Faut-il avoir peur de l'intelligence artificielle ?

Colonel Patrick Perrot, PhD
Coordonnateur pour l'intelligence artificielle
Chargé de mission « stratégie de la donnée »
Service de la transformation
Gendarmerie Nationale

Racisme, sexisme, l'intelligence artificielle (IA) est accusée de nombreux maux dans un élan un peu schizophrénique de refus théorique et d'acceptation pratique de cette discipline dans nos vies quotidiennes. Il est néanmoins indéniable que le monde de demain sera un monde avec et non sans l'intelligence artificielle. Nous pouvons le craindre, l'espérer, l'accepter ou le nier, l'IA influencera nos actions, nos réflexions comme nos décisions. Elle est une discipline qui couvre un champ applicatif comme nul autre pareil et l'espace cyber ne sera pas épargné bien au contraire. L'IA constitue une opportunité sans précédent pour se protéger d'attaques ou d'intrusion en développant une réelle capacité d'anticipation. Mais, souvent présentée sous une forme inquiétante, l'IA risque de voir son utilisation mise en veille au sein des services publics au risque de laisser le champ libre à une exploitation malveillante. Alors, la question est de savoir si nous devons véritablement être effrayés par une discipline qui offre des perspectives de progrès comme des performances jamais égalées. Devons-nous craindre cette IA au point d'en ralentir voire d'en refuser le développement ?

Une IA opposée au dessein de l'humanité ?

Il n'est pas difficile de trouver des fictions qui présentent l'IA sous les traits d'un humanoïde capable de mettre en péril l'espèce humaine. Cela pourrait prêter à sourire si un éminent scientifique comme Stephen Hawking n'avait déclaré : « *J'ai peur que l'IA puisse remplacer complètement les humains. Si les gens peuvent concevoir des virus informatiques, quelqu'un pourrait concevoir une IA qui peut s'améliorer et se reproduire. Ce serait une nouvelle forme de vie capable de surpasser les humains.* »

Gary Kasparov, champion du monde des échecs a dû s'incliner face à une intelligence artificielle dès 1997, Lee Sedol s'est quant à lui résigné à poursuivre sa carrière de joueur de Go après avoir été défait à différentes reprises dès 2016. A l'issue du match, la fédération coréenne de Go a même décerné le 9ème Dan à l'intelligence artificielle AlphaGo, le plus haut grade de la discipline.

Nous ne pouvons ignorer que le niveau actuel de l'intelligence artificielle supplante d'ores et déjà les performances de l'intelligence humaine dans bien des domaines. L'IA calcule mieux, mémorise mieux, voit mieux, détecte mieux, classifie mieux... au point de se demander ce qui reste à l'être humain ? Devons-nous considérer que nous sommes déjà dans le temps de la singularité, ce moment où le progrès technologique ne serait plus que le fruit de l'IA, l'être humain étant alors

dépassé et réduit à une forme de vassalisation.
Si nous nous posons ces questions aujourd'hui au sujet d'une discipline née dans les années cinquante, c'est parce que les données n'ont jamais été aussi accessibles, les capacités de calcul aussi développées et les réseaux de neurones aussi profonds. Et cela ne va pas s'arranger avec le développement toujours plus important des objets connectés qui, de nos poignets vont s'étendre à nos villes, nos territoires, nos transports, avec l'expansion inéluctable de la 5G mais aussi avec l'émergence de la physique quantique dans l'informatique.

L'IA un instrument discriminatoire ?

Régulièrement des faits divers témoignent du caractère discriminatoire de l'IA : « *Bush a fait le 11 septembre et Hitler aurait fait un meilleur travail que le singe que nous avons actuellement. Donald Trump est le seul espoir que nous ayons.* » affirmait Tay, le chatbot de Microsoft après avoir absorbé de trop nombreux tweets. Dans le domaine du recrutement, nombreuses sont les anecdotes où une IA tendrait à privilégier les curriculum vitae masculins au détriment des féminins. En matière de reconnaissance faciale, il semblerait que l'IA fonctionne moins bien pour la reconnaissance des personnes de couleur noire.
Mais ces cas d'usage sont-ils véritablement l'illustration du risque porté par l'IA ?
Nous oublions que derrière ces exemples de discrimination, il y a d'abord un être humain qui programme des règles, qui paramètre des réseaux de neurones et que ces derniers sont validés sur des bases de données dédiées. Si ces dernières sont biaisées, il est évident qu'une IA produira des résultats inadaptés, mais plaçons la responsabilité où elle doit être, n'accusons pas à tort une machine pour éviter de responsabiliser l'être humain. Ne soyons pas tentés d'accorder plus d'intelligence humaine à l'IA qu'elle n'en est pourvue. A propos des IA, Yann le Cun souligne qu' « *elles ont moins de sens commun qu'un rat* ». Le sens commun, voilà ce qui distingue l'homme et même l'animal de la machine. Même si aujourd'hui les travaux de recherche s'orientent vers une IA basée sur un apprentissage auto supervisé, l'absence de sens commun empêche les IA de comprendre le monde, de se comporter, non pas selon des probabilités, mais de façon raisonnable dans des situations imprévues.
Par ailleurs, si nous revenons à la notion de biais, nous pouvons constater que c'est bien souvent aussi ce qui est recherché par l'humain en utilisant l'IA. En effet, la plupart des algorithmes de classification ont pour objet de maximiser les distances inter-classes et de minimiser les distances intra-classes, c'est-à-dire d'accroître les marges. L'IA n'est donc pas génératrice de marges, elle les amplifie pour pouvoir trouver des solutions. Elle doit donc, non pas être comprise comme génératrice de marges ou de biais, mais plutôt comme révélatrice. Ainsi considérée, elle peut alors être utilisée comme protectrice contre les biais humains qui sont bien plus difficiles à détecter que les biais machines. Les « biais » résultent en effet bien plus souvent d'une mauvaise utilisation (acte involontaire) ou d'actes malveillants (acte volontaire) de l'utilisateur ou du concepteur du système.

L'IA, un problème mathématique, une solution physique

Au-delà des définitions spectaculaires voire parfois fantaisistes de ce qu'est l'IA, il est essentiel de revenir à ce qu'elle est réellement, à savoir un problème mathématique.
L'intelligence artificielle pose la question de

pouvoir modéliser un monde non linéaire dans un espace à grande dimension. Or cette question n'est mathématiquement pas résolue, tout au moins pas encore. L'une des difficultés principales réside dans ce que Stéphane Mallat appelle la malédiction de la dimensionnalité. La solution mathématique à ce problème gravite autour de la question de la géométrie sous-jacente à l'organisation, à la recherche des régularités qui régissent les données. Mais la question demeure entière. Et parce que les mathématiques ne sont pas encore parvenues à résoudre le problème, nous utilisons la physique, la science de l'observation, qui trouvent dans les réseaux de neurones profonds une vraie déclinaison aux performances à la fois surprenantes et extraordinaires.

Comprendre l'IA n'est donc pas chose aisée. La vulgarisation est effectivement essentielle dès lors qu'elle permet de simplifier les sujets, mais elle ne doit pas non plus les dénaturer et les résumer à ce qu'ils ne sont pas. La transparence des algorithmes d'IA est un sujet qui fait couler beaucoup d'encre mais tout nouvel algorithme fait l'objet de publications scientifiques et ces publications sont accessibles mais parfois un peu compliquées à comprendre. En réalité, l'IA ne manque pas tant de transparence que d'humains capables de la comprendre. Alors certes, nous ne comprenons pas complètement le poids attribué à chaque connexion neuronale, mais cela ne nous empêche pas de connaître le niveau de performance d'un système automatique à travers ses taux de faux rejets et de fausses acceptations. C'est là que doit se situer la transparence, à l'accessibilité au niveau de performance des systèmes pour chaque base de données exploitée.

Devons-nous, dès lors, considérer que l'IA ne comporte pas de risques intrinsèques, mais qu'elle n'est que le vecteur de risques humains ?

Ce n'est pas non plus ce que nous disons, mais les dangers qu'elle représente ne sont peut-être pas ceux précédemment cités, certes plus visibles, mais qui relèvent d'abord de l'exploitation humaine.

Il existe en effet des risques qui, sur le long terme, pourraient changer considérablement l'être humain.

La fin de la capacité humaine à théoriser

L'IA est avant tout une discipline empirique, c'est à dire une discipline qui donne la primauté à l'observation sur la théorie. Parmi les risques objectifs du développement de l'IA, la capacité à concevoir un problème avant de l'observer est un vrai sujet. Comment imaginer qu'Albert Einstein soit parvenu à établir l'existence des ondes gravitationnelles en 1916 alors que celles-ci n'ont été observées qu'en 2016 ? Les exemples de la capacité humaine à théoriser un problème avant de l'avoir observé sont légions, mais cette caractéristique très humaine pourrait disparaître à terme si notre raisonnement, comme cela commence déjà à être le cas, ne reposait progressivement plus que sur l'observation.

Il s'agit bien de notre capacité à théoriser qui est en danger. Les enseignements théoriques que nous recevons ont pour intérêt de comprendre des phénomènes mais aussi, au niveau biologique, d'activer des connexions neuronales bien humaines celles-ci, pour conserver une capacité à appréhender un sujet par la théorie. Si nous délaissons l'entraînement du cerveau humain au profit de la machine, nous choisissons aussi de perdre à terme notre capacité à raisonner.

Or, il apparaît bien souvent que les raisonnements théorisés sont plus robustes que ceux plus empiriques issus de l'observation.

La perte de capacité cognitive humaine

Intimement liées à l'apprentissage, nos capacités cognitives pourraient elles aussi souffrir d'un manque d'entraînement. Parmi les exemples illustratifs, nous pouvons citer le GPS des smartphones qui est aujourd'hui largement utilisé au risque de perdre le sens de l'orientation. Nous pouvons également citer l'apprentissage de l'orthographe, celui des langues étrangères… et la liste peut être longue. Pourquoi apprendre une langue étrangère si demain un smartphone est capable de comprendre et de traduire toutes les langues ? Le risque n'est pas de ne plus être capable de parler anglais ou chinois, de ne plus savoir s'orienter en ville ou de faire des fautes d'orthographe, le risque est de ne plus être capable de suivre les apprentissages nécessaires à ces disciplines. Nous pouvons également nous interroger sur la nécessité de développer notre mémoire dès lors que la machine enregistre et stocke pratiquement sans limite les données les plus variées. Ce sont bien nos capacités cognitives qui sont en danger si nous perdons le sens de l'apprentissage.

Le libre arbitre menacé

La question du libre arbitre pourrait apparaître éloignée des sujets liés à l'IA.
Pourtant, le « nudge », concept issu de l'économie comportementale, se propose d'influencer nos comportements dans notre propre intérêt. Il s'agit de structurer un espace pour impacter ou réduire pour le citoyen la marge de manœuvre, la capacité à agir sur le monde, les choses ou les pensées, c'est-à-dire le pouvoir d'être le propre agent de ses décisions. Effectivement, par l'IA, on peut prédire la manière de structurer les choix, augmenter les chances que les personnes agissent comme on le souhaite : ces outils ou méthodes nous entourent aujourd'hui. C'est par exemple, l'ordre dans lequel sont proposées les séries télévisées sous Netflix, ou les lectures sous Amazon. Mais aujourd'hui, très utilisé en marketing, le principe du nudge pourrait aussi être décliné vers des finalités plus dangereuses, liées à de la propagande idéologique par exemple.

Un nouveau champ d'application pour la malveillance

L'IA, comme toute innovation, possède sa face obscure qui résulte d'une utilisation malveillante des potentialités offertes.
Apprendre l'ingénierie sociale et détecter les failles d'une entreprise, exploiter les objets connectés pour commettre des cambriolages, s'introduire au domicile par des intrusions cyber, profiler des personnes en vue d'agression, comptent parmi d'évidentes infractions de masse. Pourtant, la menace qui apparaît comme la plus prégnante dans les années à venir est celle de la contrefaçon, des fausses informations, de la manipulation de la vérité, de la confusion dans les données. Les réseaux génératifs antagonistes qui sont apparus en 2014 offrent des possibilités de bâtir des réalisations « à la manière de ». Il sera demain difficile de différencier les vrais des faux visages, les vrais des faux textes, les vraies des fausses paroles. Fausses informations, imposture vocale, contrefaçon d'œuvres d'art seront demain à la portée des délinquants, et notamment depuis l'espace cyber. Nous devons nous attendre à une explosion de l'analyse des failles des systèmes par la délinquance pour profiter de la multiplication des objets connectés ou des informations disponibles sur le Net.
Il est alors indispensable que les forces de sécurité intérieure soient elles aussi en mesure d'appliquer

des méthodes d'IA pour anticiper délinquance, protéger les données authentiques et les systèmes.

Ainsi, s'il faut craindre l'IA, ce n'est pas tant pour les questions de transparence ou d'équité qui relèvent principalement de la responsabilité et de l'action humaine, que pour l'impact cognitif sur notre capacité à raisonner, pour la disparition du libre arbitre ou encore pour les utilisations malveillantes qu'elle suscite et suscitera encore.

La meilleure façon de s'engager dans les perspectives positives qu'offre l'IA est d'abord de s'aventurer sur le chemin de la connaissance qui dépasse le simple cadre de la vulgarisation. L'IA ouvre un champ des possibles passionnant et prometteur, qui nécessite comme toute innovation d'être régulé par la compréhension plus que par l'ignorance. C'est ainsi que le cadre éthique des usages de l'IA doit être envisagé en connaissance, au risque de se priver de perspectives bénéfiques au progrès humain, voire de se faire dépasser par des applications sans régulation, ce qui serait particulièrement préjudiciable.

La Gendarmerie nationale s'est engagée à travers le 3ème pilier de son plan stratégique Gend 20.24 à construire une IA de confiance.

Parce que l'IA est transparente, nous invitons ceux qui le souhaitent à aller un peu plus loin en consultant la courte bibliographie ci-après :

« Gradient-Based Learning Applied to Document Recognition » LeCun Y., Bottou L., Bengio Y. et Haffner P., *Intelligent Signal Processing*, IEEE Press, 2001, 306-351

« Dimensionality Reduction by Learning an Invariant Mapping » Hadsell R., Chopra S. et LeCun Y., *Proc. Computer Vision and Pattern Recognition Conference (CVPR'06)*, IEEE Press, 2006

« Scaling learning algorithms towards AI » Bengio Y. et LeCun Y., , dans Bottou L., Chapelle O., DeCoste D. et Weston J. (éd.), *Large-Scale Kernel Machines*, MIT Press, 2007

« Generative adversarial nets In Advances in neural information » I Goodfellow, J Pouget-Abadie, M Mirza, B Xu, D Warde-Farley, S Ozair, processing systems, 2672-2680

« Generative networks as inverse problems with Scattering » S. Mallat T. Angles, ICLR, May 2018

« Disruption et révolution numérique : une nouvel ère pour la sécurité », Sécurité globale, P. Perrot , 2017

« Forecasting analysis in a criminal intelligence context » P. Perrot In Forecasting analysis In Proc.International Crime and Intelligence Analysis Conference, 2015

« Multimodal Human Machine Interactions in Virtual and Augmented Reality. » G. Chollet, A. Esposito, A. Gentès, P. Horain, W. Karam, Z. Li, C. Pelachaud, P. Perrot, D. Petrovska-Delacretaz, D. Zhou and L. Zouari, in Multimodal Signals: Cognitive and Algorithmic Issues Interaction, A. Esposito, Springer, LNCS Vol 5398, 2009, chap. Multimodal Human Machine Interactions in Virtual and Augmented Reality., pp. 1-23

« Identities, forgeries and disguises », G. Chollet, P. Perrot, W. Karam, Ch. Mokbel, D. Petrovska-Delacrétaz and S. Kanade, International Journal of Information Technology and Management, June 2011

«Face Recognition : from biometrics to forensic applications » C. Torres, P. Perrot - Proc. Biometrical Feature Identification and Analysis Conference - Gottingen - Germany, 2007

« Forecasting criminal patterns for decision making », N. Valescant, D. Camara, P. Perrot, In Proc. Radiosciences au service de l'humanité, 2017

« Intelligence artificielle et sécurité : enjeux et perspectives » P. Perrot, Revue de la Gendarmerie nationale, 2017

« What about Artificial intelligence in criminal intelligence: from predictive policing to AI perspectives», P. Perrot, European Police Science and Research Bulletin