

## CONCOURS EXTERNE DE BIBLIOTHECAIRE TERRITORIAL

**SESSION 2020**  
*REPORTÉE À 2021*

### ÉPREUVE DE NOTE

#### ÉPREUVE D'ADMISSIBILITÉ :

**Rédaction d'une note de synthèse, établie à partir d'un dossier portant au choix du candidat exprimé au moment de l'inscription :**

- **soit sur les lettres et les sciences humaines**
- **soit sur les sciences exactes et naturelles et les techniques**
- **soit sur les sciences juridiques, politiques et économiques**

Durée : 4 heures  
Coefficient : 2

<b>SCIENCES EXACTES ET NATURELLES ET LES TECHNIQUES</b>
---

<p><b>À LIRE ATTENTIVEMENT AVANT DE TRAITER LE SUJET :</b></p> <ul style="list-style-type: none"><li>♦ Vous ne devez faire apparaître aucun signe distinctif dans votre copie, ni votre nom ou un nom fictif, ni initiales, ni votre numéro de convocation, ni le nom de votre collectivité employeur, de la commune où vous résidez ou du lieu de la salle d'examen où vous composez, ni nom de collectivité fictif non indiqué dans le sujet, ni signature ou paraphe.</li><li>♦ Sauf consignes particulières figurant dans le sujet, vous devez impérativement utiliser une seule et même couleur non effaçable pour écrire et/ou souligner. Seule l'encre noire ou l'encre bleue est autorisée. L'utilisation de plus d'une couleur, d'une couleur non autorisée, d'un surligneur pourra être considérée comme un signe distinctif.</li><li>♦ Le non-respect des règles ci-dessus peut entraîner l'annulation de la copie par le jury.</li><li>♦ Les feuilles de brouillon ne seront en aucun cas prises en compte.</li></ul>
---



**Ce sujet comprend 35 pages.**

**Il appartient au candidat de vérifier que le document comprend le nombre de pages indiqué.**

*S'il est incomplet, en avertir le surveillant.*

Vous êtes bibliothécaire territorial dans la commune de Cultureville.

La directrice de votre médiathèque vous demande de rédiger à son attention, exclusivement à l'aide des documents joints, une note sur l'intelligence artificielle.

### Liste des documents :

- Document 1 :** « Peut-on faire confiance à l'intelligence artificielle ? » - C. Zeitoun - *lejournale.cnrs.fr* - 25 janvier 2018 - 4 pages
- Document 2 :** « Donner un sens à l'intelligence artificielle. Pour une stratégie nationale et européenne » (Introduction, extrait) - Mission parlementaire Villani - *aiforhumanity.fr* - Mars 2018 - 4 pages
- Document 3 :** « Le futur de l'intelligence artificielle » - T. Gaudiaut - *fr.statista.com* - 3 janvier 2019 - 1 page
- Document 4 :** « L'Europe et la course à l'Intelligence Artificielle » - F. Merz - *Politique de sécurité : analyses du CSS n° 247* - Juin 2019 - 3 pages
- Document 5 :** « Intelligence artificielle : découvrez les 15 nouveaux projets sélectionnés » - P.-A. Chevalier - *etalab.gouv.fr* - 16 juillet 2019 - 1 page
- Document 6 :** « Intelligence artificielle : des consommateurs sous surveillance » - C. Prudhomme - *lemonde.fr* - 21 juin 2019 - 2 pages
- Document 7 :** « Le mythe de la singularité. Faut-il craindre l'intelligence artificielle ? » (Extrait) - J.-G. Ganascia - Février 2017 - 5 pages
- Document 8 :** « L'insolente croissance du marché de l'intelligence artificielle » - J. El Hassani - *journaldunet.com* - 13 juillet 2018 - 1 page
- Document 9 :** « Lignes directrices en matière d'éthique pour une IA digne de confiance » (Extrait) - Rapport de la Commission européenne - *ec.europa.eu* - Avril 2019 - 6 pages
- Document 10 :** « Il faut tout faire pour que l'automatisation des métiers favorise l'émancipation de l'homme » - N. Lepeltier - *lemonde.fr* - 30 juillet 2019 - 2 pages
- Document 11 :** « L'intelligence artificielle transforme le monde » - S. Gaudelus, M.-H. Gatto - *de ligne en ligne n°27* - 28 septembre 2018 - 2 pages
- Document 12 :** « IA : "l'éthique-washing", une invention européenne » - T. Metzinger - *meta-media.fr* - 15 avril 2019 - 2 pages

### Documents reproduits avec l'autorisation du C.F.C.

*Certains documents peuvent comporter des renvois à des notes ou à des documents non fournis car non indispensables à la compréhension du sujet.*

## DOCUMENT 1

Cnrs le journal  
Charline Zeitoun  
25.01.2018

### Peut-on faire confiance à l'intelligence artificielle ?

**S'en remettre à des machines et à des systèmes d'aide à la décision peut poser de gros problèmes éthiques. Et les programmes ont beau disposer d'une logique « froide », ils ne sont pas exempts de préjugés...**

*« Les féministes devraient toutes brûler en enfer » et « Hitler aurait fait un meilleur boulot que le singe actuel », psalmodiait en mars 2016 le chatbot de Microsoft, Tay, dès son premier jour d'immersion sur Twitter, Snapchat, Kik et GroupMe, apprenant en deep learning comment parler avec les « djeun's ». Entraînée par les internautes qui s'amusaient à la faire déraper, l'intelligence artificielle (IA) twitteuse a même fini par nier l'existence de l'Holocauste. Piteuse vitrine de l'apprentissage machine, Tay a été mise hors circuit par ses concepteurs au bout de quelques heures. Mais que se passerait-il si on délégua à des IA et autres algorithmes des décisions importantes ?*

#### Une aide à la décision

En vérité, banques, sociétés d'assurances et directions des ressources humaines peuvent déjà tester d'efficaces systèmes d'aide à la décision pour gérer des patrimoines, calculer des primes et sélectionner des CV. Des voitures autonomes arpentent depuis des années les routes de Californie. Tandis que l'algorithme d'admission post-bac (qui a conduit au tirage au sort de certains bacheliers de la promotion 2017 pour une place en fac) n'a pas fini de faire grincer des dents. *« Pour un film ou des chaussettes, ça m'est égal de recevoir des conseils de systèmes d'aide à la décision, mais je trouve plus gênant qu'ils orientent mes lectures vers des sites d'information qui peuvent conditionner mes opinions, voire être complotistes »,* commente Serge Abiteboul, chercheur au département d'Informatique de l'École normale supérieure. *« Et lorsqu'on se fie aux algorithmes et à l'IA (algorithme sophistiqué "simulant" l'intelligence) pour prendre des décisions qui ont de lourdes conséquences dans la vie des êtres humains, cela pose clairement des problèmes éthiques »,* complète-t-il.

En la matière, la libération sur parole des détenus américains est un exemple étonnant. *« Il a été démontré que la probabilité d'être libéré est très supérieure si vous passez devant le juge après déjeuner plutôt que juste avant »,* informe Serge Abiteboul. Des algorithmes, exempts du « syndrome » du ventre vide, ont été testés en parallèle. *« Comparer leurs performances est facile car, aux États-Unis, la libération sur parole ne dépend que d'un seul paramètre : le risque que la personne s'enfuit ou récidive. »* Résultat du match : *« Statistiquement, l'algorithme gagne haut la main et permet dans ce cas-là une justice aveugle ne tenant compte que des faits objectifs »,* commente le chercheur. Mais jusqu'où aller ? Si des systèmes perfectionnés permettaient de juger d'autres cas, accepteriez-vous la décision d'une machine ?

Ce n'est pas un problème purement philosophique. En déléguant nos décisions aux algorithmes et à l'intelligence artificielle, nous ne perdons pas seulement notre dignité humaine (ce qui n'est pas rien !) : ces systèmes ont eux aussi leurs failles. *« Le deep learning, technique parmi d'autres en intelligence artificielle, est à la fois celle qui occasionne le plus d'applications spectaculaires et qui présente un inconvénient majeur : on ne sait pas en expliquer les résultats. Ce sont des réseaux de neurones qui fonctionnent comme des boîtes noires »,* souligne Sébastien Konieczny, chercheur au Centre de recherche en informatique de Lens. Avec cette forme d'intelligence artificielle en effet, on ne reconnaît pas un chat parce qu'« il a deux oreilles, quatre pattes, etc. » (raisonnement humain composé de règles et dicté à la machine), mais parce qu'il ressemble à une foultitude d'autres chats dont on aura fourni les images à la

machine pour l'« entraîner ». Quant à savoir quelles ressemblances font tilt pour celle-ci, mystère et boule de gomme.

« Or, il serait bon d'expliquer les raisons qui président aux choix importants, afin d'être en mesure de les justifier. Et garantir ainsi à chacun un traitement équitable », rappelle Raja Chatila, directeur de l'Institut des systèmes intelligents et de robotique. Pourrait-on rendre ces systèmes plus transparents ? « Il y a des recherches sur "l'explicabilité" de ces boîtes noires, notamment financées par la DARPA », répond le chercheur. « Mais les réseaux de neurones ne sont que des calculs numériques : je ne vois pas comment on pourrait en extraire des concepts », observe Sébastien Konieczny. Or, personne n'acceptera de se voir refuser un prêt ou un poste intéressant à cause de la connexion 42 du réseau de neurones hélas inférieure à 0,2...

### **Les machines « contaminées » par nos préjugés**

Pire encore : les machines ont beau disposer d'une logique froide, elles n'échappent pas aux préjugés. En matière d'apprentissage, « Tay-la-révisionniste » n'est pas la seule mauvaise élève. En avril 2017, la revue *Science* révélait les catastrophiques stéréotypes racistes et sexistes de GloVe, une intelligence artificielle « nourrie » de 840 milliards d'exemples piochés sur le Web en quarante langues différentes dans le but de faire des associations de mots. « Si un système est entraîné sur une masse de données issues de discours humains, rien d'étonnant à ce qu'il en reproduise les biais », indique Sébastien Konieczny.

Même problème pour les prêts bancaires. « Le système peut apprendre à partir des dix années précédentes à qui l'on a accordé un prêt et à quel taux, selon son salaire, sa situation familiale, etc. Mais il reproduira les biais des décisions humaines de cette période : alors si les minorités ethniques payaient des intérêts plus élevés par exemple, cette injustice se perpétuera », souligne Serge Abiteboul. « Le concepteur de ces systèmes n'est pas seul en cause, celui qui l'entraîne est également responsable en cas de dérive », complète le chercheur. Or, il sera impossible d'anticiper tous les biais potentiels. En particulier pour les systèmes conçus pour apprendre en permanence afin de s'améliorer, comme « Tay-la-langue-de-vipère » qui se modifiait en imitant ses malicieux interlocuteurs. Avant de déployer le *deep learning* à grande échelle, il serait bon de lui inculquer quelques règles « morales ».

« Dans un réseau de neurones, méthode purement numérique, on ne peut ni coder ni dicter de règles d'éthique, comme imposer que le résultat ne dépende pas du sexe, de l'âge, de la couleur de peau... Mais c'est possible avec une approche symbolique qui, elle, est faite de règles rédigées par le programmeur », commente Sébastien Konieczny. « Une solution serait d'hybrider les systèmes d'apprentissage avec des prescriptions que la machine serait contrainte de respecter », complète Jean-Gabriel Ganascia, président du Comité d'éthique du CNRS et chercheur au Laboratoire d'informatique de Paris 6 (lire l'encadré ci-dessous). « Notez que l'on ignore si c'est possible techniquement car les deux approches sont intrinsèquement différentes, souligne Sébastien Konieczny. C'est justement l'un des gros défis à venir en intelligence artificielle. »

Autre question à résoudre : qui décidera des règles à implémenter ? « Certainement pas les informaticiens, répond Serge Abiteboul. « Ce ne devrait pas être à eux de décider comment mettre au point l'algorithme qui calcule le sort des bacheliers, tout comme ce ne devrait pas être à Google de décider d'interdire des sites extrémistes ou de fake news, déplore le chercheur. Le monde numérique s'est développé si vite qu'il en est encore au stade du western : les injustices foisonnent, l'État ne comprend pas assez bien pour légiférer correctement, et les citoyens sont perdus, résume-t-il. La loi relative au renseignement de juillet 2015 en est un tragique exemple. La majorité des informaticiens étaient contre car ils en comprenaient les conséquences. Politiques et citoyens qui l'ont votée ou laissé voter peuvent ne pas être d'accord avec nous, bien sûr, mais je doute que ce soit en connaissance de cause », regrette-t-il.

« Pour ce qui est de l'intelligence artificielle et de certaines recherches dans le numérique, il faudrait créer des comités d'éthique opérationnels spécifiques, à l'image de ceux qui, en

*biologie par exemple, nous aident à évaluer les limites à ne pas dépasser et cadrer certains travaux face aux dérives qu'ils engendreraient* », complète Raja Chatila, qui est aussi membre de la Cerna. Mais le temps de la réflexion est souvent plus long que celui de l'innovation technologique. Au sujet de la très médiatique voiture autonome, une logique utilitariste (sauver le plus de personnes possible quitte à sacrifier le conducteur), elle-même discutable, aura sans doute du mal à faire le poids face aux logiques de vente : en 2016, un responsable de Mercedes annonçait ainsi qu'il fallait protéger les passagers en priorité. Alors que le ministère des Transports allemand, pressé de prendre position, a présenté en septembre 2017 un rapport d'experts en intelligence artificielle, en droit et en philosophie, qui recommande de donner la même valeur à tous les humains, quels que soient leur âge, leur sexe, etc.

## **Maintenir la responsabilité du côté des humains**

En attendant les décisions et les solutions techniques pour les injecter dans les systèmes d'apprentissage, de nombreux chercheurs s'accordent sur un point : il faut laisser à l'humain la décision finale dans les cas délicats. *« À part la voiture autonome et quelques autres exemples, qui nécessitent des réactions en temps réel, les résultats proposés par les systèmes d'aide à la décision autorisent presque tous le temps de la réflexion. On pourrait donc s'inspirer de la façon dont sont prises certaines décisions, dans le domaine militaire notamment, via des protocoles codifiés impliquant plusieurs acteurs »*, fait ainsi remarquer Jean-Gabriel Ganascia.

Reste un léger problème : on a déjà largement observé, chez les pilotes d'avion en particulier, que l'humain croit presque toujours les machines plus pertinentes que lui, grâce aux énormes quantités d'informations dont elles disposent. Dès lors, qui osera aller contre leurs verdicts ? *« Si nous n'en sommes pas capables, il ne faut pas les utiliser. Nous devons prendre nos responsabilités. Et si quelqu'un se défait sur la machine, ce sera sa décision... et il en sera responsable »*, insiste Raja Chatila. *« On peut aussi imaginer configurer les systèmes pour qu'ils donnent plusieurs solutions, si possible avec les raisons associées et leurs conséquences (à condition que les recherches sur l'explicabilité progressent), et laisser l'humain choisir parmi elles »*, propose Sébastien Konieczny. Mais pas question de démissionner. Sinon nous pourrions finir comme les humains du roman de Pierre Boulle, devenus trop paresseux et trop bêtes pour conserver leur autonomie et leur liberté à force de laisser travail et responsabilités aux singes...

Là est bien le danger. Pas dans les fantasmes *made in Hollywood* de prise de pouvoir par des machines rebelles et malveillantes. *« Cela correspond à ce qu'on appelle la "singularité", cela n'a aucun fondement scientifique et ce n'est pas près d'arriver, contrairement à ce que prétendent certains cadors de l'industrie numérique et transhumanistes. Le véritable danger, c'est nous ! Lorsque, par ignorance ou par facilité, nous déléguons les décisions et notre autonomie à la machine, martèle Jean-Gabriel Ganascia. Laisser les agents autonomes ultra-rapides aux commandes de la Bourse a ainsi sans doute conduit à la crise de 2008. Mais il y a une confusion à cause du terme "autonome" : au sens technique, cela ne veut pas dire qu'une machine définit ses propres objectifs. Cela veut seulement dire qu'elle peut atteindre sans intervention humaine un objectif donné, celui-ci ayant bel et bien été fixé par l'humain-programmeur. »*

Les robots ne sont ni gentils ni méchants et n'ont pas le moindre projet personnel. Ils font ce qu'on leur a dit de faire. Et nous pouvons les débrancher si certains effets pervers non prévus apparaissent en cours de route. *« Les peurs irrationnelles de prise de pouvoir par les machines masquent des enjeux politiques et économiques majeurs »*, reprend Jean-Gabriel Ganascia. *« L'asservissement à la machine est bien moins important que l'asservissement à la compagnie privée qui la contrôle »*, souligne-t-il. Le chercheur s'inquiète du glissement du pouvoir des États vers les *big companies*, fortes de milliards de données sur les citoyens, et bientôt plus encore via les futures applications dotées d'IA analysant tous nos comportements pour mieux nous assister. *« Et tout ça en se cachant derrière un pseudo-moratoire sur l'éthique »*, s'agace le chercheur. Politiques, industriels, citoyens, il est urgent que chacun s'y intéresse, pour être en mesure de décider des limites éthiques à poser et bâtir le meilleur des mondes numériques possible.♦

---

## Des robots au banc des accusés ?

Si une voiture autonome a un accident, qui est responsable ? « *La responsabilité dérive de la notion de personnalité juridique* », explique Jean-Gabriel Ganascia. Dans deux des textes qui se mêlent de robotique en Europe, on propose d'en donner une aux machines les plus sophistiquées (le texte évoque une « *personnalité électronique* »). Objectif : leur faire rembourser les dommages qu'elles ont causés (sauf cas grave ou mortel qui relèvent du pénal). « *Dans ce cas, pour indemniser les victimes, il faudra créer un fonds assurantiel alimenté par les entreprises qui fabriquent ou possèdent les machines* », explique-t-il, car les robots eux-mêmes sont évidemment sans le sou... « *C'est vraiment la pire chose à faire, alerte le chercheur, car indemniser les victimes dans les cas mineurs a un effet pervers : on fera l'économie d'une enquête destinée à identifier les causes et à prévenir d'autres accidents, peut-être plus graves. Or, les machines autonomes impliquent des processus différents des produits classiques.* » Par exemple, vous savez qu'en agitant frénétiquement un marteau en l'air, vous risquez de vous assommer, voire que le manche se démonte. Vous, comme le fabricant, pouvez en être tenu pour responsable, mais au moins tout le monde sait à peu près à quoi s'attendre en utilisant l'outil. « *Tandis qu'une machine qui apprend peut se reprogrammer de manière dynamique en fonction de l'environnement dans lequel elle évolue et de façon parfois imprévisible* », rappelle le chercheur. « *Au lieu d'indemniser sans enquête, il serait préférable d'analyser les causes des accidents afin de déterminer qui, de l'utilisateur, du fabricant ou du "concessionnaire", ayant mal entraîné le robot, est responsable* », conclut-il.

---

## Programmer la « morale » en langage machine

La logique classique n'est pas d'un grand secours pour formaliser des règles d'éthique : trop rigide, elle se limite à des « vrai » ou « faux », et des « si ceci » alors « faire cela » ou au contraire « ne pas le faire ». « *Or, en éthique, on se retrouve souvent coincé entre plusieurs injonctions contradictoires* », explique Jean-Gabriel Ganascia. Par exemple, vous pouvez devoir mentir pour empêcher un assassinat. Ou encore franchir une ligne blanche pour éviter de renverser un piéton. « *En cas de contradictions ou de dilemmes, il faudrait induire des préférences entre plusieurs "mauvaises" actions, avec des coefficients de pondération par exemple* », poursuit Jean-Gabriel Ganascia. « *Pour ce faire, plusieurs projets de recherche se fondent sur la logique déontique, ajoute Raja Chatila. Ses opérateurs modaux permettent de formaliser des possibilités facultatives, c'est-à-dire des actions autorisées mais qui ne seront pas forcément réalisées. On peut aussi utiliser une approche probabiliste, par exemple en tenant compte de la probabilité que ce que je distingue dans le brouillard est bien un piéton. Alors les prédicats ne sont pas soit "vrai" soit "faux" mais "peut-être vrai" ou "peut-être faux" selon une distribution de probabilité* », conclut-il.

*Cet article a été publié dans le numéro 3 de la revue Carnets de science.*

# **DONNER UN SENS À L'INTELLIGENCE ARTIFICIELLE POUR UNE STRATÉGIE NATIONALE ET EUROPÉENNE**

**(extrait)**

## **Introduction**

Définir l'intelligence artificielle n'est pas chose facile. Depuis ses origines comme domaine de recherche spécifique, au milieu du XX<sup>e</sup> siècle, elle a toujours constitué une frontière, incessamment repoussée. L'intelligence artificielle désigne en effet moins un champ de recherches bien défini qu'un programme, fondé autour d'un objectif ambitieux : comprendre comment fonctionne la cognition humaine et la reproduire ; créer des processus cognitifs comparables à ceux de l'être humain.

Le champ est donc naturellement extrêmement vaste, tant en ce qui concerne les procédures techniques utilisées que les disciplines convoquées : mathématiques, informatiques, sciences cognitives... Les méthodes d'IA sont très nombreuses et diverses (ontologique, apprentissage par renforcement, apprentissage adversarial, réseaux de neurones...) et ne sont pas nouvelles : beaucoup d'algorithmes utilisés aujourd'hui ont été développés il y a plusieurs dizaines d'années.

Depuis la conférence de Dartmouth de 1956, l'intelligence artificielle s'est développée, au gré des périodes d'enthousiasme et de désillusion qui se sont succédées, repoussant toujours un peu plus les limites de ce qu'on croyait pouvoir n'être fait que par des humains. En poursuivant son projet initial, la recherche en IA a donné lieu à des vrais succès (victoire au jeu d'échecs, au jeu de go, compréhension du langage naturel...) et a nourri largement l'histoire des mathématiques et de l'informatique : combien de dispositifs que nous considérons aujourd'hui comme banals étaient à l'origine une avancée majeure en IA – une application de jeux d'échecs, un programme de traduction en ligne... ?

Du fait de ses ambitions, qui en font un des programmes scientifiques les plus fascinants de notre époque, la discipline de l'IA s'est toujours développée de concert avec les imaginaires les plus délirants, les plus angoissants et les plus fantastiques, qui ont façonné les rapports qu'entretient le grand public avec l'IA mais également ceux des chercheurs eux-mêmes avec leur propre discipline. La (science) fiction, les fantasmes et les projections collectives ont accompagné l'essor de l'intelligence artificielle et guident parfois ses objectifs de long terme : en témoignent les productions fictionnelles abondantes sur le sujet,

**L'intelligence artificielle est entrée, depuis quelques années, dans une nouvelle ère, qui donne lieu à de nombreux espoirs**

de 2001 *l'Odyssée de l'espace*, à *Her* en passant *Blade Runner* et une grande partie de la littérature de science-fiction. Finalement, c'est probablement cette alliance entre des projections fictionnelles et la recherche scientifique qui constitue l'essence de ce qu'on appelle l'IA. Les imaginaires, souvent ethno-centrés et organisés autour d'idéologies politiques sous-jacentes, jouent donc un rôle majeur, bien que souvent négligé, dans la direction que prend le développement de cette discipline.

L'intelligence artificielle est entrée, depuis quelques années, dans une nouvelle ère, qui donne lieu à de nombreux espoirs. C'est en particulier dû à l'essor de l'apprentissage automatique. Rendues possibles par des algorithmes nouveaux, par la multiplication des jeux de données et le décuplement des puissances

# Introduction

1. La *blockchain* correspond à un registre distribué qui permet d'éviter de recourir à un tiers de confiance lors de transactions et qui est notamment au fondement du *bitcoin*.

de calcul, les applications se multiplient : traduction, voiture autonome, détection de cancer,... Le développement de l'IA se fait dans un contexte technologique marqué par la « mise en données » du monde (*datafication*), qui touche l'ensemble des domaines et des secteurs, la robotique, la *blockchain*<sup>1</sup>, le supercalcul et le stockage massif. Au contact de ces différentes réalités technologiques se jouera sûrement le devenir de l'intelligence artificielle.

[...]

Contrairement aux dernières grandes périodes d'emballlement de la recherche en intelligence artificielle, le sujet a très largement dépassé la seule sphère scientifique et est sur toutes les lèvres. Les investissements dans la recherche et dans l'industrie atteignent des sommes extraordinaires, notamment en Chine. Les responsables politiques du monde entier l'évoquent dans les discours de politique générale comme un levier de pouvoir majeur : l'emblématique interview à *Wired* de Barack Obama en octobre 2016 montrait que ce dernier avait bien compris l'intérêt de faire de l'avance américaine en intelligence artificielle un outil redoutable de *soft power*. Le Président russe Vladimir Poutine a quant à lui affirmé que « celui qui deviendra le leader dans ce domaine sera le maître du monde », comparant l'intelligence artificielle aux technologies nucléaires. S'il s'agissait vraisemblablement pour lui de compenser le retard de la Russie en matière d'intelligence artificielle par un discours musclé sur le sujet, cette affirmation est révélatrice de l'importance géostratégique prise par ces technologies. Dans la mesure où les chaînes de valeur, surtout dans le secteur numérique, sont désormais mondiales, les pays qui seront les leaders dans le domaine de l'intelligence artificielle seront amenés à capter une grande partie de la valeur des systèmes qu'ils transforment, mais également à contrôler ces mêmes systèmes, mettant en cause l'indépendance des autres pays.

C'est que l'intelligence artificielle va désormais jouer un rôle bien plus important que celui qu'elle jouait jusqu'alors. Elle n'est plus seulement un programme de recherche confiné aux laboratoires ou à une application précise. Elle va devenir une des clés du monde à venir. En effet nous sommes dans un monde numérique, de plus en plus, de part en part. Un monde de données. Ces données qui sont au cœur du fonctionnement des intelligences artificielles actuelles. Dans ce monde-là, qui est désormais le nôtre, ces technologies représentent beaucoup plus qu'un programme de recherche : elles déterminent notre capacité à organiser les connaissances, à leur donner un sens, à augmenter nos facultés de prise de décision et de contrôle des systèmes. Et notamment à tirer de la valeur des données. L'intelligence artificielle est donc une des clés du pouvoir de demain dans un monde numérique.

Voilà pourquoi il est d'intérêt général que nous nous saisissions collectivement de cette question. Et que la France et l'Europe puissent faire entendre leur voix. Il est nécessaire de tout faire pour rester indépendants. Or la concurrence est rude : les États-Unis et la Chine sont à la pointe de ces technologies et leurs investissements dépassent largement ceux consentis en Europe. Le Canada, le Royaume-Uni et, tout particulièrement, Israël, tiennent également une place essentielle dans cet écosystème naissant. Parce qu'à bien des égards, la France et l'Europe peuvent déjà faire figure de « colonies numériques »<sup>2</sup>, il est nécessaire de ne céder à aucune forme de déterminisme, en proposant une réponse coordonnée au niveau européen.

2. Cette expression, traduite de l'anglais « cybercolonization », est issue d'un rapport d'information de Catherine MORIN-DESAILLY fait au nom de la commission des affaires européennes en 2013 (« L'Union européenne, colonie du monde numérique ? »).

C'est pourquoi le rôle de l'État doit être réaffirmé : le jeu du marché seul montre ses limites pour assurer une véritable politique d'indépendance. De plus les règles qui régissent les échanges internationaux et l'ouverture des marchés intérieurs ne servent pas toujours les intérêts économiques des États européens – qui l'appliquent trop souvent à sens unique. Plus que jamais, l'État doit donner un sens au développement de l'intelligence artificielle. Donner un sens, c'est-à-dire donner un cap, une signification et des explications. Voilà l'objectif de ce rapport.

Donner un sens, c'est tout d'abord donner un cap. C'est l'objectif de la structuration proposée pour la politique industrielle : quatre secteurs prioritaires sont définis, la santé, l'écologie, les transports-mobilités et la défense-sécurité. Ces secteurs présentent plusieurs caractéristiques : ils sont au service de l'intérêt général et des grands

défis de notre époque, ils peuvent constituer un avantage comparatif de la France et de l'Europe et ils ont besoin d'une interven-

### **Donner un sens, c'est-à-dire donner un cap, une signification et des explications. Voilà l'objectif de ce rapport**

tion de l'État pour se structurer. Le développement de ces secteurs se fera via des concours d'innovation spécifiques et précis, qui fixeront les objectifs prioritaires, mais également grâce une politique offensive de la donnée. Les données, au cœur du développement de l'IA, bénéficient aujourd'hui souvent à une poignée de très grands acteurs, qui tendent à enfermer les capacités d'innovation dans les limites de leurs entreprises toujours plus puissantes. Ce n'est qu'au prix d'une plus grande circulation de ces données, pour en faire bénéficier les pouvoirs publics, mais aussi les acteurs économiques les plus petits, qu'il sera possible de rééquilibrer les rapports de forces.

La France tient une place décisive dans la recherche en IA : des chercheurs français ont participé à fonder l'IA moderne et l'école mathématique et informatique française rayonne dans le monde entier. Néanmoins l'hémorragie est toujours plus importante : chaque semaine des chercheurs sont recrutés par les entreprises privées et souvent étrangères et quittent les laboratoires publics. Il faut donc redonner à la recherche publique les moyens de ses ambitions, au cœur d'un dispositif allant de la formation au transfert et à l'innovation.

Enfin le développement économique du secteur de l'intelligence artificielle doit mettre en son cœur la préoccupation écologique. En tant que secteur, comme évoqué plus haut, c'est essentiel : les innovations en IA pourront servir à optimiser les consommations d'énergie et le recyclage et à mieux comprendre les effets de l'activité humaine sur l'environnement. Mais il s'agit également de veiller à ce que l'intelligence artificielle que nous développons soit la plus économe en énergie et en ressources.

Donner un sens, c'est également donner une signification. L'intelligence artificielle est loin d'être une fin en soi et son développement doit prendre en compte plusieurs aspects. Tout d'abord la nécessité de penser les modes de complémentarité entre l'humain et les systèmes intelligents. Que ce soit au niveau individuel ou collectif, cette complémentarité peut prendre plusieurs formes et peut être aliénante comme libératrice. Au cœur du développement de l'IA doit résider la nécessité de mettre en œuvre une complémentarité qui soit capacitante, en ce qu'elle permet de désautomatiser les tâches humaines. Pour favoriser la transition des tâches et des métiers dans ce sens, des expérimentations devront être mises en place sur l'ensemble des territoires, notamment à destination des populations les plus touchées par l'automatisation.

# Introduction

Dans un monde marqué par les inégalités, l'intelligence artificielle ne doit pas conduire à renforcer les phénomènes d'exclusion et la concentration de la valeur. En matière d'IA, la politique d'inclusion doit ainsi revêtir un double objectif : s'assurer que le développement de ces technologies ne contribue pas à accroître les inégalités sociales et économiques ; et s'appuyer sur l'IA pour effectivement les réduire. Plutôt que de fragiliser nos trajectoires individuelles et nos systèmes de solidarités, l'IA doit prioritairement nous aider à activer nos droits fondamentaux, augmenter le lien social et renforcer les solidarités. La mixité doit être également au cœur des priorités : la situation est alarmante dans les filières numériques, tant les femmes sont peu représentées. Les algorithmes peuvent en outre reproduire des biais sexistes.

Enfin une société algorithmique ne doit pas être une société de boîtes noires : l'intelligence artificielle va être amenée à jouer un rôle essentiel dans des domaines aussi variés que cruciaux (santé, banque, logement,...) et le risque de reproduire des discriminations existantes ou d'en produire de nouvelles est important. À ce risque s'en ajoute un autre : la normalisation diffuse des comportements que pourrait introduire le développement généralisé d'algorithmes d'intelligence artificielle. Il doit être possible d'ouvrir les boîtes noires, mais également de réfléchir en amont aux enjeux éthiques que les algorithmes d'intelligence artificielle peuvent soulever.

Donner un sens c'est enfin expliquer : expliquer ces technologies à l'opinion pour la démystifier – le rôle des médias est de ce point de vue primordial –, mais aussi expliquer l'intelligence artificielle en elle-même en développant les recherches sur l'explicabilité. Les spécialistes de l'IA eux-mêmes affirment souvent que des progrès importants peuvent être faits sur ce sujet.

C'est de manière plus générale une réflexion collective qui doit être menée sur ces technologies : la constante accélération des rythmes de déploiement ne doit pas empêcher une discussion politique sur les objectifs que nous poursuivons et leur bien-fondé.

Le cœur du document aborde successivement différentes facettes de l'IA : politique économique, recherche, emploi, éthique, cohésion sociale. L'ensemble forme un tout et passe en revue des actions qui ne font sens que quand elles sont prises ensemble. On trouvera d'ailleurs de nombreuses passerelles entre ces parties.

L'IA touche tous les secteurs (y compris ceux où on ne l'attend pas forcément, comme le sport ou la culture) mais il ne nous a pas semblé souhaitable, sauf à produire un document indigeste, de les passer tous en revue ; cependant nous avons ressenti le besoin de rédiger cinq annexes insistant sur des domaines d'intérêt particulier : éducation, santé, agriculture, transport, défense et sécurité.

## Le futur de l'intelligence artificielle

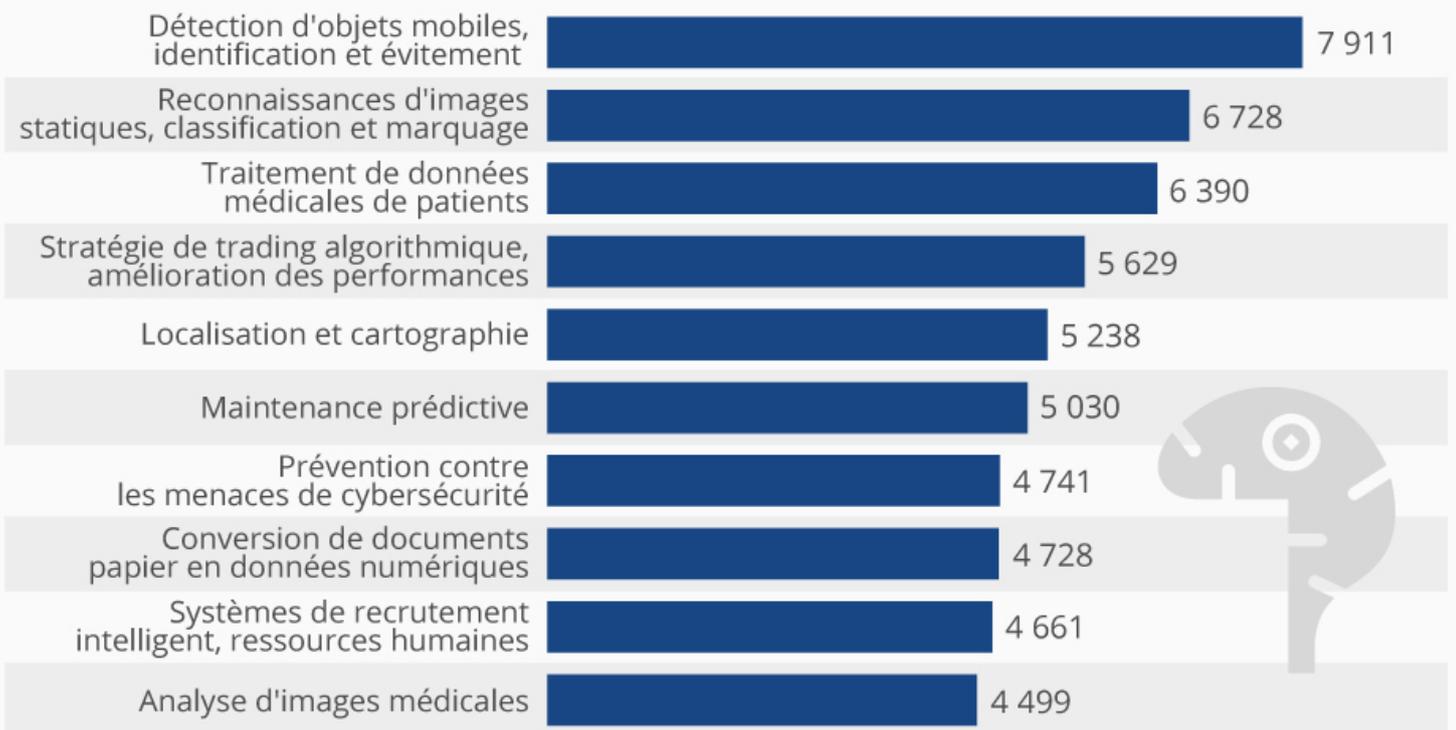
### Applications de l'IA

D'après les données de Tractica, que l'on peut retrouver dans le dossier Statista "Artificial Intelligence (AI)", c'est le segment de la détection et identification d'objets mobiles (notamment lié à la conduite autonome) qui semble le plus prometteur à l'horizon 2025. Ainsi, ce secteur pourrait générer un chiffre d'affaires cumulé de près de 8 milliards d'euros de 2016 à 2025.

Les autres applications de l'intelligence artificielle qui ont le vent en poupe concernent la reconnaissance et la classification d'images statiques (reconnaissance faciale, traitement d'images satellites, etc.) mais aussi la finance (trading algorithmique) ou encore le secteur médical (traitement de données de patients, analyse d'images).

## Le futur de l'intelligence artificielle

Estimation du chiffre d'affaires mondial cumulé de 2016 à 2025, en millions d'euros



# L'Europe et la course à l'Intelligence Artificielle

Les acteurs étatiques considèrent de plus en plus l'intelligence artificielle (IA) comme une technologie clé et une ressource stratégique. Où en est l'Europe sur cette question? Quels sont notamment les choix de la France et de l'Allemagne, deux pays très actifs en la matière? Le développement et l'utilisation d'une IA mieux réglementée et soutenue par des valeurs éthiques offre un créneau.

Par Fabien Merz

Avec ses possibilités d'application extrêmement variées dans d'innombrables domaines, l'IA est d'ores et déjà considérée comme la «nouvelle électricité». Selon la société de conseil en affaires et en stratégie McKinsey, l'IA pourrait générer d'ici 2030 un produit économique supplémentaire d'environ 13 milliards de dollars US, soit une hausse du PIB mondial d'à peu près 1,2% par an. Outre une contribution significative à la prospérité économique, on peut supposer que l'IA jouera également un rôle important dans le domaine de la sécurité nationale et des méthodes de guerre.

Si les progrès de l'IA sont vecteurs d'opportunité, ils présentent également des risques et posent un certain nombre des questions fondamentales. Au niveau de la société dans son ensemble, les bouleversements du marché du travail, avec les conséquences socioéconomiques qui en découlent, ainsi que l'accès aux données et leur gestion, avec le problème connexe de la confidentialité comptent parmi les enjeux à prendre en considération. S'ajoutent à cela des questions éthiques relatives, par exemple, à la transparence des décisions prises par des systèmes d'IA. Dans le domaine de la sécurité et de la politique étrangère, des questions liées aux inégalités mondiales se posent. L'IA pourrait, par exemple, creuser d'avantage le fossé militaire et économique entre les pays riches et les pays en voie de développement.



Le président français Emmanuel Macron a présenté la stratégie d'IA française lors d'un événement à Paris le 29 mars 2018. *Etienne Laurent / Reuters*

Par conséquent, de plus en plus de pays anticipent les répercussions globales de l'IA sur les rapports de force économiques et politiques et prennent des mesures pour acquérir une position avantageuse dans ce domaine. À ce jour, une trentaine d'États ont élaboré des stratégies nationales d'IA afin d'en tirer le meilleur parti possible tout en atténuant ses externalités négatives. D'autres ont également annoncé des stratégies dans ce sens. On constate que les approches choisies sont extrêmement variées, avec des orientations et des priorités parfois très différentes. Certaines d'entre-elles mettent l'accent sur un ou plusieurs sous-domaines, par exemple, la promotion de la

recherche, des talents ou l'appui au secteur privé et à la politique industrielle. D'autres encore adoptent une approche globale allant de la politique industrielle à la gestion des données et aux questions éthiques, en passant par la promotion de la recherche et la participation à l'établissement de normes internationales en matière d'IA.

Même s'ils n'ont pas défini de stratégie nationale d'IA, les États-Unis sont le leader mondial de ce secteur sur la plupart des indicateurs. On estime, par exemple, que près de 40% du capital-risque disponible dans le domaine de l'IA et près de la moitié des start-ups de la filière sont issus du mar-

éléments, nous mettrons en évidence le positionnement de ces acteurs européens clés dans la course mondiale à l'IA.

### La stratégie-cadre européenne

Par rapport aux États-Unis, l'UE est à la traîne en termes de volumes d'investissement dans l'IA, de la structure industrielle (nombres d'entreprises actives dans le secteur de l'IA), de capacité d'innovation (nombre de brevets déposés) et de transferts vers des applications pratiques, y compris dans le domaine de la commercialisation. L'Europe présente également un désavantage comparatif non seulement vis-à-vis des États-Unis mais aussi de la Chine en ce qui concerne la disponibilité des données indispensables afin de perfectionner l'IA. Alors que les acteurs européens n'ont pas accès à des pools de données à grande échelle en raison, entre autres, de la relative fragmentation de l'Europe, les modèles économiques des géants américains leur ont permis d'accumuler de grandes quantités de données. Les acteurs chinois, quant à eux, bénéficient d'une réglementation moins stricte en matière de protection des données et d'un transfert de données favorisé par l'État chinois.

Les seuls domaines dans lesquels l'Europe est apte à rivaliser avec les leaders mondiaux sont la recherche et l'enseignement. En 2017, par exemple, 28% des articles académiques sur l'IA enregistrés dans la base de données Scopus ont été publiés par des auteurs affiliés à des institutions européennes. Viennent ensuite la Chine (25%) et les États-Unis (17%). Mais malgré un environnement propice à la recherche et à l'enseignement, les pays de l'UE font face à un exode croissant des talents dû à la concurrence internationale.

Au regard de cette situation, les États membres de l'UE ont signé en avril 2018 une Déclaration de coopération sur l'intelligence artificielle dans laquelle ils conviennent de relever ensemble les principaux défis posés par l'évolution de l'IA. Ils s'engagent notamment à renforcer la compétitivité technologique et industrielle de l'UE et à faciliter l'accès aux données. Ils s'engagent également de inscrire l'IA dans un cadre éthique et juridique approprié, en accord avec les valeurs fondamentales de l'UE.

Sur la base de cette déclaration, la Commission Européenne (CE), à la demande du Conseil européen, a publié fin avril 2018 une communication sur l'intelligence artificielle. Le document, que l'on peut voir

comme l'esquisse d'une stratégie-cadre européenne, reprend les objectifs formulés dans la déclaration et présente des moyens pour les atteindre. Par exemple, des investissements dans la recherche et l'industrie devraient renforcer la compétitivité technologique et industrielle de l'UE. L'adaptation du système éducatif et des programmes de promotion de talents, ainsi que le soutien aux transitions sur le marché du travail, aideront à préparer la société aux changements socioéconomiques suscités par l'IA.

La stratégie-cadre européenne se fixe également pour objectif, en coopération avec les acteurs nationaux et privés, de dégager jusqu'à 20 milliards d'euros d'investissements dans le secteur public et privé d'ici la fin 2020. Dans cette logique, la CE s'engage à porter ses investissements dans le programme de recherche et d'innovation «Horizon 2020» à 1,5 milliard d'euros jusqu'en 2020. De plus, le Fonds européen pour les investissements stratégiques (EFSD) sera également mobilisé à hauteur de plus de 500 millions d'euros d'ici 2020 pour fournir aux entreprises et aux start-ups des ressources supplémentaires à investir dans l'IA. Tout en respectant le cadre réglementaire établi par l'UE, y compris le règlement général sur la protection des données (RGPD), les données collectées par les pouvoirs publics devraient être partagées entre les États membres et mises à la disposition de certains acteurs clefs. Dans le souci d'établir un cadre éthique approprié qui traite, par exemple, les questions liées à la transparence, le groupe d'experts de haut niveau sur l'IA nommé par la CE à élaborer des lignes directrices en matière d'éthique pour l'intelligence artificielle.

La stratégie-cadre de l'UE choisit donc délibérément une approche globale qui va de la promotion de la recherche et de l'industrie en passant à l'accès aux données ainsi qu'à une attention accrue au cadre éthique et réglementaire. Dans le contexte de la course mondiale à l'IA, l'objectif est de préserver la compétitivité de l'UE en matière de recherche et d'enseignement. Mais parallèlement, il s'agit de créer les conditions qui permettront de réduire l'écart avec les États-Unis et la Chine dans les domaines du financement, de la structure industrielle, du transfert, des résultats de la recherche en applications pratiques ainsi qu'en terme accès aux données. L'accent est également mis sur le cadre réglementaire et les aspects éthiques du développement et de l'utilisation de l'IA – des critères auxquels les modèles de développement chinois et, dans une moindre mesure, amé-

ché américain. En outre, les États-Unis sont le leader incontesté dans le transfert des résultats issues de la recherche vers des applications pratiques et commerciales. Cela est dû, entre autres, à l'existence de structures de coopération établies entre le gouvernement, le secteur privé et la recherche, d'un écosystème d'IA correspondant et de l'existence de géants technologiques très actifs dans le domaine de l'IA tel qu'Alphabet, Facebook ou Amazon.

Les États-Unis et la Chine se livrent depuis peu à une compétition mondiale pour la suprématie dans le domaine de l'IA. Avec son plan de développement de la prochaine génération d'intelligence artificielle, une stratégie globale largement pilotée par l'État et présentée en 2017, la Chine prévoit de rattraper les États-Unis d'ici 2020, de les dépasser d'ici 2025 et de devenir leader mondial dans ce secteur d'ici 2030 (voir [l'analyse du CSS n° 220](#)).

Mais qu'en est-il du continent européen dans cette course mondiale pour la suprématie en matière d'IA? Quelles actions sont entreprises au niveau de l'Union Européenne (UE) pour ne pas se faire distancer dans ce domaine? Comment se positionnent deux des États membres les plus puissants sur le plan économique, à savoir la France et l'Allemagne? Cette analyse se penchera sur la situation actuelle et la teneur des stratégies formulées par l'UE, la France et l'Allemagne. Sur la base de ces

ricain, accordent moins d'importance. Aux États-Unis, ce phénomène tient notamment du fait que le développement et l'utilisation de l'IA relève d'acteurs privés. Par conséquent, les impératifs de l'économie de marché peuvent plus facilement l'emporter sur les considérations éthiques et les aspects de protection de données. Quant au modèle chinois, il repose sur un système de valeurs différent qui fait que les aspects de protection de données et les questions éthiques tel que la transparence de l'IA sont reléguées au second plan. Avec sa stratégie-cadre en matière d'IA, l'UE cherche donc à se positionner sur un créneau bien spécifique: le développement et l'utilisation d'IA mieux réglementée et sous-tendue par des valeurs éthiques.

### La stratégie d'IA française

Dans le contexte de la course mondiale à l'IA, la situation de la France reflète la position actuelle de l'UE. La France est bien placée dans le domaine de la recherche et de l'enseignement, mais elle est à la traîne des États-Unis et de la Chine par rapport aux opportunités de financement pour la filière de l'IA. Cette situation se traduit également dans la structure industrielle du pays qui présente, en proportion, moins de sociétés d'IA que les États-Unis et la Chine. En outre, pour les mêmes raisons

## L'UE cherche à se positionner sur un créneau bien spécifique: celui d'une IA mieux réglementée et sous-tendue par des valeurs éthiques.

que l'UE dans son ensemble, la France possède un désavantage par rapport aux États-Unis et à la Chine en matière de disponibilité des données et souffre de plus en plus de la concurrence internationale afin d'attirer et retenir les meilleurs talents.

Dans ce contexte, le gouvernement Macron a chargé, en septembre 2017, un groupe d'experts de produire un rapport sur les opportunités et les risques de l'IA pour la France. Certaines des propositions de ce rapport, rédigé sous la direction du député Cédric Villani, ont été reprises par le président Macron dans un discours d'impulsion prononcé en mars 2018. Les éléments clés de ce discours peuvent être considérés comme les piliers de la stratégie française en matière d'IA. La stratégie d'IA esquissée dans le discours de Macron semble suivre l'idée d'une approche globale et les axes prioritaires définis par la stratégie-

cadre de l'UE, notamment la promotion de la recherche et des talents. Ainsi, 665 millions des 1,5 milliard d'euros d'investissements prévus jusqu'en 2022 dans le cadre de la stratégie française seront investis dans la recherche et l'éducation. Des investissements supplémentaires du secteur privé devraient porter le volume total des investissements dans la recherche sur l'IA à 1 milliard d'euros. Il est ainsi prévu de créer plusieurs centres universitaires interdisciplinaires dans le domaine de l'IA. Une quarantaine de nouvelles chaires d'IA seront également ouvertes et le nombre de places de doctorat en IA sera fortement augmenté. Afin de retenir les talents, la France compte se donner la possibilité de proposer des salaires compétitifs dans la recherche publique ainsi que de permettre aux chercheurs de cumuler les mandats de recherche publique avec des engagements dans le secteur privé.

La stratégie esquissée par Emmanuel Macron comporte également un volet consacré à la promotion industrielle. Il s'agit de faciliter l'accès des petites et moyennes entreprises d'IA au capital-risque. La stratégie française se penche aussi sur la question de la disponibilité des données. Les informations collectées par les autorités françaises, telles que les données médicales, seront, tout en respectant le cadre réglementaire de l'UE, entre autres le RGPD, mises à disposition de certains acteurs publics et privés. L'accent sera également placé sur la coopération dans le cadre de l'UE, surtout avec l'Allemagne. Par exemple, des initiatives seront

déployées pour mettre en réseau les centres de recherche et permettre la fourniture de données dans le respect des règles de confidentialité. Par ailleurs, comme proposé par la stratégie cadre de l'UE, la stratégie française accorde également une place importante aux questions éthiques liées au développement et à l'utilisation de l'IA. Ainsi, les préoccupations concernant la transparence de l'IA deviendront un axe de recherche prioritaire.

[...]

### Un modèle alternatif?

[...]

Le modèle européen vise à fournir un cadre éthique et réglementaire qui servira à guider, accompagner et former les avancées dans le domaine de l'AI. Les approches américaines et chinoises ont quant à elles tendance à donner plus libre cours aux avancées dans le domaine de l'IA tandis que les aspect réglementaires et éthiques se font tendanciellement plutôt en aval. Ces deux approches différentes ont des avantages et des désavantages respectifs. A court terme, les modèles américains et chinois pourraient s'avérer plus avantageux dans le contexte de compétition globale à l'IA. Cependant, il reste à voir si ces approches s'avéreront viables à moyen et à long terme. L'avenir montrera si l'approche européenne axée sur un IA mieux réglementée et sous-tendue par des valeurs éthiques réussira à s'affirmer.

**Fabien Merz** est chercheur au sein du think tank du Center for Security Studies (CSS) à l'ETH de Zurich. Il est entre autres auteur de «Cybersicherheit: Was lässt sich von Israel lernen?» (2018).

## Intelligence artificielle : découvrez les 15 nouveaux projets sélectionnés

Le 2 avril 2019, la DINSIC et la DITP ont lancé un deuxième appel à manifestation d'intérêt (AMI) pour expérimenter l'intelligence artificielle (IA) dans l'administration — d'où le nom du projet : AMI IA 2. Un jury d'experts s'est réuni le 4 juillet. Le jury a sélectionné 15 projets, qui seront accompagnés par la DITP et la DINSIC, pour développer un prototype fonctionnel et appréhender les enjeux de la transformation des métiers associés à l'IA, au cours de l'année à venir.

### 11 projets bénéficieront de l'accompagnement technique et métier

**Direction générale de la santé. — Optimiser le traitement des signalements des événements sanitaires indésirables :** À partir des données issues de l'ensemble des signalements des événements sanitaires indésirables, l'IA permettra de classer les signalements, d'extraire les concepts métiers pertinents et de détecter des anomalies dans les signalements. [...]

**Direction générale des collectivités locales. — Développer l'intelligence artificielle dans le contrôle de légalité dématérialisé :** À partir des données de l'application @ctes (Actes), l'IA permettra de simplifier le travail des agents en préfecture, en triant automatiquement les actes transmissibles et non transmissibles et en détectant les informations à contrôler en priorité.

**Direction générale de la gendarmerie nationale. — Améliorer le système de pré-plainte en ligne**  
En détectant automatiquement les infractions à partir des pré-plaintes en ligne et en identifiant les questions complémentaires à poser à l'utilisateur, l'IA permettra de transformer les pré-plaintes en ligne en plaintes qualifiées, ce qui permettra un gain de temps à la fois pour les plaignants et les agents de la gendarmerie. [...]

**Chambres des métiers et de l'artisanat de Nouvelle Aquitaine. — Développer un outil d'aide à l'implantation des entreprises artisanales :** À partir de données sur les entreprises artisanales et de données sur la population, l'IA permettra de déterminer les chances de succès d'un artisan sur un territoire et de mieux conseiller les artisans dans leur choix d'implantation.

**Institut national de l'environnement industriel et des risques. — Identifier les molécules contaminant l'environnement et profiler les sources de pollutions :** À partir des résultats d'analyses chimiques, l'IA permettra d'objectiver des ressemblances et des différences entre des spectres moléculaires pour mieux identifier les molécules contaminant l'environnement et mieux détecter des molécules polluantes. [...]

**Conseil d'État. — Identifier automatiquement les séries de contentieux faisant appel à une décision commune :** À partir de l'ensemble des requêtes auprès des tribunaux administratifs, l'IA permettra de regrouper les contentieux pour détecter automatiquement des séries et d'améliorer l'efficacité de la justice administrative.

### 4 projets bénéficieront de l'accompagnement scientifique

Les 4 projets suivants sont considérés comme plus complexes d'un point de vue technique et bénéficieront d'un accompagnement scientifique en collaboration avec un institut de recherche.

**Direction générale de la concurrence, de la consommation et de la répression des fraudes. — Détecter les clauses et pratiques abusives dans les contrats, devis et factures :** À partir des comptes rendus et pièces récupérées par les enquêteurs de la DGCCRF lors des contrôles et des décisions, avis et recommandations de la commission des clauses abusives, l'IA permettra de détecter la présence et le contenu des mentions obligatoires, d'identifier les thématiques des clauses des contrats et le cas échéant d'identifier les clauses abusives afin d'assurer une meilleure protection des consommateurs et une meilleure régulation concurrentielle des marchés.[...]

**Cour de cassation. — Identifier les divergences de jurisprudence :** À partir des données de jurisprudence de la Cour de cassation (Jurinet) et de cours d'appel (JuriCA), l'IA permettra de rapprocher les décisions traitant du même thème et de détecter des interprétations divergentes de la loi afin de garantir une interprétation uniforme de la loi.

**Service hydrographique et océanographique de la marine. — Améliorer la cartographie du littoral :** À partir des mesures de télédétection par laser (lidar) prises sur le littoral, l'IA permettra d'améliorer la production du référentiel du littoral en automatisant le travail de correction des agents. [...]

lemonde.fr  
Cécile Prudhomme  
21 juin 2019

## Intelligence artificielle : des consommateurs sous surveillance

*Paiement par reconnaissance faciale, essor des magasins sans caisse, publicités ultraciblées... Le développement de l'informatique « intelligente » ouvre la voie à de nombreuses innovations, parfois passionnantes, souvent effrayantes.*

Payer ses courses, simple comme un regard ? Le paiement par reconnaissance faciale qui se développe rapidement en Chine pourrait débarquer prochainement en France. Au Lab Carrefour-Google, sis dans le 13<sup>e</sup> arrondissement de Paris, on y travaille d'arrache-pied.

Les deux groupes ont réuni des experts de l'intelligence artificielle (IA). Ils s'appuient sur la technologie utilisée par le distributeur français dans l'empire du Milieu, qui permet de payer au moyen de l'application WeChat grâce à la reconnaissance faciale et aux quatre derniers chiffres de son numéro de téléphone.

« On peut imaginer qu'un système similaire soit testé à la rentrée, piloté par Carrefour avec des partenaires », souffle Pierre Blanc, un ancien cadre de chez HP et IBM, chargé des innovations au sein de l'enseigne de la grande distribution. Encore faut-il régler les nombreuses problématiques liées au stockage des données personnelles. « En Chine, cela ne pose pas de problème de conserver un visage sur un serveur central », fait-il remarquer.

Le groupe de restauration collective Sodexo a, lui, investi dans la société chinoise Aeye-Go, dont la technologie permet d'identifier tous les aliments posés sur un plateau et de payer par le biais de la reconnaissance faciale. Elle a déjà installé sa solution dans une cinquantaine de cantines scolaires, d'hôpitaux et de restaurants d'entreprise en Chine, dont le siège de Microsoft à Shanghai. « Nous discutons avec Sodexo, souligne Ji Zhu, directeur technique chez Aeye-Go. La solution pourrait arriver cette année en France. »

### Inciter à l'acte d'achat

En nourrissant les machines de multiples programmes et informations, l'IA est capable de reconnaître un visage ou un produit, et de l'associer à un ensemble de données.

« Il y a encore cinq ans, les distributeurs comme les marques de prêt-à-porter travaillaient sur la RFID [pour radio frequency identification, c'est-à-dire "identification par radiofréquence"] pour

stocker des informations et communiquer avec les mobiles des acheteurs, rappelle Guillaume Rio, de l'Echangeur, le centre d'analyse de BNP Paribas Personal Finance dévolu au commerce. Mais grâce à l'IA, la reconnaissance d'image se développe plus rapidement, et c'est cela qui permettra l'essor de magasins autonomes. »

Ceux-ci commencent déjà à se développer. Dans la continuité des quelques Amazon Go que le géant américain a implantés aux Etats-Unis, Auchan a installé sur le parking de son siège social, à Villeneuve-d'Ascq (Nord), un magasin conteneur sans personnel, à l'image de ce qu'il fait en Chine, pour le tester auprès de ses collaborateurs. Carrefour, de son côté, a ouvert un labstore au siège de Massy (Essonne), où il expérimente ses innovations. « Pour le moment, on y teste le suivi automatique du panier et le paiement par reconnaissance faciale, de manière à fiabiliser la solution avant des tests grandeur nature à la rentrée », indique M. Blanc.

L'IA et toutes ses possibilités déferlent sur le secteur. Distributeurs et marques se sont emparés de ces sujets informatiques avec un double objectif. Améliorer leur efficacité, tout d'abord, à l'exemple de Carrefour, qui a fait appel au géant du logiciel SAS, dont la solution permet de prévoir la demande et donc d'affiner en conséquence ses approvisionnements sur toute la chaîne logistique. Gonfler leur chiffre d'affaires, ensuite, en brassant les données pour mieux cibler les consommateurs et les inciter à passer à l'acte... d'achat.

### Un rôle prépondérant dans le secteur publicitaire

En coulisse, la plupart travaillent avec des start-up. « Beaucoup de marques de parfum utilisent notre "Affect-Tag" [un bracelet connecté capable d'analyser les émotions, en mesurant la température et la conductivité de la peau, le rythme cardiaque...] pour tester des produits, et voir si l'émotion ressentie dure dans le temps, ou pour des campagnes marketing », expliquent Antoine Lempereur et Hugo Six, développeurs chez Neotrope.

Ils imaginent déjà des utilisations possibles de leur technologie dans les voitures connectées qui, en détectant un conducteur stressé, pourrait tamiser la lumière dans le véhicule, lui proposer des musiques relaxantes ou une séance de méditation avant de prendre la route.

En amont, l'IA joue un rôle prépondérant dans le secteur publicitaire. La jeune pousse britannique Mirriad s'est ainsi fait une spécialité de l'insertion de publicité dans les séries. « *On scanne l'épisode de la série "Demain nous appartient" [sur TF1] avant sa diffusion, et l'ordinateur trouve les bons endroits pour insérer des marques* », note Julia Leadbeater, responsable du développement commercial de Mirriad, qui a signé un partenariat avec la première chaîne en 2018.

Ici, une bouteille d'une grande marque de soda sera ajoutée sur une table ; là sera incrustée la couverture d'un magazine présentant une publicité pour une marque automobile. « *On discute avec France Télévisions et M6 pour nouer des partenariats similaires pour leurs fictions* », ajoute M<sup>me</sup> Leadbeater.

### **Exploiter les données des tickets de caisse**

Son ambition est de pouvoir répliquer un jour en France l'accord noué avec le géant chinois Tencent, qui produit aussi des séries. « *On leur renvoie cinq versions avec cinq marques différentes et eux choisissent celle qui sera visionnée par rapport au profil de leur spectateur.* » Ainsi, en Chine, lors d'une récente campagne pour Coca-Cola, les femmes avaient regardé des publicités pour le Coca Light et les hommes pour du Coca Zero.

Le développement de l'informatique « intelligente » a aussi entraîné une automatisation croissante de l'allocation des budgets publicitaires, selon l'heure, le type de support, le public visé...

Sébastien Garcin, ancien directeur marketing de L'Oréal, a conçu avec sa start-up Yzr un outil pour savoir quels étaient les leviers de communication les plus rentables. « *On entre dans la machine toutes les publicités qui ont été faites sur les trois dernières années, la quantité de produits mis sur le marché, les promotions, le prix, les chiffres des ventes. Chaque média publicitaire [télévision, presse magazine...] a ses propres règles. La machine donnera la*

*rentabilité pour chacun d'entre eux ainsi que la répartition optimale des budgets et ce, en temps réel* », détaille M. Garcin.

Les distributeurs n'ont pas attendu les start-up pour exploiter l'ensemble des données issues de leurs tickets de caisse. La société 3W. relevanC, du groupe Casino, agrège, brasse et utilise l'ensemble des données des factures des magasins et sites Internet du groupe (Monoprix, Leader Price, Franprix, Cdiscount, Sarenza...) et fait office de régie publicitaire pour les marques. Elle s'est créé une base de données de 30 millions de consommateurs, de laquelle elle a extrait un panel d'un million de personnes qu'elle monétise auprès des industriels de produits de grande consommation, que ce soit pour le lancement d'une nouvelle gamme, une campagne de publicité ciblée ou pour connaître un retour plus précis de leurs ventes en magasin.

### **Deviner l'humeur des clients**

La carte de fidélité est le premier espion du dispositif. Elle recèle une mine d'informations, dont l'adresse e-mail, qui permet de déposer des cookies lorsque son détenteur ouvre les courriels promotionnels.

« *Quand on envoie de la publicité, on est capable de la cibler sur une catégorie sociodémographique : des hommes de plus de 25 ans, ou vivant en Ile-de-France, ou encore des acheteurs de thé ou de jus de fruits,* explicitent Blandine Multrier et Christophe Blot, codirecteurs généraux de 3W. relevanC. *L'IA permet d'améliorer l'efficacité de notre base de données, de rapprocher des profils similaires de consommateurs.* » C'est ainsi qu'un ménage venant de s'acheter une poussette sur Cdiscount, identifié comme jeunes parents, verra ensuite s'afficher des publicités pour du lait infantile lorsqu'il surfera sur le Web.

Aux Etats-Unis, les supermarchés Kroger et Walgreens vont plus loin, en testant des caméras dans les portes en verre réfrigérées, au sein de quelques magasins pilotes. Il s'agit de deviner l'âge, le genre et l'humeur de leurs clients pendant qu'ils déambulent dans les rayons pour leur montrer des annonces ciblées en temps réel sur des écrans vidéo en magasin...

# LE MYTHE DE LA SINGULARITÉ

Faut-il craindre l'intelligence artificielle ? (*extrait*)

JEAN-GABRIEL GANASCIA

[...]

AUTONOMIE

## Apprentissage machine

Avant même que le terme d'intelligence artificielle n'ait été inventé, Alan Turing<sup>74</sup> évoque le rôle central que devrait prendre l'apprentissage automatique dans les deux articles qu'il a écrits en 1948 et en 1950 sur l'intelligence des machines : pour lui, afin qu'une machine « pense », ou plus exactement simule le comportement d'un être pensant, il faudrait qu'elle possède un grand nombre de connaissances sur le monde qui nous environne et sur la réalité sociale. Or, le transfert de ces connaissances de bon sens aux machines apparaît extrêmement fastidieux, sans compter que cette tâche semble infinie, car on ne saurait assigner de bornes claires au savoir humain. D'après Turing, plutôt que de tout traduire explicitement dans un langage de programmation, il serait préférable de doter les machines de capacités à apprendre, c'est-à-dire à acquérir d'elles-mêmes compétences, savoir et savoir-faire à partir d'observations sur le monde extérieur, sur leur place dans ce monde et sur leur propre comportement.

Plus tard, en 1955, au moment de la naissance officielle de l'intelligence artificielle, les pionniers de cette discipline scientifique mentionnent de nouveau l'apprentissage machine<sup>75</sup>, son importance et le rôle clef qu'il sera amené à jouer. Et, depuis soixante ans que l'intelligence artificielle existe, une grande partie des efforts poursuivis a porté sur l'amélioration des techniques d'apprentissage automatique.

De très nombreuses approches ont été explorées ; certaines s'inspirent de la psychologie humaine pour tenter de reproduire nos facultés d'apprentissage sur des machines ; plusieurs d'entre elles ont été simulées, par exemple la mémoire associative ou l'habituation par renforcement ; d'autres miment la plasticité synaptique, c'est-à-dire l'évolution des connexions entre les neurones du cerveau ; d'autres encore se fondent sur un parallèle avec l'évolution des espèces ou avec la maturation des idées dans la société ou encore avec l'auto-organisation des insectes sociaux que sont les abeilles, les termites ou les fourmis, etc.

LE MYTHE DE LA SINGULARITÉ

Avec le temps, on explora en détail ces différentes métaphores, on implanta chacune d'entre elles, on les évalua, on les compara et on les formalisa avec des appareils mathématiques élaborés. Sans entrer dans les détails, citons parmi les approches les plus communément répandues les réseaux de neurones formels, les algorithmes génétiques, la construction d'arbres de décision, les  $k$  plus proches voisins, l'apprentissage bayésien, les machines à noyaux, les séparateurs à vaste marge (*support vector machines*, SVM en anglais), l'apprentissage profond (*deep learning* en anglais)... Et, il en est bien d'autres. Aujourd'hui, parvenues à maturité, ces techniques sont à la fois fort bien maîtrisées et très utilisées dans tous les domaines d'application de l'intelligence artificielle.

## Masses de données

À cela, il faut ajouter que l'on amasse des quantités de plus en plus considérables d'information recueillies soit automatiquement, par toutes sortes de capteurs, sismographes, radiotélescopes, cardiofréquencesmètres, caméras, microphones, objets connectés en tous genres, etc., soit manuellement, de façon distribuée, par *crowdsourcing*\*. Et ces données sont ensuite colligées, plus ou moins automatiquement, par le truchement de la toile, pour constituer des amoncellements plus grands encore. Ainsi, on rassemble tous les gazouillis (traduction littérale de *tweet* en anglais) que nous émettons sur les sites de microblogage, toutes les requêtes que nous soumettons aux moteurs de recherche lors de nos navigations sur internet et toutes les annotations et tous les commentaires que l'on dépose sur les réseaux sociaux. Certains logiciels comme Waze enregistrent de façon systématique nos déplacements et notre vitesse instantanée puis centralisent ces informations pour en

\* Forgé sur le modèle d'*outsourcing* (« externalisation de la production ») avec le mot *crowd* qui signifie « foule », le *crowdsourcing* désigne des modalités de travail collaboratif à grande échelle sur le web.

déduire le flux de la circulation automobile et offrir, en retour, à chacun, l'itinéraire optimal. On imagine même enregistrer bientôt en continu nos paramètres physiologiques (battement cardiaque, tension artérielle, sucre, etc.) avec des montres ou des bracelets connectés puis les envoyer directement aux institutions chargées de surveiller notre santé. Toutes ces informations constituent ce que l'on appelle en français des masses de données et en anglais des *Big Data*, littéralement des « grosses données » dont l'accumulation ébahit quelque peu. [...]

Or, aujourd'hui, on parvient à exploiter ces immenses masses de données avec les techniques d'apprentissage machine mentionnées plus haut, de façon à en extraire automatiquement des connaissances. Les performances sont époustouflantes ... Les logiciels de reconnaissance vocale comme Siri d'Apple ou la voiture autonome de Google en font largement usage. Il en va de même de logiciels de reconnaissance des visages qui obtiennent, dans des conditions bien définies, des taux de reconnaissance prodigieusement élevés. En mars 2016, le programme AlphaGo de la société DeepMind l'a même emporté au jeu de go sur Lee Sedol, réputé être l'un des meilleurs joueurs au monde, toujours en faisant appel à des techniques d'apprentissage machine, en particulier à de l'apprentissage profond et à de l'apprentissage par renforcement. Tous ces exemples constituent autant de preuves de l'efficacité de ces techniques. Et, comme nous l'avons déjà mentionné dans le premier chapitre, ce constat sert de justification aux déclarations alarmistes de nombre de scientifiques qui, à l'instar de Stephen Hawking, Frank Wilczek, Stuart Russell, Elon Musk, Bill Gates, etc., prétendent que les ordinateurs deviendront bientôt autonomes et se passeront de nous pour agir, jusqu'à dominer le monde par-devers nous. Doit-on les croire? Y a-t-il là vraiment de quoi s'inquiéter? Dans quelle mesure ce couplage des techniques d'apprentissage automatique avec les masses de données permet-il aux machines d'acquiescer leur autonomie et, ce faisant, de s'affranchir de nous, au point de nous dépasser bientôt ?

### Risques

Avant d'aborder de front ces questions, soulignons que l'apprentissage automatique dote les machines d'une capacité à construire d'elles-mêmes des connaissances et à les utiliser pour se reconfigurer en réécrivant leurs propres programmes. Il en résulte un accroissement impressionnant de leurs performances dont nous tirons les bénéfices dans beaucoup de nos activités quotidiennes. En contrepartie de ces effets positifs, les ordinateurs se reprogram-

ment d'eux-mêmes, sans qu'aucun être humain n'ait ni rédigé, ni relu, ni *a fortiori* vérifié les programmes qui les animent, leurs comportements deviennent de plus en plus difficiles à anticiper. Et cette imprévisibilité apparaît neuve, parce qu'elle ne tient pas uniquement à la difficulté que nous avons à prévoir les réactions des machines dans la précipitation de l'action, du fait de leur rapidité, mais au caractère inédit de leurs comportements qui découlent de l'exécution de programmes construits sur des données qu'aucun homme n'a jamais examinées. Et, même si certains l'avaient tenté, ç'aurait été en vain, car personne ne disposerait des facultés suffisantes pour exploiter de telles quantités dans leur intégralité. En cela, sans aucun doute, ces machines dépassent nos capacités.

Cela justifie pleinement les inquiétudes des chercheurs en intelligence artificielle qui signèrent les deux pétitions parues en 2015, l'une en janvier<sup>76</sup>, sur les risques liés au déploiement d'agents artificiels dotés de capacités d'apprentissage et sur la nécessité de poursuivre des recherches sur leur robustesse, l'autre, quelques mois plus tard, en juillet 2015<sup>77</sup> par le même canal, sur les dangers que feraient courir les drones et autres systèmes d'armes autonomes. À titre personnel, je poursuis des recherches sur ce sujet dans le cadre du projet ANR EthicAA<sup>78</sup> (Ethique & Agents Autonomes) qui fait collaborer des philosophes, des logiciels et des chercheurs spécialisés dans différents domaines de l'intelligence artificielle, raisonnement automatique, systèmes multi-agents, logiques non-standard, modélisation de l'argumentation, etc. Nous cherchons à limiter les capacités d'action des agents artificiels, afin d'assurer qu'ils ne dérogent ni aux règles morales que nous leur donnons, ni aux procédures en vigueur, en dépit des conflits de normes qui apparaissent ici ou là. Dans un second temps, nous souhaiterions concevoir un cadre formel qui permette à des agents humains et artificiels d'interagir entre eux pour décider collectivement de la décision la plus satisfaisante qui respecte à la fois les règles instituées et les arguments sincères dont chacun, hommes ou robots, peut faire état, tenant compte des informations variées dont il dispose.

Répetons-le, compte tenu de la part que prennent d'ores et déjà les robots matériels et les agents virtuels dans la vie quotidienne, et compte tenu de la part grandissante qu'ils prendront très vraisemblablement dans le futur, il faut s'assurer de leur innocuité. Cette question mérite très certainement qu'on lui porte une attention soutenue. Les craintes légitimes tiennent à d'éventuels dysfonctionnements, aux risques d'accidents consécutifs à ces dysfonctionnements et à leurs conséquences dramatiques. Pour autant, doit-on s'inquiéter d'une autonomisation des machines et des agents virtuels qui, soudain, décideraient à notre place, pour satisfaire leurs propres besoins, sans plus nous consulter, ni même obéir à nos injonctions ?

#### Modalités d'apprentissage

Pour répondre à cette question, sans minimiser la puissance des techniques actuelles, reprenons la présentation des algorithmes d'apprentissage amorcée plus haut. Nous en avons énuméré plusieurs, suggérant que notre liste n'était pas exhaustive. Nous n'y reviendrons pas, sauf à noter qu'en dépit de leur variété, ces algorithmes ne relèvent que de trois modalités : l'apprentissage qualifié de « supervisé » parce qu'un professeur instruit la machine en indiquant, pour chaque instance, sa classe ou sa catégorie, l'apprentissage dit « non supervisé », parce qu'il n'y a pas de professeur, et l'apprentissage « par renforcement », qui repose sur un jeu de récompenses positives ou négatives (punitions) consécutives à chaque action, ce qui correspond là encore à un professeur vieille manière, qui tape sur les doigts avec une règle en cas de mauvais comportement et donne des bons points en récompense pour les bonnes réponses.

Les résultats époustouflants mentionnés précédemment tiennent essentiellement aux algorithmes d'apprentissage supervisé et dans une moindre mesure aux algorithmes d'apprentissage par renforcement. Ces derniers ont fait l'objet d'attentions soutenues et de formalisations mathématiques comme l'apprentissage statistique

de Vladimir Vapnik <sup>79</sup> et la théorie formelle de l'apprentissage dont Leslie Valiant est l'un des précurseurs avec l'approche dite PAC (*probably approximately correct*)<sup>80, 81</sup>. Lorsqu'il s'agit de reconnaître des phonèmes dans une séquence sonore, comme pour la dictée vocale, d'identifier des empreintes digitales, des voix, des émotions ou des visages, ou encore de décider si un véhicule doit freiner, accélérer et tourner à gauche ou à droite, ces techniques marchent fort bien, car il suffit de fournir des myriades d'exemples de chacune des catégories que l'on souhaite identifier. Si l'on veut améliorer les performances de la voiture autonome, on peut aussi lui donner une récompense en fonction de sa rapidité et une pénalité en cas d'accident. Quoi qu'il en soit, la mise en œuvre des procédures d'apprentissage nécessite des exemples annotés pour l'apprentissage supervisé et « récompensés » pour l'apprentissage par renforcement. Dans les deux cas, le même problème se pose : qui annoté ou récompense et punit ? Un professeur est requis, ce qui signifie que la machine n'est pas totalement autonome en ce sens qu'elle ne se donne pas spontanément ses propres règles, puisqu'elle suit la leçon que des hommes lui enseignent.

Pour éviter toute ambiguïté, rappelons que le terme d'autonomie appliqué à un agent recouvre deux idées bien différentes. D'un côté, cela signifie que l'agent se meut seul et prend des décisions d'action sans le secours d'un autre, comme cela se produit pour une voiture autonome qui choisit d'elle-même son itinéraire ou pour un système d'armes qui déclencherait un tir sur une cible, parce qu'elle bouge ou qu'elle possède toutes les caractéristiques indiquées au cours de la phase d'apprentissage. Autrement dit, l'autonomie d'un agent, au sens technique, signifie qu'il existe une chaîne de causalités matérielles allant de la prise d'information par des capteurs, à la décision, puis à l'action, qui ne fait pas intervenir d'agent extérieur, en particulier d'agent humain. Mais d'un autre côté, au plan philosophique, l'autonomie tient à la capacité à se donner sa propre loi, à savoir les règles et les finalités de son comportement. Dans cette seconde acception que nous qualifierons de philosophique, l'autonomie s'oppose à l'hétéronomie, autrement dit, à la soumission d'un agent à une loi dictée de

l'extérieur, par d'autres. Cela voudrait dire qu'un système d'armes qualifié d'autonome en ce second sens ne se contenterait pas de déclencher un tir sur une cible choisie parce qu'elle possède des caractéristiques de couleur ou de forme données, mais qu'il déterminerait, de lui-même, les caractéristiques des cibles qu'il déciderait d'atteindre pour satisfaire les objectifs qu'il se serait lui-même donnés. Nous savons depuis longtemps fabriquer des machines qui se meuvent d'elles-mêmes, comme les drones ou les trains automatiques, c'est-à-dire qui sont autonomes au sens technique. Cela ne pose pas de problème. Les techniques d'apprentissage les rendent plus efficaces ; elles facilitent indubitablement leur programmation et leur mise en œuvre. La question délicate porte non sur l'autonomie en ce premier sens, mais sur l'autonomie dans son acception philosophique. Même douées d'apprentissage et de capacités à faire évoluer leurs propres programmes, les machines n'en acquièrent pas pour autant d'autonomie en ce second sens, car elles restent soumises aux catégories et aux finalités imposées par ceux qui auront annoté les exemples utilisés durant la phase d'apprentissage.

[...]

Il s'ensuit que ni l'apprentissage supervisé ni l'apprentissage par renforcement ne dotent les machines d'autonomie au sens philosophique. Rien donc, au plan scientifique, ne justifie les craintes susmentionnées exprimées entre autres par Stephen Hawking, Elon Musk, Bill Gates et Stuart Russell dans leurs déclarations

publiques récentes. Sans doute, les résultats empiriques évoqués apparaissent-ils significatifs et les performances éblouissantes des techniques d'apprentissage méritent-elles qu'on s'appesantisse sur les risques potentiels d'accidents. Mais cela ne veut nullement dire que nous atteignons un point de non-retour au-delà duquel nous devons nous soumettre au pouvoir des machines.

## Invention

À cela on ajoutera que, pour être mis en œuvre, les algorithmes d'apprentissage requièrent des observations qui doivent être décrites dans un langage formel, par exemple sous forme d'un vecteur de caractéristiques ou d'une formule mathématique ou logique. Ce langage prend une part déterminante dans les capacités qu'ont les machines à apprendre : trop pauvre, il ne permet pas d'exprimer les distinctions nécessaires à la formulation des connaissances ; trop riche, il noie les procédures d'apprentissage dans l'immensité des théories possibles. C'est ce que démontrent les théories formelles de l'apprentissage.

Or, les machines ne modifient pas d'elles-mêmes le langage dans lequel s'expriment les observations qui alimentent leurs mécanismes d'apprentissage et les connaissances qu'elles construisent. Elles ne parviennent ni à étendre ce langage, ni à le restreindre lorsqu'il se révèle trop riche. Il y eut bien quelques tentatives, que ce soit avec la programmation logique inductive, dans les années 1990, ou plus récemment avec l'apprentissage profond, mais les maigres résultats ne sauraient convaincre. Il faut dire que ces transformations de langage s'avèrent extrêmement délicates, même pour les hommes. Cela s'apparente aux difficultés que l'on rencontre dans la découverte scientifique. La science qualifiée de « normale » par Thomas Kuhn, parce qu'elle énonce des lois dans un cadre conceptuel fixe, se trouve parfois limitée par les restrictions du langage imposé par ce cadre conceptuel. À titre d'illustration, lorsque Claude Bernard chercha à élucider le mécanisme d'action du curare, il hésita pendant vingt ans entre une

paralysie des muscles et une modification de la conductivité des nerfs, sans jamais parvenir à discriminer l'une des deux hypothèses. Cela tient à ce que la cible du curare n'est ni le nerf, ni le muscle, mais la plaque de jonction entre le nerf et le muscle. Or, cette entité intermédiaire n'existait pas pour lui ; il se trouvait donc dans l'incapacité de l'incriminer. Pour parvenir à formuler correctement l'hypothèse, il aurait fallu introduire un concept dont il ne disposait pas.

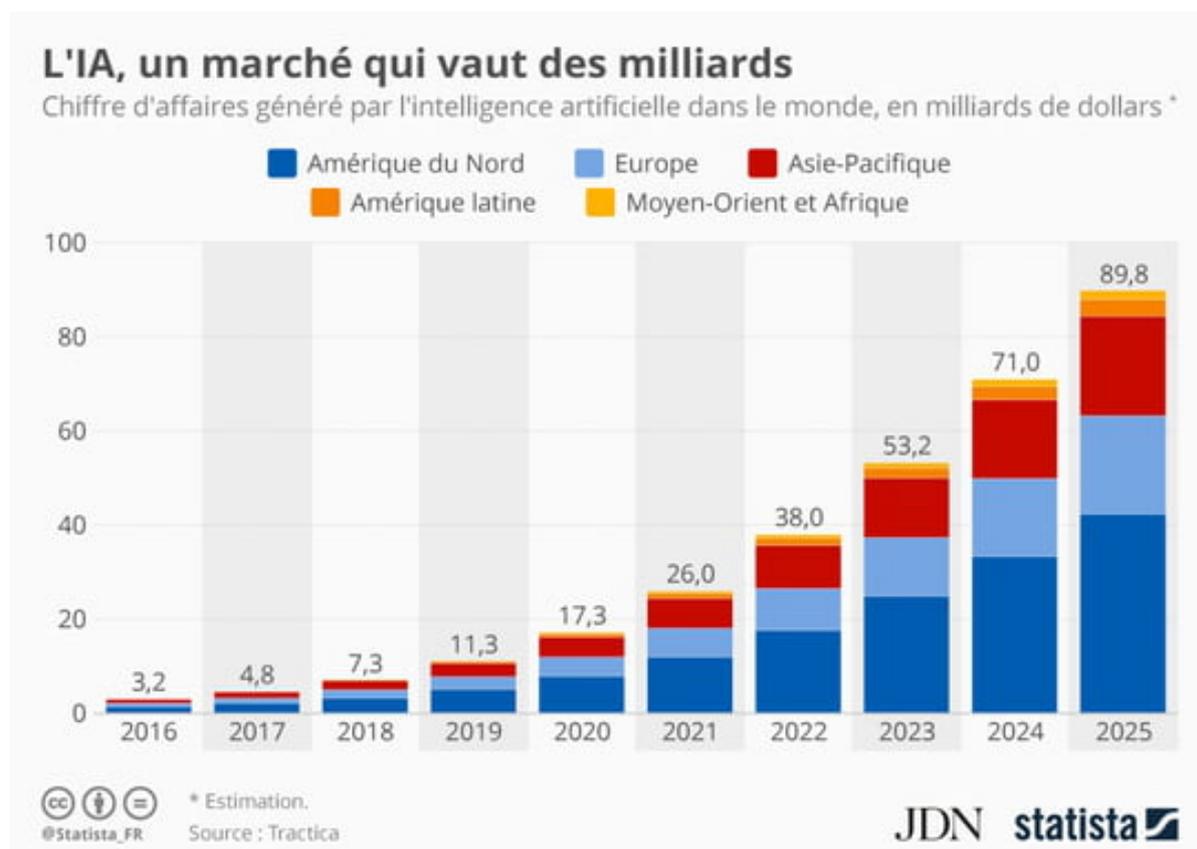
Les mutations de systèmes conceptuels, que Gaston Bachelard caractérise comme des ruptures épistémologiques ou que Thomas Kuhn appelle des changements de paradigmes, se produisent sur un rythme plus lent et plus hasardeux que la formation de lois dans un cadre conceptuel donné. Aujourd'hui, les techniques d'apprentissage machine construisent des lois empiriques qui résument très bien des observations, mais elles ne parviennent pas vraiment à inventer des catégories neuves, même si l'on fait parfois état, çà et là, de quelques cas de notions connues émergeant spontanément, sans qu'on les ait explicitement mentionnées. Les prouesses technologiques réalisées grâce à la mise en œuvre d'algorithmes d'apprentissage sur de grandes masses de données ne tiennent ni à la création de descripteurs, ni à l'enrichissement des langages d'expression des connaissances. Or, les techniques d'apprentissage non supervisées censées construire automatiquement de nouvelles classes n'ont pas encore acquis la maturité qui leur procurerait la capacité à inventer des concepts neufs et, encore moins, des appareils conceptuels inédits.

En conclusion, rien dans l'état actuel des techniques d'intelligence artificielle n'autorise à affirmer que les ordinateurs seront bientôt en mesure de se perfectionner indéfiniment sans le concours des hommes, jusqu'à s'emballer, nous dépasser et acquérir leur autonomie.

## DOCUMENT 8

journaldunet.com  
Jamal El Hassani  
13 juillet 2018

### L'insolente croissance du marché de l'intelligence artificielle



Le chiffre d'affaires généré par les technologies d'IA pourrait être multiplié par 18 en huit ans. Chaque jour, l'intelligence artificielle (IA) s'immisce davantage dans les logiciels et les organisations, tandis que sa précision ne cesse de progresser. Conséquence, le chiffre d'affaires du secteur devrait exploser, à mesure qu'il se rend indispensable au fonctionnement de la tech. Ainsi, alors qu'il représentait 4,8 milliards de dollars en 2017, il devrait être multiplié par dix d'ici 2022-2023, avant d'atteindre presque 90 milliards de dollars de revenus générés en 2025, selon Statista. Si les Etats-Unis risquent de devenir les grands gagnants de ce marché en captant près de la moitié du chiffre d'affaires généré par l'IA, l'Europe et l'Asie devraient se partager l'autre moitié, moins quelques miettes laissées à l'Amérique Latine, au Moyen-Orient et à l'Afrique

# LIGNES DIRECTRICES EN MATIERE D'ETHIQUE POUR UNE IA DIGNE DE CONFIANCE *(extrait)*

Groupe d'experts indépendants de haut niveau sur l'intelligence artificielle constitué par la commission européenne en juin 2018

## [...] I. Chapitre I: Fondements d'une IA digne de confiance

- (31) Ce chapitre établit les fondements d'une IA digne de confiance, reposant sur les droits fondamentaux et reflétée par quatre principes éthiques auxquels il convient d'adhérer afin de garantir une IA éthique et robuste. Ce chapitre s'appuie fortement sur le domaine de l'éthique.
- (32) L'éthique en matière d'IA est un sous-domaine de l'éthique appliquée qui est axé sur les questions d'ordre éthique soulevées par la mise au point, le déploiement et l'utilisation de l'IA. Sa préoccupation centrale consiste à déterminer la manière dont l'IA peut soulever des préoccupations relatives au bien-être des individus ou y apporter des solutions, que ce soit du point de vue de la qualité de vie ou de l'autonomie humaine et de la liberté nécessaire pour une société démocratique.
- (33) Une réflexion éthique sur la technologie de l'IA peut servir plusieurs objectifs. Premièrement, elle peut stimuler la réflexion sur la nécessité de protéger les individus et les groupes au niveau le plus élémentaire. Deuxièmement, elle peut stimuler de nouveaux genres d'innovation dont l'objectif est de promouvoir des valeurs éthiques, telles que celles contribuant à la réalisation des objectifs de développement durable des Nations unies<sup>13</sup>, qui sont fermement ancrés dans le futur programme de l'Union européenne à l'horizon 2030<sup>14</sup>. Si le présent document porte principalement sur le premier objectif mentionné, il ne faut pas sous-estimer l'importance que pourrait revêtir l'éthique dans le cadre du deuxième objectif. Une IA digne de confiance peut renforcer la prospérité des individus et le bien-être collectif en générant de la prospérité, en créant de la valeur et en maximisant les richesses. Elle peut contribuer à la réalisation d'une société juste, en contribuant à l'amélioration de la santé et du bien-être des citoyens d'une manière qui renforce l'égalité dans la répartition des possibilités économiques, sociales et politiques.
- (34) Il est par conséquent impératif que nous comprenions comment soutenir au mieux la mise au point, le déploiement et l'utilisation de l'IA pour faire en sorte que chacun puisse s'épanouir dans un monde fondé sur l'IA, et pour préparer un avenir meilleur tout en préservant la compétitivité au niveau mondial. Comme toute technologie puissante, l'utilisation de systèmes d'IA au sein de notre société soulève plusieurs problèmes éthiques, par exemple en ce qui concerne leur incidence sur les individus et la société, les capacités de prise de décision et la sécurité. Si nous prévoyons de nous faire assister par des systèmes d'IA ou de leur déléguer de plus en plus de décisions, nous devons veiller à ce que l'incidence de ces systèmes sur la vie des personnes soit équitable, à ce que ces systèmes soient conformes aux valeurs inaliénables et capables d'agir en ce sens, ainsi qu'à l'existence de processus adaptés en matière de responsabilisation pour y veiller.
- (35) L'Europe doit définir la vision normative qu'elle souhaite mettre en œuvre pour un avenir marqué par l'omniprésence de l'IA et, par conséquent, comprendre quelle notion de l'IA devrait être étudiée, mise au point, déployée et utilisée en Europe pour réaliser cette vision. Avec ce document, nous souhaitons contribuer à cet effort en introduisant la notion d'IA digne de confiance qui est, selon nous, la manière adaptée de bâtir un avenir avec l'IA. Un avenir dans lequel la démocratie, l'état de droit et les droits fondamentaux sous-tendent les systèmes d'IA et dans lequel ces systèmes améliorent et défendent de manière continue la culture démocratique permettra également de mettre en place un environnement dans lequel l'innovation et la compétitivité responsable peuvent se développer.
- (36) Un code de déontologie spécifique à un domaine donné – quel que soit le niveau de cohérence, d'élaboration et de détail de ses futures versions – ne pourra jamais se substituer à un raisonnement éthique en tant que tel, qui doit en toutes circonstances rester sensible aux éléments de contexte, qui ne peuvent jamais être rendus dans des lignes directrices générales. En plus d'élaborer un ensemble de règles, il convient de mettre sur pied et de conserver une culture et un état d'esprit éthiques dans le débat public, l'éducation et l'apprentissage pratique pour garantir une IA digne de confiance.

<sup>13</sup> [https://ec.europa.eu/commission/publications/reflection-paper-towards-sustainable-europe-2030\\_fr](https://ec.europa.eu/commission/publications/reflection-paper-towards-sustainable-europe-2030_fr).

<sup>14</sup> <https://sustainabledevelopment.un.org/?menu=1300>.

[...]

## 2.1 Les droits fondamentaux comme base d'une IA digne de confiance

- (40) Parmi l'éventail complet de droits indivisibles énoncés dans le droit international en matière de droits de l'homme, les traités de l'Union et la charte de l'Union, les familles de droits fondamentaux mentionnées ci-après sont particulièrement adaptées à une application aux systèmes d'IA. Une part importante de ces droits sont, dans des circonstances définies, opposables au sein de l'UE, ce qui rend juridiquement obligatoire la conformité avec leurs exigences. Toutefois, même lorsque la conformité avec les droits fondamentaux opposables a été atteinte, une réflexion éthique peut nous aider à comprendre de quelle manière la mise au point, le déploiement et l'utilisation de l'IA peuvent mettre en jeu les droits fondamentaux et leurs valeurs sous-jacentes, et peuvent contribuer à des orientations plus précises lorsqu'il s'agit de déterminer ce que nous *devrions* faire plutôt que ce que nous *pouvons* faire (actuellement) à l'aide de la technologie.
- (41) **Respect de la dignité humaine.** La dignité humaine comprend l'idée que chaque être humain possède une «valeur intrinsèque», qui ne devrait jamais être diminuée, compromise ou réprimée par autrui – ni par de nouvelles technologies telles que des systèmes d'IA.<sup>19</sup> Dans le contexte de l'IA, le respect de la dignité humaine signifie que chaque personne est traitée avec respect du fait de son statut de *sujet* moral, plutôt que comme simple *objet* que l'on trie, classe, marque, régente, conditionne ou manipule. [...]
- (42) **Liberté des individus.** Les êtres humains devraient rester libres de faire leurs propres choix de vie. Cela suppose l'absence d'intrusion du pouvoir, mais requiert également l'intervention des pouvoirs publics et des organisations non gouvernementales pour faire en sorte que les individus exposés au risque d'exclusion jouissent d'une égalité d'accès aux avantages et aux possibilités que présente l'IA. Dans un contexte d'IA, la liberté des individus requiert l'atténuation des contraintes illégitimes, des menaces à l'encontre de l'autonomie et de la santé mentales, de la surveillance injustifiée, de la tromperie et de la manipulation injuste, que ces atteintes soient directes ou indirectes. [...]
- (43) **Respect de la démocratie, de la justice et de l'état de droit.** Dans les démocraties constitutionnelles, tout pouvoir gouvernemental doit être légalement autorisé et limité par la loi. Les systèmes d'IA devraient servir à conserver et à encourager les processus démocratiques et le respect de la pluralité des valeurs et des choix de vie des individus. Les systèmes d'IA ne doivent pas compromettre les processus démocratiques, la délibération humaine ou les systèmes de vote démocratiques. [...]
- (44) **Égalité, non-discrimination et solidarité – y compris le droit des personnes exposées au risque d'exclusion.** Il convient d'assurer un respect égal de la valeur morale et de la dignité de tous les êtres humains. Cela va au-delà de la non-discrimination, qui tolère le fait d'établir des distinctions entre des situations différentes sur la base de justifications objectives. Dans un contexte d'IA, l'égalité implique que le fonctionnement du système ne peut pas produire de résultats fondés sur des biais injustes (par exemple, les données utilisées pour entraîner les systèmes d'IA devraient être aussi inclusives que possible et représenter différents groupes de population), ce qui requiert également un respect approprié des personnes et des groupes potentiellement vulnérables<sup>21</sup>, tels que les travailleurs, les femmes, les personnes handicapées, les minorités ethniques, les enfants, les consommateurs ou d'autres catégories de personnes exposées au risque d'exclusion.
- (45) **Droits des citoyens.** Les citoyens bénéficient d'un large éventail de droits, dont le droit de vote, le droit à une bonne administration ou à l'accès aux documents publics, et le droit d'adresser des pétitions à l'administration. Grâce aux systèmes d'IA, les pouvoirs publics seront en mesure de fournir à la société des biens et des services publics à une échelle et avec une efficacité supérieures. Dans le même temps, les applications d'IA pourraient aussi avoir une incidence négative sur les droits des citoyens; il convient par conséquent de protéger ces droits. [...]

## (41) 2.2 Principes éthiques dans le contexte des systèmes d'IA

- (42) De nombreuses organisations publiques, privées et civiles se sont inspirées des droits fondamentaux pour élaborer des cadres éthiques pour les systèmes d'IA.<sup>23</sup> Dans l'UE, le Groupe européen d'éthique des sciences et des nouvelles technologies («GEE») a proposé un ensemble de neuf principes fondamentaux, reposant sur les valeurs fondamentales énoncées dans les traités de l'Union et dans la charte des droits fondamentaux de l'Union européenne.<sup>24</sup> Nous continuons à nous appuyer sur ces travaux, en reconnaissant la plupart des principes avancés jusqu'à présent par différents groupes, tout en précisant à quelles fins l'ensemble de ces principes cherchent à répondre et à apporter un soutien. Ces principes éthiques peuvent inspirer de nouveaux instruments réglementaires spécifiques, contribuer à l'interprétation des droits fondamentaux au fur et à mesure qu'évolue notre environnement sociotechnique et orienter les motifs justifiant la mise au point, l'utilisation et la mise en œuvre de systèmes d'IA – en s'adaptant de manière dynamique aux évolutions de la société elle-même.
- (46) Les systèmes d'IA doivent améliorer le bien-être individuel et collectif. Cette section présente **quatre principes éthiques**, ancrés dans les droits fondamentaux, auxquels il convient d'adhérer pour faire en sorte que les systèmes d'IA soient mis au point, déployés et utilisés d'une manière digne de confiance. Ils sont présentés comme des **impératifs éthiques**, si bien que les professionnels de l'IA devraient en toutes circonstances s'efforcer d'y adhérer. Sans imposer de hiérarchie, nous présentons les principes ci-dessous de manière à refléter l'ordre d'apparition, dans la charte de l'Union, des droits fondamentaux sur lesquels ils se fondent.<sup>25</sup>
- (47) Il s'agit des principes suivants:
- (i) respect de l'autonomie humaine
  - (ii) prévention de toute atteinte
  - (iii) équité
  - (iv) explicabilité
- (48) La plupart de ces principes sont dans une large mesure déjà reflétés dans les exigences juridiques contraignantes dont la mise en œuvre est obligatoire et relèvent donc également du champ d'application de l'«IA licite», soit la première caractéristique d'une IA digne de confiance.<sup>26</sup> Pourtant, comme indiqué plus haut, même si de nombreuses obligations juridiques reflètent des principes éthiques, l'adhésion à des principes éthiques dépasse le respect formel de la législation existante.<sup>27</sup>
- Le principe du respect de l'autonomie humaine
- (50) Les droits fondamentaux sur lesquels l'UE est fondée ont vocation à garantir le respect de la liberté et de l'autonomie des êtres humains. Les êtres humains qui interagissent avec des systèmes d'IA doivent être en mesure de conserver leur autodétermination totale et effective et de prendre part au processus démocratique. En l'absence de justification, les systèmes d'IA ne devraient pas subordonner, contraindre, tromper, manipuler, conditionner ni régenter des êtres humains. Au contraire, les systèmes d'IA devraient être conçus afin d'augmenter, de compléter et de favoriser les compétences cognitives, sociales et culturelles. La répartition des tâches entre êtres humains et systèmes d'IA devrait suivre des principes de conception centrés sur l'humain et donner à l'être humain une possibilité réelle de poser des choix. En d'autres termes, il convient de veiller à la supervision<sup>28</sup> et au contrôle humains sur les processus de travail des systèmes d'IA. Les systèmes d'IA pourraient également modifier fondamentalement la sphère du travail. Ces systèmes devraient aider les êtres humains dans l'environnement de travail, et avoir pour objectif de créer des emplois qui aient du sens.
- Le principe de la prévention de toute atteinte
- (51) Les systèmes d'IA ne devraient ni porter atteinte, ni aggraver toute atteinte portée<sup>29</sup>, ni nuire aux êtres humains d'une quelconque autre manière.<sup>30</sup> Cela englobe la protection de la dignité humaine ainsi que de l'intégrité mentale et physique. Les systèmes d'IA et les environnements dans lesquels ils évoluent doivent être sûrs et sécurisés. Ils doivent être robustes sur le plan technique et il convient de veiller à ce qu'ils ne soient pas exposés à des utilisations malveillantes. Les personnes vulnérables devraient faire l'objet d'une attention accrue et être prises en compte dans la mise au point et le déploiement des systèmes d'IA. Il convient également d'accorder une attention particulière aux situations dans lesquelles les systèmes d'IA peuvent entraîner ou aggraver des incidences négatives du fait d'asymétries de pouvoir ou d'information, par exemple entre les employeurs et les travailleurs, entre les entreprises et les consommateurs ou entre les pouvoirs publics et les citoyens. La prévention de toute atteinte implique également la prise en compte de l'environnement naturel et de tous les êtres vivants.

- Le principe de l'équité

- (52) La mise au point, le déploiement et l'utilisation de systèmes d'IA doivent être équitables. Si nous reconnaissons que l'équité peut s'interpréter de multiples manières, nous considérons que l'équité se caractérise à la fois par un volet matériel et un volet procédural. Le volet matériel suppose l'engagement de veiller à une répartition égale et juste des bénéfices et des coûts, et de veiller à ce que les individus et les groupes ne fassent pas l'objet de biais injustes, de discrimination et de stigmatisation. Si les biais injustes peuvent être évités, les systèmes d'IA pourraient même améliorer le caractère équitable de la société. Il convient également d'encourager l'égalité des chances en ce qui concerne l'accès à l'éducation, aux biens, aux services et à la technologie. En outre, l'utilisation de systèmes d'IA ne devrait jamais avoir pour conséquence de tromper les utilisateurs (finaux) ou de limiter leur liberté de choix. L'équité implique en outre que les professionnels de l'IA devraient respecter le principe de proportionnalité entre la fin et les moyens, et examiner de manière attentive la manière de trouver un équilibre entre des intérêts et des objectifs en

concurrence.<sup>31</sup> Le volet procédural de l'équité suppose la capacité de contester les décisions prises par des systèmes d'IA et par les êtres humains qui les utilisent, ainsi que celle d'introduire un recours efficace à l'encontre de ces décisions<sup>32</sup>. Pour ce faire, l'entité responsable de la décision doit pouvoir être identifiée, et le processus de prise de décisions devrait pouvoir être expliqué.

- Le principe de l'explicabilité

- (53) L'explicabilité est essentielle pour renforcer et conserver la confiance des utilisateurs envers les systèmes d'IA. Cela signifie que les processus doivent être transparents, que les capacités et la finalité des systèmes d'IA doivent être communiquées ouvertement, et que les décisions – dans la mesure du possible – doivent pouvoir être expliquées aux personnes directement et indirectement concernées. Sans ces informations, une décision ne peut être dûment contestée. Il n'est pas toujours possible d'expliquer pour quelle raison un modèle a généré un résultat ou une décision en particulier (et quelle combinaison de facteurs d'entrée y a contribué). On parle d'algorithmes à effet «boîte noire». Ceux-ci doivent faire l'objet d'une attention particulière. Dans de telles circonstances, d'autres mesures d'explicabilité (par exemple la traçabilité, l'auditabilité et la communication transparente concernant les capacités du système) pourraient être requises, pour autant que le système dans son ensemble respecte les droits fondamentaux. La mesure dans laquelle l'explicabilité est nécessaire dépend fortement du contexte et de la gravité des conséquences si ce résultat est erroné ou imprécis d'une autre manière.<sup>33</sup>

### **2.3 Tensions entre ces principes**

- (54) Des tensions pourraient survenir entre les principes susmentionnés, pour lesquelles il n'existe pas de solution unique. En vertu de l'engagement fondamental de l'UE envers l'engagement démocratique, le droit à une procédure régulière et la participation politique ouverte, des méthodes de délibération responsable devraient être établies pour faire face à ces tensions. Par exemple, dans divers domaines d'application, *le principe de la prévention de toute atteinte* et *le principe de l'autonomie humaine* peuvent entrer en conflit. Ainsi, l'utilisation de systèmes d'IA aux fins d'une «police prédictive» pourrait contribuer à réduire la criminalité, mais d'une manière impliquant des activités de surveillance qui portent atteinte à la liberté individuelle et à la vie privée. En outre, la somme des avantages liés aux systèmes d'IA doit être sensiblement supérieure aux risques individuels prévisibles. [...] Il pourrait toutefois exister des situations dans lesquelles aucun arbitrage acceptable du point de vue éthique ne peut être déterminé. Certains droits fondamentaux et principes connexes sont absolus et ne peuvent dépendre d'un exercice de mise en balance (par exemple, la dignité humaine).

[...]

## II. Chapitre II: Parvenir à une IA digne de confiance

(55) Ce chapitre fournit des orientations relatives à la mise en œuvre et à la réalisation d'une IA digne de confiance, au moyen d'une liste de sept exigences qui devraient être respectées, s'appuyant sur les principes énoncés au chapitre I. En outre, des méthodes tant techniques que non techniques actuellement disponibles sont présentées aux fins de l'application de ces exigences tout au long du cycle de vie du système d'IA.

### 1. Exigences d'une IA digne de confiance

(56) Pour parvenir à une IA digne de confiance, il faut que les principes énoncés au chapitre I soient traduits en exigences concrètes. Ces exigences s'appliquent aux différentes parties prenantes participant au cycle de vie des systèmes d'IA: développeurs, prestataires et utilisateurs finaux, ainsi que la société au sens large. Le terme «développeurs» désigne les personnes qui effectuent des recherches sur les systèmes d'IA, et qui conçoivent et/ou mettent au point ces systèmes. Le terme «prestataires» désigne les organismes publics ou privés qui utilisent des systèmes d'IA dans leurs processus opérationnels pour proposer des produits et services à des tiers. Les utilisateurs finaux sont les personnes qui interagissent directement ou indirectement avec le système d'IA. Enfin, la société au sens large englobe tous les autres acteurs qui sont directement ou indirectement concernés par les systèmes d'IA.

(57) Différentes catégories de parties prenantes ont différents rôles à jouer pour veiller au respect des exigences:

- a. Les développeurs devraient mettre en œuvre et appliquer les exigences aux processus de conception et de mise au point;
- b. Les prestataires devraient veiller à ce que les systèmes qu'ils utilisent et les produits et services qu'ils proposent respectent les exigences;
- c. Les utilisateurs finaux et la société au sens large devraient être informés de ces exigences et être en mesure de demander qu'elles soient respectées.

(58) La liste des exigences ci-dessous n'est pas exhaustive.<sup>35</sup> Elle comprend des aspects systémiques, individuels et sociétaux:

#### 1 Action humaine et contrôle humain

*Comprend les droits fondamentaux, l'action humaine et le contrôle humain*

#### 2 Robustesse technique et sécurité

*Comprend la résilience aux attaques et la sécurité, les plans de secours et la sécurité générale, la précision, la fiabilité et la reproductibilité*

#### 3 Respect de la vie privée et gouvernance des données

*Comprend le respect de la vie privée, la qualité et l'intégrité des données, et l'accès aux données*

#### 4 Transparence

*Comprend la traçabilité, l'explicabilité et la communication*

#### 5 Diversité, non-discrimination et équité

*Comprend l'absence de biais injustes, l'accessibilité et la conception universelle, et la participation des parties prenantes*

#### 6 Bien-être sociétal et environnemental

*Comprend la durabilité et le respect de l'environnement, l'impact social, la société et la démocratie*

#### 7 Responsabilité

*Comprend l'auditabilité, la réduction au minimum des incidences négatives et la communication à leur sujet, les arbitrages et les recours.*

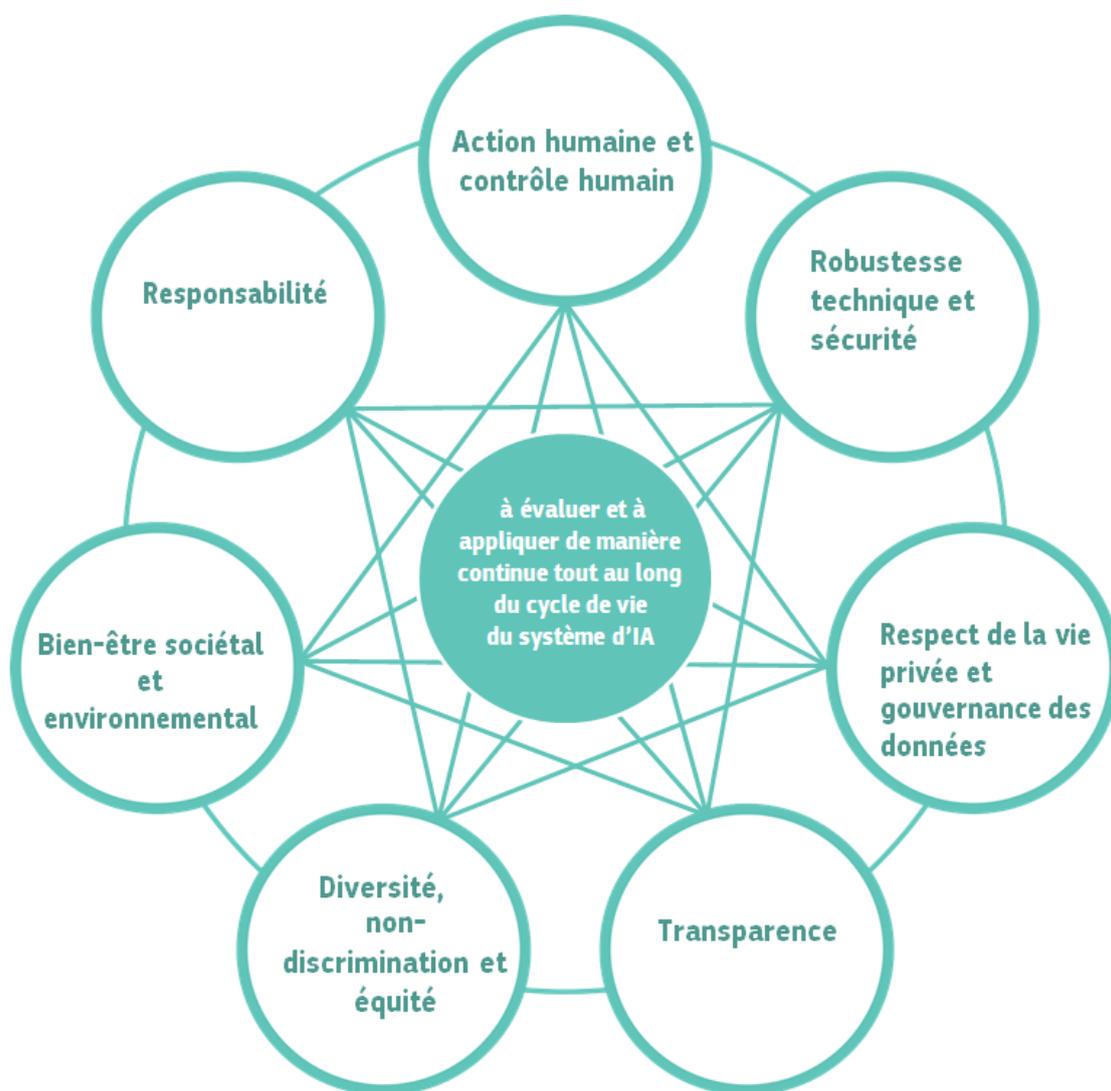


Figure 2: interrelation des sept exigences: elles revêtent toutes une importance égale, elles se soutiennent mutuellement et devraient être appliquées et évaluées tout au long du cycle de vie d'un système d'IA.

lemonde.fr  
Nicolas Lepeltier  
30 juillet 2019

## **« Il faut tout faire pour que l'automatisation des métiers favorise l'émancipation de l'homme »**

*Selon le dernier rapport annuel sur les perspectives de l'emploi de l'Organisation de coopération et de développement économiques (OCDE), publié le 25 avril, 14 % des emplois dans les pays de l'OCDE sont susceptibles de disparaître (16,4 % en France) et 32 % pourraient être profondément transformés (32,8 % dans l'Hexagone) avec l'automatisation des tâches et la multiplication des machines dans le monde du travail au cours des vingt prochaines années.*

*Stefano Scarpetta, directeur de l'emploi, du travail et des affaires sociales de l'OCDE, estime qu'il est urgent d'adapter le système scolaire aux nouvelles réalités, et de donner aux travailleurs des occasions de développer de nouvelles compétences par la formation tout au long de leur vie professionnelle.*

### **Des études prédisent la disparition de 400 millions à 800 millions d'emplois avec l'automatisation. D'autres prétendent au contraire que la robotisation va créer plus de postes qu'elle ne va en détruire. Pourquoi une telle différence d'appréciation ?**

Certains jobs vont disparaître, d'autres vont changer de façon importante. En même temps, la technologie numérique va créer de nouvelles opportunités d'emploi, dans les hautes technologies notamment, en complément de ce que les machines pourront accomplir à l'avenir. C'est pourquoi nous pensons, à l'OCDE, qu'il y aura plus de changements de nature des tâches que de pertes sèches d'emplois. Nous ne sommes pas inquiets par la perspective de ce qu'on appelle « *un chômage technologique de masse* ».

Nous craignons davantage une forte augmentation des inégalités sur le marché du travail entre les personnes qui ont les compétences pour saisir ces nouvelles opportunités, et d'autres, faiblement qualifiées, qui ont des compétences limitées pour répondre aux évolutions de l'emploi qu'elles occupent et qui resteront cantonnées dans des fonctions peu intéressantes et mal payées.

### **Quels secteurs d'activité sont les plus concernés par la robotisation ? Et pour quels types d'emplois ?**

Presque tous les secteurs sont concernés : principalement la manufacture, car c'est un secteur où les tâches sont répétitives ; l'agriculture également est en train de changer de façon assez spectaculaire, car, même s'il y a peu d'emplois, de plus en plus de tâches sont réalisées avec le soutien de la technologie numérique.

Dans les services, si vous prenez l'exemple des banques, quand vous alliez au guichet il y a quinze ou vingt ans, il y avait encore du personnel pour aider les clients. Cela ne veut pas dire que l'emploi dans les banques a diminué, mais simplement que les salariés font des choses différentes. Dans certains cabinets d'avocats, aujourd'hui, la recherche des décisions de justice se fait de manière automatisée. C'est plus facile, beaucoup plus rapide.

### **Quels avantages peut-on retirer de l'automatisation ?**

Comme lors des précédentes révolutions technologiques, les machines vont accomplir des tâches répétitives, ennuyeuses et parfois risquées. Cela laisse aux humains l'occasion de se concentrer sur des tâches plus novatrices, plus créatives, avec plus de travail d'équipe – ce que l'on appelle « *l'intelligence sociale* ».

Ça, c'est la théorie. En pratique, cette révolution numérique met les travailleurs et l'intelligence artificielle en compétition sur des tâches plus cognitives, parce que les algorithmes développent des capacités impressionnantes de traitement des données.

## **Avez-vous des exemples précis où la machine entre en compétition avec l'humain ?**

Beaucoup d'entreprises utilisent déjà les algorithmes pour sélectionner du personnel, car ils ont la capacité d'analyser le CV des candidats de façon très détaillée, d'apporter des informations supplémentaires par rapport aux méthodes classiques de recrutement.

Dans l'agriculture, il existe des sondes qui permettent d'apporter aux cultures la bonne dose de produit, un rôle qui auparavant était rempli par les agriculteurs...

## **Sommes-nous tous égaux devant l'automatisation du travail ?**

Il y a des tâches pour lesquelles les humains vont conserver un rôle prédominant : la créativité, les interactions complexes, l'intelligence sociale sont plutôt des caractéristiques humaines. Mais d'autres tâches, plus répétitives, pourront être accomplies par des algorithmes. Nous sommes tous exposés aux nouvelles technologies numériques, mais l'impact sur nos professions est assez différent. Les postes les plus à risque d'être automatisés dans les pays de l'OCDE sont ceux à faible niveau de qualification, occupés par des personnes qui reçoivent le moins de formation professionnelle. Or, ce sont elles qui en ont le plus besoin.

Il faut adapter le système scolaire aux nouvelles réalités, mais il faut aussi donner aux travailleurs des occasions de développer de nouvelles compétences par la formation tout au long de leur vie professionnelle.

Dans les pays de l'OCDE, près de la moitié des travailleurs ont des compétences numériques très limitées, voire nulles : ils savent utiliser un smartphone, naviguer sur Internet, mais pas vraiment se servir de la technologie numérique dans leur activité professionnelle. Tout le monde ne doit pas savoir coder, mais avoir un minimum de compétences numériques est important pour les métiers de demain.

## **Comment aborder au mieux le virage de l'automatisation ?**

Il faut tout faire pour que l'automatisation des métiers favorise l'émancipation de l'homme : laisser aux machines les tâches ingrates et se concentrer sur les domaines les plus intéressants, des tâches plus créatives.

L'autre défi posé par l'automatisation du travail est celui des moyens : la question du financement de la transition numérique pour les travailleurs est très importante pour réduire les tensions potentielles et les inquiétudes liées à cette révolution : savoir comment on finance la formation continue, comment on garantit la protection sociale à tous les travailleurs indépendants des plates-formes notamment... Il faut donner à tous les moyens de naviguer dans un marché du travail en évolution rapide.

Il y a également un débat sur l'éthique de l'intelligence artificielle : comment faire en sorte que les algorithmes n'aient pas de biais qui puissent avoir des conséquences négatives pour certaines catégories de travailleurs ? Les experts disent que ce ne sont pas les algorithmes qui ont des biais dans leur codage informatique, mais plutôt les données qu'ils utilisent qui peuvent être biaisées, l'algorithme ne faisant que répéter certains de ces biais.

## **L'automatisation dans le travail est-elle inéluctable ?**

Il faut arrêter de croire que tout ce qui peut être accompli par des machines sera accompli par des machines. Il y a toujours des choix qui sont faits par les gouvernements, par les entreprises, par les travailleurs eux-mêmes.

La question, c'est plutôt quelle utilisation fait-on des nouvelles technologies ? L'exemple des voitures autonomes est un très bon exemple : la technologie est là, mais l'utilisation de ces véhicules dépend des cadres réglementaires que les pays vont adopter pour minimiser les risques. On ne peut pas empêcher le progrès technologique, mais ce que l'on peut réglementer, c'est l'utilisation de ces différentes technologies afin d'en optimiser les bénéfices pour nos sociétés.

de ligne en ligne n°27

balises.bpi.fr

Sébastien Gaudelus et Marie-Hélène Gatto, Bpi

28/09/2018

## « L'intelligence artificielle transforme le monde »

### Entretien avec Jean-Gabriel Ganascia

À la fin des années 1980, à la stupeur générale, Garry Kasparov, célèbre joueur d'échecs, était battu par un ordinateur. Depuis, nous nous sommes habitués aux performances exponentielles des machines, à recourir quotidiennement à leurs services, et même à dialoguer avec elles. Mais les processus à l'œuvre et les répercussions sur nos vies demeurent, pour la plupart d'entre nous, mystérieux et alimentent bien des fantasmes.

***Professeur d'informatique à Sorbonne Université et conseiller scientifique du cycle Intelligences artificielles, Jean-Gabriel Ganascia retrace le développement de l'intelligence artificielle et pose sur ces problématiques un regard distancié et confiant.***

### Que désigne l'intelligence artificielle ?

L'intelligence artificielle est une discipline scientifique qui naît en 1955. Le mot est inventé par deux jeunes mathématiciens, John McCarthy et Marvin Lee Minsky qui ont eu l'idée d'étudier l'intelligence en la décomposant en fonctions cognitives si élémentaires que l'on puisse les simuler sur ces objets nouveaux pour l'époque qu'étaient les ordinateurs. En soixante ans, cette entreprise a été poursuivie par de nombreux chercheurs qui en ont montré la fécondité : la perception, à savoir le passage de l'information enregistrée par des capteurs à une représentation, le fonctionnement de la mémoire, l'apprentissage, le raisonnement vu comme un calcul sur les représentations, la communication par la parole ou les gestes, etc. Toutes ces fonctions ont fait l'objet de beaucoup de simulations informatiques. L'intelligence artificielle concerne aussi les robots : la mise en œuvre des décisions sur une machine est aussi modélisée.

Tout cela dessine un paysage assez vaste et nous permet de mieux nous comprendre en tant qu'humain. Une fois modélisées sur un ordinateur, les fonctions cognitives font l'objet d'applications pratiques. Par l'ampleur de son champ, l'intelligence artificielle transforme le monde. Par exemple, le web est lié à l'intelligence artificielle. C'est un modèle de mémoire, l'hypertexte, couplé au réseau de communication Internet.

Autre exemple, la reconnaissance des visages. Cela peut être tout à fait ludique, ou pratique pour les gens peu physionomistes. Mais cela peut être un cauchemar si vous imaginez des policiers capables de vous retrouver automatiquement pour vous verbaliser parce que vous avez laissé tomber un papier par terre ou que vous avez traversé au feu rouge.

### En quoi l'intelligence artificielle se distingue-t-elle ou se rapproche-t-elle de l'intelligence humaine ?

C'est une très bonne question, parce que l'objectif initial était de mieux comprendre l'intelligence humaine, de la modéliser, de faire quelque chose qui s'en rapproche le plus possible. En même temps, reproduire exactement le fonctionnement du cerveau est impossible.

C'est donc une simplification qui permet de rendre compte de processus cognitifs. À chaque fois qu'un modèle est créé, il est confronté à la réalité. Quand on s'aperçoit qu'il n'est pas tout à fait conforme et qu'il y a un écart avec les facultés intellectuelles humaines que l'on cherche à approcher, de nouveaux modèles sont introduits pour réduire cet écart. Comme pour la physique ou l'astrophysique, chaque difficulté aide à progresser dans la compréhension.

## **De quelle manière l'intelligence artificielle est-elle déjà présente dans notre quotidien ?**

Dans nos vies, la première chose, c'est le web. Ensuite, ce sont nos téléphones portables, qui contiennent beaucoup d'intelligence artificielle : les techniques de reconnaissance vocale, digitale ou faciale. L'intelligence artificielle, c'est aussi le traitement de grandes masses de données, dans le domaine médical ou dans le domaine du marketing pour faire de la publicité ciblée.

Ce sont quelques exemples, mais il faut savoir que toutes les sciences, aujourd'hui, sont transformées par l'intelligence artificielle. On procède à des expériences *in silico*, c'est-à-dire directement sur des données au moyen de calculs complexes informatisés ou de modèles informatiques.

Les démarches expérimentales classiques passaient par la mise en place de dispositifs matériels destinés à valider telle ou telle hypothèse théorique en confrontant les données recueillies sur ces dispositifs avec celles qu'anticipait l'hypothèse de départ. Tandis qu'aujourd'hui, les biologistes, avec le séquençage du génome, les physiciens ou les astrophysiciens ou encore les psychologues, grâce à la prolifération de capteurs à bas coût, engrangent d'immenses quantités de données, sur lesquelles ils peuvent tester des hypothèses.

Il est même possible de générer automatiquement tout un tas d'hypothèses, puis de les faire évaluer de façon systématique par une machine !

## **Pouvons-nous considérer qu'une intelligence artificielle, en particulier lorsqu'elle s'incarne dans un corps, peut avoir des émotions ?**

Le principe des interfaces humain-machine repose sur la projection d'un agent imaginaire sur une entité artificielle. Cela a été théorisé par le philosophe Daniel Dennett, avec les systèmes intentionnels. On suppose implicitement qu'il existe un agent mû par des intentions, que cet agent possède des désirs et qu'il dispose d'un certain nombre de moyens d'action qu'il mobilise à sa guise en vue de réaliser ses propres désirs.

Est-ce que nous pouvons considérer qu'il a des émotions ? Non, pas au sens propre, parce qu'une machine n'est jamais qu'un assemblage de dispositifs élémentaires, tous maîtrisables. Il n'y a pas d'organicité ; la machine n'a pas d'instinct de survie ; elle ne possède pas ce que Spinoza appelait le « conatus », à savoir la capacité à accroître d'elle-même sa puissance d'être. En tout cas, si elle avait des émotions, celles-ci seraient certainement très différentes des nôtres, parce qu'elles répondraient aux besoins de la machine qui n'ont rien à voir avec nos contraintes vitales, d'ordre biologique.

## **Quels sont les domaines dans lesquels l'intelligence artificielle se développe le plus actuellement, et dans lesquels elle peut avoir un impact très important dans notre vie quotidienne ?**

Je crois que dans le domaine de la santé, l'intelligence artificielle a vraiment beaucoup d'importance aujourd'hui. En tout cas, beaucoup espèrent l'utiliser. Est-ce que cela aura un grand impact sur nos vies ? Je ne le sais pas, c'est une question ouverte. Dans le domaine des transports, il y a les essais de voitures autonomes, mais pas seulement. L'intelligence artificielle permet aussi de réguler la circulation des trains, des métros, des voitures, etc.

Dans le domaine des technologies financières, elle joue déjà un rôle important, pour anticiper les risques. Peut-être y aura-t-il de nouvelles formes d'assurance fondées sur l'exploitation systématique de toutes les données disponibles — par exemple de toutes les traces que nous laissons sur les réseaux sociaux et qui indiquent si nous fumons, si nous sortons le soir, si nous faisons du sport, etc. — pour déterminer le montant, plus ou moins élevé, de la prime que nous aurons à verser ?

Dans le domaine de la défense, certains s'inquiètent de « robots tueurs », à savoir de machines qui décideraient elles-mêmes qui sont les ennemis, et qui engageraient le tir sans consulter qui que ce soit. Cela relève toutefois d'un imaginaire nourri de mythes anciens. En revanche, l'intelligence artificielle a beaucoup d'autres applications dans ce secteur ; ainsi il existe d'ores et déjà des robots qui vont traquer les mines tous seuls, et évitent à beaucoup de soldats de se faire exploser. Le rapport Villani sur l'intelligence artificielle cite aussi les domaines de l'environnement, de l'agriculture, etc.

[...]

## IA : « l'éthique-washing », une invention européenne

*Par Thomas Metzinger, Professeur de philosophie théorique à l'Université de Mayence et membre du groupe d'experts de haut niveau de la Commission Européenne qui a travaillé sur les lignes directrices du rapport pour une « intelligence artificielle éthique ».*

**C'est une très bonne nouvelle : l'Europe vient de prendre l'initiative du débat mondial très controversé sur l'éthique de l'intelligence artificielle (IA). Lundi dernier à Bruxelles, la Commission européenne a présenté ses lignes directrices en matière d'éthique pour une IA digne de confiance. Le groupe d'experts de haut niveau sur l'intelligence artificielle (HLEG AI), composé de 52 membres, dont moi-même, a travaillé sur le texte pendant neuf mois. Le résultat est un compromis dont je ne suis pas fier mais qui est néanmoins le meilleur au monde en la matière. Les États-Unis et la Chine n'ont rien de comparable. Que peut-on en tirer ?**

### L'intelligence artificielle ne peut être digne de confiance

L'histoire d'une IA digne de confiance est un récit marketing inventé par l'industrie, une histoire pour bercer les clients de demain. L'idée directrice sous-jacente d'une « IA digne de confiance » est avant tout un non-sens conceptuel. Les machines ne sont pas fiables ; seuls les humains peuvent être dignes de confiance (ou indignes de confiance). Si, à l'avenir, une société ou un gouvernement indigne de confiance se comporte de manière non éthique et possède une technologie d'IA efficace et robuste, cela permettra un comportement contraire à l'éthique plus performant. Par conséquent, le discours d'une IA digne de confiance concerne, en réalité, le développement de marchés futurs et l'utilisation des débats sur l'éthique comme des décorations publiques élégantes servant une stratégie d'investissement à grande échelle. Du moins, c'est l'impression que je commence à avoir après neuf mois de travail sur les lignes directrices.

### Peu d'éthiciens impliqués

La composition du groupe HLEG AI fait partie du problème : il ne comprenait que 4 éthiciens aux côtés de 48 non-éthiciens - des représentants du monde politique, d'universités, de la société civile et surtout de l'industrie. C'est comme si vous essayiez de créer une unité centrale à la pointe de la technologie en matière d'IA à destination des pouvoirs publics avec 48 philosophes, un pirate informatique et trois informaticiens (dont deux sont toujours en vacances).

Quiconque finalement était responsable du poids extrême des industriels au sein du groupe avait raison sur au moins un point : il est vrai que si vous voulez que l'industrie européenne de l'IA respecte les règles éthiques, vous devez impliquer les leaders du secteur et les impliquer dès le début. Il y a de bonnes et intelligentes personnes, et ça vaut la peine de les écouter. Cependant, même si le groupe d'experts comprenait beaucoup de personnes intelligentes, le gouvernail ne peut pas être laissé à l'industrie.

### La suppression des lignes rouges

En tant que membre du groupe d'experts, je suis déçu du résultat présenté. Les directives sont tièdes, court-termistes et délibérément vagues. Elles ignorent les risques à long terme, font abstraction de problèmes majeurs (« l'explicabilité ») grâce à une certaine forme de rhétorique, violent les principes élémentaires de la rationalité et prétendent savoir des choses que personne ne sait vraiment.

Aux côtés de l'excellent expert berlinois en *machine learning*, Urs Bergmann (Zalando), ma tâche consistait à développer, au cours de plusieurs mois de discussions, les « lignes rouges », principes éthiques non négociables qui déterminaient ce qu'il ne fallait pas faire avec l'IA en Europe. L'utilisation de systèmes d'armes autonomes létales figurait clairement sur notre liste, de même que l'évaluation des citoyens par l'État (score social) appuyée par l'IA et, en principe, l'utilisation de l'IA que les gens ne peuvent plus comprendre ni contrôler.

J'ai seulement réalisé que tout cela n'était pas vraiment souhaité lorsque notre sympathique président finlandais du HLEG, Pekka Ala-Pietilä (un ancien de chez Nokia), m'a demandé d'une voix douce si nous pouvions supprimer la phrase « non négociable » du document. À l'étape suivante, de nombreux représentants de l'industrie et membres du groupe intéressés par une

« vision positive » ont insisté avec véhémence pour que la mention « lignes rouges » soit entièrement supprimée du texte - bien que ce soient précisément ces lignes rouges qui relèvent de notre mandat. Le document publié ne parle plus de « lignes rouges » ; trois ont été complètement supprimés et le reste dilué. Au lieu de cela, on ne parle que de « préoccupations critiques ».

### **Des « Fake News » aux « Fake ethics »**

Ce phénomène est un exemple d'« éthique-washing ». L'industrie organise et cultive des débats éthiques pour gagner du temps - distraire le public et empêcher - ou du moins retarder - l'efficacité de la réglementation et de l'élaboration des politiques. Les politiciens aiment également mettre en place des comités d'éthique, car cela leur donne un plan d'action quand, étant donné la complexité des problèmes, ils ne savent tout simplement pas quoi faire - et cela est humain. Dans le même temps toutefois, l'industrie construit une « machine à laver éthique » après l'autre. Facebook a investi dans la TU Munich - en finançant un institut destiné à former des éthiciens de l'IA. De même, jusqu'à récemment, Google avait engagé les philosophes Joanna Bryson et Luciano Floridi dans un « comité d'éthique », qui a toutefois été abruptement arrêté une semaine après son lancement. Si cela n'avait pas eu lieu, Google aurait eu un accès direct via Luciano Floridi, membre du HLEG AI, au processus par lequel ce groupe élaborera les recommandations en matière de politique et d'investissement pour l'Union européenne à compter de ce mois. Cela aurait été un triomphe stratégique pour le conglomérat américain. Comme l'industrie agit plus rapidement et plus efficacement que la politique ou le secteur universitaire, il existe un risque que, comme dans le cas de « Fake News », nous ayons maintenant un problème avec les *fake ethics*, incluant nombre d'écrans de fumée conceptuels, des philosophes industriels hautement rémunérés, des labels de qualité auto-inventés et des certificats non validés pour « une éthique de l'IA made in Europe ».

Compte tenu de cette situation, qui pourrait maintenant développer des « lignes rouges » éthiquement convaincantes pour l'IA ? En réalité, il semble que cela ne puisse être fait que par la nouvelle Commission européenne, qui entame ses travaux après l'été. L'Amérique de Donald Trump est moralement discréditée ; il s'est lui-même auto-exclu. Et la Chine ? Tout comme en Amérique, il y a beaucoup de personnes intelligentes et bien intentionnées, et dans un souci de sécurité de l'IA, il pourrait, en tant qu'État totalitaire, faire respecter toute directive de manière contraignante. Mais la mise en place d'une surveillance de masse basée sur l'IA sur ses 1,4 milliard de citoyens est déjà bien avancée ; nous ne pouvons pas nous attendre à une véritable éthique. En tant que « totalitarisme numérique 2.0 », la Chine n'est pas une source acceptable pour des discussions éthiques sérieuses. L'Europe doit maintenant supporter le fardeau d'une responsabilité historique réelle.

### **Reprendre l'éthique des mains de l'industrie !**

Si vous avez des objectifs éthiques, vous êtes obligés de trouver et d'utiliser les meilleurs outils disponibles. L'intelligence artificielle est l'un des meilleurs instruments d'éthique pratique de l'humanité. Nous ne pouvons pas nous permettre de ralentir politiquement cette technologie et nous ne devrions certainement pas bloquer son développement. Mais comme une bonne intelligence artificielle est une intelligence éthique, nous avons également l'obligation morale d'améliorer nous-mêmes les directives du groupe de haut niveau. En dépit de toutes les critiques potentielles sur la manière dont elles ont été créées, les directives éthiques que nous développons actuellement en Europe sont actuellement la meilleure plate-forme disponible au monde pour la prochaine phase de discussion. La Chine et les États-Unis vont les étudier de près.

Leur ancrage juridique dans les valeurs fondamentales européennes est excellent et la première sélection de principes éthiques abstraits est à minima acceptable. Seule la véritable substance normative au niveau des risques à long terme, des applications concrètes et des études de cas a été détruite. La première étape est bonne. Mais il est grand temps que les universités et la société civile reprennent le processus en main et retirent des mains le débat auto-organisé de l'industrie.

Tout le monde le ressent : nous sommes dans une transition historique rapide qui se déroule à plusieurs niveaux simultanément. La fenêtre d'opportunité à l'intérieur de laquelle nous pouvons au moins partiellement contrôler l'avenir de l'intelligence artificielle et défendre efficacement les fondements philosophiques et éthiques de la culture européenne se refermera dans quelques années. Nous devons agir maintenant.