NSF), mais encore du projet *Todai Robot*¹ conduit depuis 2011 pour dix ans à l'Université de Tokyo au Japon, avec le soutien de l'Institut national de l'informatique japonais et l'entreprise Fujitsu.

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 liaison 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**², 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 Inria de Saclay, l'effort financier nécessaire serait de l'ordre de **100 millions d'euros par an sur dix ans**, financements publics et privés compris.

De très nombreuses coopérations public-privé existent et fonctionnent et il est loisible de s'en inspirer (Microsoft-Inria, CNRS avec IBM, Rhodia, Thalès, Michelin...)

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*

¹ En novembre 2016, ce robot a échoué pour la quatrième année consécutive à intégrer la prestigieuse Université de Tokyo. Ses concepteurs tirent à chaque rentrée scolaire les leçons de cet échec pour faire progresser la machine.

² 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.

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.

Proposition n° 7 : Encourager la constitution de champions européens en intelligence artificielle et en robotique, tout en poursuivant le soutien aux PME spécialisées, en particulier les start-up

Il faut **encourager la constitution de champions européens en intelligence artificielle et en robotique**, sans retomber dans les erreurs du projet Quaero¹ mais plutôt en suivant les traces des succès du modèle d'Airbus, ce qui pourra impliquer de poser la question du droit de la concurrence dans l'Union européenne, très contraignant en matière de concentrations. Cette proposition s'ajoute à celle de poursuivre les mesures de soutien aux PME spécialisées, en particulier les start-up.

Faire émerger des champions européens et soutenir notre tissu de PME passe aussi par un écosystème rendu encore plus favorable grâce à la mise à dispositions de très grands volumes de données (vue au paragraphe précédent), l'accès facilité à des supercalculateurs, dont les puissances de traitement pourraient être mises en réseau sur le *cloud*, l'harmonisation du

¹ Le projet, imaginé par la France en 2004 en tant qu'un des cinq premiers PMII (programmes mobilisateurs pour l'innovation industrielle), lancé en 2008 avec le soutien de l'Allemagne et abandonné en 2013, visait à développer des « outils intégrés de gestion des contenus multimédias », dont des extensions multimédias pour des moteurs de recherche de nouvelle génération qui devaient permettre de rechercher par le contenu non seulement du texte, mais aussi des images, du son et de la vidéo. L'ambition de constituer les fondements technologiques européens de futurs moteurs de recherche a donc rapidement avorté, l'américain Google demeurant dans un quasi-monopole (en dépit des qualités éthiques du moteur de recherche français « Qwant »). Le projet Quaero s'est donc soldé par un échec mais il n'est pas nécessaire d'espérer pour entreprendre ni de réussir pour persévérer, cf. <http://www.quaero.org/>

droit applicable, par exemple en matière de protection des données personnelles dans l'Union européenne.

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, mais aussi à la croissance des entreprises françaises de ce secteur, en particulier de ses start-up, doit être étudiée, afin de faire émerger une industrie française de l'intelligence artificielle qui développera des produits innovants compétitifs et exportables.** Là aussi, en plus de l'encouragement et de la facilitation de l'accès aux dispositifs de soutiens, des protections pourraient être utiles.

Comme l'affirment justement Thierry de Montbrial et Thomas Gomart, dans un ouvrage qu'ils viennent de publier en février 2017, nous devons défendre « *notre intérêt national* » et vos rapporteurs y ajoutent « *notre intérêt européen* ».

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...** De manière caricaturale, la recherche appliquée en intelligence artificielle 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.

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, le cas échéant selon un processus itératif.

Il sera possible de s'appuyer sur l'expérience des 138 cours spécifiques en intelligence artificielle, 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, appuyés sur l'excellence d'Inria, du LAAS de Toulouse, de l'ENS, de l'Institut Mines-Télécom... Ils ne seraient pas nécessairement géographiquement localisés, 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 à travers des réseaux d'excellence en IA.**

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.

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**, à l'image de celui présenté à vos rapporteurs au siège de l'organisme en matière de cybersécurité, 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** et contribuer à l'excellence de la recherche en France.

Les « **hackathons** »¹ qui sont à la fois le principe, le moment et le lieu d'événements dans lesquels des développeurs se réunissent pour faire de la programmation informatique collaborative, sont des pistes à creuser pour assurer une telle promotion de la recherche et de l'innovation en intelligence artificielle.

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° 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

¹ Ce mot valise a été formé sur la base de « hack » (activité de manipulation d'un système informatique) et de « marathon ». La référence au « marathon » provient du fait que le travail des développeurs est souvent conduit sur plusieurs jours et sans interruption.

plus globalement pour tout le secteur de l'informatique et pour les filières scientifiques en général, ne suffit pas à justifier la situation présente. Il s'agit d'ailleurs d'un axe de travail du CNSTI.

Dans le secteur de l'informatique, la vigilance et la prise en compte des **problématiques de biais et de discriminations dans les algorithmes et les données** seront d'autant plus grandes que les acteurs du domaine seront issus d'une plus grande diversité.

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 ainsi que d'accroître leurs effets bénéfiques pour la société, pourraient participer à l'effort collectif nécessaire.

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

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**¹ 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.

¹ 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 . 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

Il peut être noté qu'un **plan numérique pour l'éducation** a été lancé par le Président de la République en mai 2015¹, 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**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 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 ».

¹ 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://colenumerique.education.gouv.fr/plan-numerique-pour-l-education/>

² Cf. <http://www.education.gouv.fr/cid89179/projet-de-programme-pour-un-enseignement-d-exploration-d-informatique-et-de-creation-numerique.html>

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*¹ 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 électronique² 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 »³, diffusée par la chaîne No Life, ou 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

¹ *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*

² Voir le site officiel du salon : <http://www.ces.tech/>

³ L'émission « La Faute à l'algo » est écrite et réalisée par Michel Blockelet et Jill-Jênn 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>

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 »¹ 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.

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.

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. Les pistes à explorer pourraient s'inspirer de régimes législatifs ou déontologiques existants.

La **vérification des publicités**, en lien avec l'Autorité de régulation professionnelle de la publicité (ARPP, ex-Bureau de vérification de la

¹ 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.

publicité), serait par exemple un premier pas vers une plus grande maîtrise de la communication médiatique sur le sujet. La **référence à l'intelligence artificielle des produits et des services** offerts sur le marché gagnerait à être encadrée par un principe déontologique : toute entreprise étant tentée d'utiliser l'IA comme un élément de stratégie marketing, les abus devraient être évités par de bonnes pratiques ou une charte professionnelle non contraignante.

Votre rapporteur Claude de Ganay invite, en particulier, 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 apparaît 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) eut rédigé les premiers textes décrivant les interactions sociales rendues possibles par l'intermédiaire d'un réseau d'ordinateurs¹. 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**

¹ 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 ».

de cible mais bien définir des objectifs réalistes garantissant des résultats effectifs.

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 **superintelligence** ne fait pas partie de ces risques à court et moyen termes. **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** 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, en fin de compte, **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**.

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

RÉUNION DE L'OPECST DU 14 MARS 2017 : ADOPTION DU RAPPORT

M. Jean-Yves Le Déaut, député, président de l'OPECST. – 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 Étienne, 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, premier vice-président de l'OPECST. – 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, comme dans la médecine,

un moyen 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, membre de l'OPECST, 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 Étienne : 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), d'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 à 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.

À la suite de 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. Devant 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, après 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 ». 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 cachent 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 remarquable 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 multicouches, 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, telles les architectures profondes ou multicouches (plusieurs dizaines ou centaines de couches), dans lesquelles 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'insiste, ce que sait l'intelligence artificielle, c'est l'homme qui le lui a appris.

M. Claude de Ganay, député, membre de l'OPECST, 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^e 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 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 compétents et d'un tissu de start-up très dynamiques : 240 d'entre elles sont spécialisées en intelligence artificielle. Ce tissu de start-up, encouragé par l'initiative French Tech, 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-up 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 prémices 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 homme-machine 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 courses*), 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 étonné par les 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 au 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.

Comme déjà signalé dans le rapport, le positionnement de la CERNA, qui étudie les questions éthiques du point de vue de la recherche et de la technologie, est à mettre en synergie avec celui récemment dévolu à la CNIL en matière d'instruction des questions éthiques dans leur dimension plus sociétale.

Aux termes de la loi pour une République numérique, la CNIL, notre autorité de contrôle en matière de protection des données personnelles, est en effet 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 ». 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, d'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 quelque 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* (IEET), 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 vingt-trois 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 à 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 notamment de ce point. L'initiative « Transalgo » d'Inria porte largement sur ce sujet.

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 d'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. 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 à 2020 et qu'elles la 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 homme-machine 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'emporte 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, 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. En effet, 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 comme le bien-être, la santé, la dépendance, le handicap, les infrastructures civiles, la 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 d'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. En réaction 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 détaillée 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 termes. 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'anticipation 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 possible. 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, premier 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, sénateur, membre de l'OPECST. – 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, sénateur, membre de l’OPECST. – 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, membre de l’OPECST. – 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...

Plutôt que 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.

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'Allistene¹

- 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'École 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

¹ 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, 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 ;
- 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. 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 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^e Alain Bensoussan, avocat, président de l'association pour les droits des robots et M^e 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-Télécom), chercheur en électronique et informatique, membre de

l'Académie des sciences, surtout connu pour ses travaux sur les turbo-codes, très utilisés en téléphonie mobile¹ ;

- Mme Laurence Devillers² 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 (Limsi 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*³ ;
- M. Dominique Sciamma directeur de l'école de Design « *Strate* » à Sèvres ;
- M. David Sadek directeur de la recherche de l'Institut Mines-Télécom, spécialiste en intelligence artificielle ;
- Mme Verity Harding, directrice des affaires publiques de *Deep Mind* (Londres, Royaume-Uni) ;

¹ 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 Télécom, 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.

² 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.

³ 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.

-
- M. Paul Strachman représentant d'ISAI aux États-Unis¹, 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 ; membre de l'Académie des technologies ;
- M. Yves Demazeau, président de l'Association française pour l'intelligence artificielle (AFIA) ;
- M. Bertrand Braunschweig, directeur du Centre 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 ;
- M. Gilles Babinet, entrepreneur, *digital champion* auprès de la Commission européenne ;
- M. Henri Verdier, directeur interministériel du numérique ;

¹ 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-up.

- 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 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 Inria ;
- M. Jean Ponce, professeur et directeur du département d'informatique de 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 avec le soutien du CNRS, d'Inria, du CEA et de l'Institut Mines-Télécom)

- 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 ;
 - M. Claude Kirchner, directeur de recherche Inria, président du COERLE ;
 - Mme 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) ;
- M. Lionel Tardy, député de la Haute-Savoie
- M. Laurent Celerier, capitaine de vaisseau, commandement opérationnel de cyberdéfense (COMCYBER) ;
- Mme Nozha Boujemaa, directrice de recherche Inria ;
- M. Jean Ponce, professeur et directeur du département d'informatique de l'École normale supérieure (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. Éric 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. Éric 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* ;

-
- M. Johann Christoph Ebell, directeur exécutif du projet HBP ;
 - M. Jean-Pierre Changeux, chercheur au HBP, École normale supérieure, Collège de France, Institut Pasteur ;
 - M. Bernd Stahl, directeur de l'éthique du HBP, directeur du centre de calcul et responsabilités sociales. Professeur à l'Université De Montfort (Royaume-Uni) ;
 - Mme 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)* ;

-
- Mme Susan Gowans, Administrateur, *Leverhulme Centre for the Future of Intelligence* ;
 - Professeur 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

- Professeur Alex "Sandy" Pentland, professeur d'informatique, chaire Marvin Minsky, MIT Media Lab ;
- Professeur David H. Autor, co-director of the School Effectiveness and Inequality Initiative economist and professor of economics, MIT ;
- Professeur 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 Miailhes 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) ;
- Professeur Thomas A. Kochan, professor of industrial relations, work and employment, MIT Sloan School of Management ;

-
- Mme Chloé 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 ;
 - M. 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 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.

BIBLIOGRAPHIE

- *Panorama de l'intelligence artificielle, ses bases méthodologiques, ses développements*¹, dirigé par **Pierre Maquis, Odile Papini et Henri Prade**, Cépaduès éditions, 2014.
- *Intelligence artificielle : problèmes et méthodes*, **Gérard Tisseau et Jacques Pitrat**, Presses universitaires de France, 1996.
- *De l'intelligence humaine à l'intelligence artificielle*, **Hugues Bersini**, Ellipses, 2006.
- *Les fondements de l'informatique*, **Hugues Bersini et Marie-Paule Spinette-Rose**, Vuibert, 2014.
- *Brèves réflexions d'un informaticien obtus sur la société à venir*, **Hugues Bersini**, Académie royale de Belgique, 2017.
- *Les Sciences cognitives*, **Jean-Gabriel Ganascia**, Flammarion, 1998.
- *L'intelligence artificielle*, **Jean-Gabriel Ganascia**, Le Cavalier Bleu, 2007.
- *Le mythe de la Singularité*, **Jean-Gabriel Ganascia**, Le Seuil, 2017.
- *Des robots et des hommes : mythes, fantasmes et réalité*, **Laurence Devillers**, Plon, 2017.
- *Le temps des algorithmes*, **Serge Abiteboul et Gilles Dowek**, Le Pommier, 2017.
- *À quoi rêvent les algorithmes : nos vies à l'heure des big data*, **Dominique Cardon**, Le Seuil, 2015.
- *Droit des robots, Comparative Handbook : robotic technologies law*, **Alain Bensoussan et Jérémy Bensoussan**, éditions Larcier, 2016.
- *En compagnie des robots*, **Alain Bensoussan, Jean-Gabriel Ganascia, Yannis Constantinidès, Kate Darling, John McCarthy et Olivier Tesquet**, Premier Parallèle, 2016.
- *Dictionnaire politique d'internet et du numérique*, **Christophe Stener**, Books on Demand, 2016.
- *Code Informatique, fichiers et libertés*, **Alain Bensoussan**, éditions Larcier, 2014.
- *Droit de l'informatique et de la télématique*, **Alain Bensoussan**, Berger-Levrault, 1985.

¹ 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.

- *Traité de droit et d'éthique de la robotique civile*, **Nathalie Nevejans**, Les Études Hospitalières édition, 2017.
- *La Machine de Turing*, **Jean-Yves Girard et Alan Turing**, Le Seuil, Collections Points Sciences, 1999.
- *La révolution transhumaniste*, **Luc Ferry**, Plon, 2016.
- *Superintelligence*, **Nick Bostrom**, OUP Oxford, 2016.
- *Deep Learning*¹, Yoshua Bengio, Aaron Courville et Ian Goodfellow, The MIT Press, 2016.
- *Learning in Graphical Models (Adaptive Computation and Machine Learning)*, **Michael Jordan**, A Bradford Book, 1998.
- *Probabilistic Graphical Models : Principles and Techniques*, **Daphne Koller et Nir Friedman**, The MIT Press, 2009.
- *Comprendre le Deep Learning : Une introduction aux réseaux de neurones*, **Jean-Claude Heudin**, Science eBook, 2016.
- *Immortalité numérique : Intelligence artificielle et transcendance*, **Jean-Claude Heudin**, Science eBook, 2016.
- *Les 3 Lois de la robotique : Faut-il avoir peur des robots ?*, **Jean-Claude Heudin**, Science eBook, 2015.
- *Robots et avatars : Le rêve de Pygmalion*, **Jean-Claude Heudin**, Odile Jacob, 2009.
- *Les créatures artificielles : des automates aux mondes virtuels*, **Jean-Claude Heudin**, Odile Jacob, 2008.
- *L'Homme neuronal*, **Jean-Pierre Changeux**, Fayard, 1983.
- *Un cerveau très prometteur : Conversation autour des neurosciences*, **Jean-Michel Besnier, Francis Brunelle et Florence Gazeau**, Le Pommier, 2015.
- *Les systèmes multi-agents : vers une intelligence collective*, **Jacques Ferber**, InterEditions, 1997.
- *L'âge de la multitude : Entreprendre et gouverner après la révolution numérique*, **Henri Verdier et Nicolas Colin**, Éditions Armand Colin, 2015.
- *Transformation digitale : l'avènement des plateformes*, **Gilles Babinet**, éditions Le Passeur, 2016.
- *Big Data, penser l'Homme et le monde autrement*, **Gilles Babinet**, éditions Le Passeur, 2015.
- *L'ère numérique, un nouvel âge de l'humanité*, **Gilles Babinet**, éditions Le Passeur, 2014.

¹ Ouvrage disponible en ligne : <http://www.deeplearningbook.org/>

-
- *L'industrie du Futur à travers le monde*, **Thibaut Bidet-Mayer**, fiche à télécharger sur le site <http://www.la-fabrique.fr/fr/publication/l-industrie-du-futur-a-travers-le-monde>
 - *L'Industrie 4.0 : Les défis de la transformation numérique du modèle industriel allemand*, **Dorothée Kohler** et **Jean-Daniel Weisz**, La Documentation Française, 2016
 - *Le Numérique – Une chance pour l'école*, **Joël Boissière**, **Simon Fau** et **Francesco Pedró**, Éditions Armand Colin, 2013.
 - *Comment le numérique transforme les lieux de savoir : Le numérique au service du bien commun et de l'accès au savoir pour tous*, **Bruno Devauchelle**, FYP éditions, 2012.
 - *L'école, le numérique et la société qui vient*, **Philippe Meirieu**, **Denis Kambouchner** et **Bernard Stiegler**, Fayard, 2012.
 - *Dans la disruption : Comment ne pas devenir fou ?*, **Bernard Stiegler**, Les liens qui libèrent Éditions, 2016.
 - *Apprendre avec le numérique*, **Franck Amadieu** et **André Tricot**, Retz, 2014.
 - *Le jour où mon robot m'aimera*, **Serge Tisseron**, Albin Michel, 2015.
 - *Vivre avec les robots, Essai sur l'empathie artificielle*, **Paul Dumouchel** et **Luisa Damiano**, Le Seuil, 2016.
 - *L'homme nu : la dictature invisible du numérique*, **Marc Dugain** et **Christophe Labbé**, Plon, 2016.
 - *L'Homme artificiel, Golems, robots, clones, cyborgs*, **Michel de Pracontal**, Denoël, 2002.
 - *La souveraineté numérique*, **Pierre Bellanger**, Stock, 2014.
 - *La silicolonisation du monde : l'irrésistible expansion du libéralisme numérique*, **Éric Sadin**, L'échappée Éditions, 2016.
 - *La vie algorithmique : critique de la raison numérique*, **Éric Sadin**, L'échappée Éditions, 2015.
 - *L'humanité augmentée : L'administration numérique du monde*, L'échappée Éditions, 2015.
 - *La Révolution transhumaniste*, **Luc Ferry**, Plon, 2016.
 - Le livre blanc d'Inria sur l'intelligence artificielle, coordonné par **Bertrand Braunschweig**¹
 - *Dix Questions sur l'intelligence artificielle et la technologie*, posées à l'académicien **Gérard Sabah**, dans une brochure de l'Académie des technologies.

¹ https://www.inria.fr/content/download/103897/1529370/./AI_livre-blanc_n01.pdf

- *Quelles priorités éducatives pour 2017-2027 ?*, note de **France Stratégie**, mai 2016.
- *Les risques de l'automatisation pour l'emploi*, **M. Arntz, T. Gregory et U. Zierahn**, rapport OCDE, 2016.
- *Automatisation, numérisation et emploi*, rapport du Conseil d'orientation pour l'emploi (COE), 2016.
- *L'école sous algorithme*, note de la fondation **Terra Nova**, mars 2016.
- *La déconnexion des élites. Comment Internet dérange l'ordre établi ?* **Laure Belot**, Les Arènes, 2015.
- *Intelligence artificielle et robotique : Confluences de l'Homme et des STIC*, cahier de l'Agence nationale de la recherche (ANR), mars 2012.
- *Is Google making us stupid ?*, article de **Nicholas Carr** dans la revue *The Atlantic*, juin 2008.
- *Social media, texting, and personality: A test of the shallowing hypothesis*, **Logan E. Anisette, Dr. Kathryn Lafreniere**, Elsevier, 2016.
- *It's 2001. Where Is HAL? »*, **Marvin Minsky**, Dr. Dobb's Technetcast, 2001.
- *Introduction to Multi-Agent Systems*, **Michael Wooldridge**, John Wiley & Sons, 2002.
- *Social Physics*, **Alex Pentland**, Penguin Press, 2014.
- *Logical Foundations of Artificial Intelligence*, **Michael R. Genesereth et Nils J. Nilsson**, Stanford University, Morgan Kaufmann Publishers, 1988
- *What to think about machines that think*, **John Brockman**, Harper Perennial, 2015.
- *A (Very) Brief History of Artificial Intelligence*, **Bruce G. Buchanan**, AI Magazine, 2005.
- *The Perceptron: A Probabilistic Model For Information Storage And Organization in the Brain*, **Frank Rosenblatt**, Psychological Review, 1958.
- *What the Frog's Eye Tells the Frog's Brain*, **J.Y. Lettvin, H. R. Maturanat, S. Mc Culloch, W. H. Pitts**, Proceedings of the Institute of Radio Engineers (IRE), 1959.
- *Artificial Intelligence: A Modern Approach*, **Stuart Russell et Peter Norvig**, Prentice Hall, 1995.

LES RAPPORTS DU SÉNAT

- *Les robots et la loi*, **Jean-Yves Le Deaut** et **Bruno Sido**, rapporteurs, OPECST, Sénat n° 570 (2015-2016).
- *Projet de loi pour une République numérique*, **Christophe-André Frassa**, rapporteur, Sénat n° 534 (2015-2016).
- *Le numérique au service de la santé*, **Catherine Procaccia** et **Gérard Bapt**, rapporteurs, OPECST, Sénat n° 465 (2014-2015).
- *Sécurité numérique et risques : enjeux et chances pour les entreprises*, **Anne-Yvonne Le Dain** et **Bruno Sido**, rapporteurs, OPECST, Sénat n° 271 (2014-2015).
- *La place du traitement massif des données (big data) dans l'agriculture : situation et perspectives*, **Jean-Yves Le Deaut**, **Anne-Yvonne Le Dain** et **Bruno Sido**, rapporteurs, OPECST, Sénat n° 614 (2014-2015).
- *Les drones et la sécurité des installations nucléaires*, **Jean-Yves Le Deaut** et **Bruno Sido**, rapporteurs, OPECST, Sénat n° 267 (2014-2015).
- *Les nouvelles mobilités sereines et durables*, **Denis Baupin** et **Fabienne Keller**, rapporteurs, OPECST, Sénat n° 293 (2013-2014).
- *Les enjeux des nouvelles technologies d'exploration et de traitement du cerveau*, **Alain Claeys** et **Jean-Sébastien Vialatte**, rapporteurs, OPECST, Sénat n° 476 (2011-2012).
- *Proposition de résolution sur la régulation des marchés financiers*, **Nicole Bricq**, rapporteur, Sénat n° 369 (2011-2012).
- *La gouvernance mondiale de l'Internet*, **Claude Birraux** et **Jean-Yves Le Deaut**, rapporteurs, OPECST, Sénat n° 219 (2005-2006).
- Les conséquences de l'évolution scientifique et technique dans le secteur des télécommunications, **Pierre Laffitte** et **René Tregouet**, rapporteurs, OPECST, Sénat n° 159 (2001-2002).
- Images de synthèse et monde virtuel : techniques et enjeux de société, **Claude Huriet**, rapporteur, OPECST, Sénat n° 169 (1997-1998).
- Les techniques des apprentissages essentiels pour une bonne insertion dans la société de l'information, **Franck Sérusclat**, rapporteur, OPECST, Sénat n° 383 (1996-1997).
- La France et la société de l'information : un cri d'alarme et une croisade nécessaire, **Pierre Laffitte**, rapporteur, OPECST, Sénat n° 213 (1996-1997).
- *Les nouvelles techniques d'information et de communication : l'Homme cybernétique*, **Franck Sérusclat**, rapporteur, OPECST, Sénat n° 232 (1994-1995).