

A S S E M B L É E   N A T I O N A L E

X I V <sup>e</sup>   L É G I S L A T U R E

# Compte rendu

## Office parlementaire d'évaluation des choix scientifiques et technologiques

Examen du projet de rapport présenté par M. Claude de Ganay, député, et Mme Dominique Gillot, sénatrice, sur « *Les enjeux économiques et sociétaux de l'intelligence artificielle* »

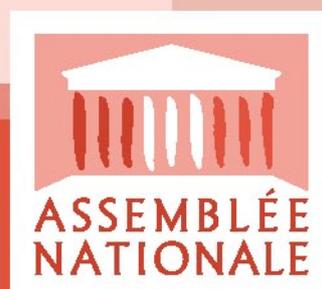
Mardi 14 mars 2017

Séance de 16 h 30

Compte rendu n° 88

SESSION ORDINAIRE DE 2016-2017

**Présidence  
de M. Jean-Yves  
Le Déaut,  
député,  
*Président***



## Office parlementaire d'évaluation des choix scientifiques et technologiques

Mardi 14 mars 2017

Présidence de M. Jean-Yves Le Déaut, député, président

*La séance est ouverte à 16 h 45*

– **Examen du projet de rapport présenté par M. Claude de Ganay, député, et Mme Dominique Gillot, sénatrice, sur « Les enjeux économiques et sociétaux de l'intelligence artificielle »**

**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 transformée, 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 dominent donc, mais la recherche et les entreprises chinoises montent en puissance. La Chine a ainsi pris la tête des publications en *deep learning* depuis trois ans. L'entreprise Baidu a développé le principal moteur de recherche chinois, site le plus consulté en Chine et le cinquième plus consulté au niveau mondial : indexant près d'un milliard de pages, l'entreprise dispose d'un flux de données permettant d'envisager des applications dans de nombreux domaines. Ses résultats algorithmiques sont impressionnants, malgré son existence récente. Le système de reconnaissance d'image de Baidu a ainsi battu celui de Google depuis 2015. Le recrutement du chercheur de Stanford, Andrew Ng, par Baidu en 2014 en tant que responsable de l'intelligence artificielle alors qu'il en était le responsable chez Google est emblématique. De même, en 2017, Baidu débauche Qi Lu, au poste de numéro deux, alors qu'il était auparavant vice-président chez Microsoft et directeur des projets Bing, Skype et Microsoft Office et auparavant directeur de la recherche de Yahoo.

Le 13<sup>e</sup> plan quinquennal chinois comprend une liste de quinze « grands projets » qui structurent les priorités scientifiques avec des investissements de plusieurs milliards d'euros. Ce plan vise à dynamiser la recherche chinoise en IA et à concurrencer les États-Unis. Parmi ces projets, ceux en lien avec l'IA représentent un montant de 100 milliards de yuans sur trois ans.

Autre caractéristique, l'interdisciplinarité est indispensable en intelligence artificielle, alors que la discipline demeure éclatée en une cinquantaine de sous-domaines de recherche, un tableau les décrit dans notre rapport.

Par ailleurs, la recherche en intelligence artificielle est soumise à une contrainte d'acceptabilité sociale assez forte, notamment sous l'effet de représentations catastrophistes, comme en témoignent différents sondages d'opinion, eux aussi rappelés dans le rapport.

Plusieurs interventions médiatiques et pétitions ont cherché en 2015 à interpeler l'opinion à propos des risques qui seraient inhérents à l'intelligence artificielle. L'existence d'une association française contre l'intelligence artificielle (AFCIA) est révélatrice d'un certain climat d'angoisse puisque la France serait le seul pays où une telle association existerait. L'AFCIA juge « *illégitime et dangereuse la recherche scientifique visant à créer des organismes à intelligence artificielle supra-humaine* » et considère que le seul moyen « d'éviter un avenir funeste pour l'humanité est d'obtenir l'interdiction légale de la recherche en intelligence artificielle à l'échelle mondiale ». Se définissant comme association de lobbying, elle vise à obtenir cette interdiction auprès des pouvoirs publics. Jacques Attali s'est, à la fin de l'année 2016, prononcé pour un moratoire sur les technologies d'intelligence artificielle, ce qui nous a beaucoup surpris.

Dernière caractéristique : la multiplication des initiatives visant la prise en compte de principes éthiques dans la recherche et les usages de l'intelligence artificielle. Cela vaut pour la recherche publique, comme pour la recherche privée, en Europe comme en Amérique.

Concernant la recherche française en IA, notre pays dispose d'importants atouts à faire valoir, même si la communauté française de l'intelligence artificielle est encore insuffisamment organisée, connue et visible. La reconnaissance internationale des travaux des chercheurs français doit beaucoup à des organismes comme 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 calculette 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 récuse 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 *Innorobo* 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 détaillé à la fin du mois.

Ni quête vaine ni projet de remplacement de l'homme par la machine, l'intelligence artificielle représente une chance à saisir pour nos sociétés et nos économies. La France doit relever ce défi. Les progrès en intelligence artificielle sont d'abord et avant tout bénéfiques. Ils comportent aussi des risques, il serait malhonnête de le nier. Mais ces risques peuvent et doivent être identifiés, anticipés et maîtrisés.

L'avènement d'une super-intelligence ne fait pas partie de ces risques à court et moyen 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'utilisateur, 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.*

*La séance est levée à 18 h 40*

### **Membres présents ou excusés**

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

Réunion du mardi 14 mars 2017 à 16 h 30

##### Députés

*Présents.* - M. Claude de Ganay, M. Jean-Yves Le Déaut

*Excusés.* - M. Bernard Accoyer, M. Alain Claeys, M. Alain Marty, Mme Maud Olivier, Mme Dominique Orliac, M. Jean-Louis Touraine

##### Sénateurs

*Présents.* - Mme Delphine Bataille, M. Michel Berson, Mme Dominique Gillot, Mme Fabienne Keller, M. Christian Namy, Mme Catherine Procaccia, M. Bruno Sido

*Excusés.* - M. Patrick Abate, M. Gilbert Barbier, M. Alain Houpert, M. Jean-Pierre Leleux, M. Franck Montaugé