

L'intelligence artificielle éthique est-elle une utopie ?

Démarche

J'ai choisi de traiter le thème de l'intelligence artificielle car c'est un sujet d'actualité passionnant mais également déconcertant. La puissance de calcul des ordinateurs de plus en plus importante permet des applications de l'IA qui semblaient inconcevables il y a encore dix ou vingt ans. Des faits d'actualités polémiques, mettant en cause la confidentialité des données traitées par les algorithmes et les IA, soulèvent de multiples questions d'ordre éthique, juridique et même philosophique. Réfléchir à l'éthique de l'IA c'est réfléchir principalement aux fondements éthiques des concepteurs. Il m'a semblé pertinent de réfléchir à ces questionnements dans notre manière de penser l'IA en tant que manager de demain. Les thématiques abordées par l'avènement de l'IA concernent tout le monde et chacun peut s'insurger contre les pratiques qui ne sembleraient pas justes ou éthiques.

Nous serons tous confrontés demain à des choix sensibles faisant intervenir l'IA et nous pouvons déjà nous demander comment nous allons utiliser l'IA et s'en servir pour faire des choix éthiques. Pour illustrer mon propos et le rendre plus parlant, j'ai choisi des exemples dans différents domaines. Mon essai a pour objectif de mettre en évidence certaines limites de l'IA qui m'ont fait réfléchir sur la possibilité ou non, de construire une IA éthique.

Résumé

D'abord, je présenterai la place de l'éthique dans la conception de l'IA aujourd'hui. J'expliquerai en quoi les craintes autour de la technologie ne sont pas nouvelles, puis comment l'IA nous influence aujourd'hui et tend à prendre de plus en plus de place dans notre vie professionnelle.

Ensuite, j'expliquerai les raisons qui font du questionnement éthique dans l'IA une priorité : les algorithmes sont le fruit de formalisations d'objectifs en termes mathématiques qui peuvent être sources de biais. De plus, j'aborderai la grande lacune cognitive des algorithmes face à l'homme.

Enfin, je proposerai quelques solutions pour rendre l'IA plus éthique. La sensibilisation des concepteurs, ainsi qu'une mixité dans les équipes peut contribuer à réduire les biais. La sensibilisation des utilisateurs, est également nécessaire pour que chacun ait en tête les limites des algorithmes.

Bibliographie indicative

- Bernheim A., Vincent F. (2019). *L'intelligence artificielle, pas sans elles ! Faire de l'intelligence artificielle un levier pour l'égalité*, Belin.
- Carr N. (2017). *Remplacer l'humain : critique de l'automatisation de la société*, L'Échappée.
- Casilli A. A. (2019), *En attendant les robots : enquête sur le travail du clic*, Seuil.
- Département du transport. (2019) *Threat Image Protection (TIP) Management Software*, [En ligne]. Disponible sur < www.gov.uk/government/publications/approved-threat-image-projection-management-software > [Accédé le 24 janvier 2021]
- Encyclopédie Larousse. (2021). *Intelligence artificielle et sciences cognitives*. [En ligne]. Disponible sur < www.larousse.fr/encyclopedie/divers/intelligence_artificielle/187257 > [Accédé le 24 janvier 2021]
- Levesque H. J. (2013). *On our best behaviour*, intervention à l'International Joint Conference on Artificial Intelligence, Chine.
- Povyakalo A. A. (2013). *How to discriminate between computer-aided and computer-hindered decisions: a case study in mammography*, Medical Decision Making, 33, n°1
- Villani C. (2018). *Donner un sens à l'intelligence artificielle*. [En ligne]. Disponible sur < www.aiforhumanity.fr > [Accédé le 24 janvier 2021]

Introduction

Depuis les années 1950 et le développement de l'informatique, nous assistons à l'avènement de l'intelligence artificielle (IA) qui trouve aujourd'hui de nombreuses applications dans de multiples domaines de notre vie quotidienne. Au-delà des représentations futuristes des films de science-fiction, l'IA est aujourd'hui encore loin des robots humanoïdes conscients et indépendants cherchant à exterminer l'espèce humaine. Elle prend la forme d'applications ou d'algorithmes qui nous aident au quotidien. Bon nombre d'entre nous sont aujourd'hui amenés à utiliser des outils utilisant l'IA : l'assistant vocal de notre smartphone pour connaître la météo du lendemain, le GPS calculant l'itinéraire le plus rapide pour se rendre sur notre lieu de travail ou encore les chatbots sur certains sites web qui répondent à nos questions.

L'utilisation de l'IA peut aussi intervenir dans des domaines plus sensibles. La médecine, où les décisions prises doivent être les plus éclairées possibles, est souvent l'un des cas extrêmes invoqué pour montrer les limites de l'IA. Une machine peut-elle être aussi compétente qu'un médecin pour réaliser un diagnostic médical ? On constate que la question éthique se pose : la recommandation faite par un algorithme a-t-elle autant de valeur et de rationalité qu'une recommandation faite par un individu humain ?

En 2018, l'ancien Premier Ministre Édouard Philippe a confié à Cédric Villani, mathématicien et député, la rédaction d'un rapport visant à « donner un sens à l'intelligence artificielle », en coordonnant des stratégies nationales et européenne. Cette démarche montre que ces questionnements éthiques ne concernent plus les seuls philosophes de l'IA, mais aussi les représentants politiques et notre gouvernement. Ce sont des questions sur lesquelles nous pouvons d'ores et déjà réfléchir afin de nous demander vers quel type d'IA éthique nous souhaitons évoluer.

I. Quelle est la place de l'éthique dans l'intelligence artificielle aujourd'hui ?

« La création d'une intelligence artificielle serait le plus grand événement de l'histoire de l'humanité. Mais il pourrait aussi être l'ultime. » - Stephen Hawking

A. La crainte de l'intelligence artificielle n'est pas nouvelle

L'IA peut être définie aujourd'hui comme « l'ensemble des théories et techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence humaine » (Larousse, 2021).

En 1997, l'entreprise américaine IBM impressionne la communauté scientifique et marque un jalon dans l'histoire de l'IA avec son super ordinateur Deep Blue, qui bat le champion du monde en titre du jeu d'échecs, Garry Kasparov. Cet événement majeur était prometteur car il était le fruit d'années d'études et d'expérimentations sur l'apprentissage automatique par les ordinateurs, le *machine learning*. Mais cela a également soulevé de grosses inquiétudes car jusqu'à présent, personne ne pensait qu'un ordinateur puisse battre un humain à un jeu aussi complexe et stratégique que les échecs, qui plus est, le champion du monde. La peur d'être remplacé par les robots s'est renforcée et la crainte d'une intelligence qui pourrait surpasser l'intelligence humaine s'est développée.

De tous temps, que ce soit avec la machine à vapeur ou la taylorisation de la chaîne de production, l'Homme a toujours craint de voir ses créations technologiques le dépasser et prendre le contrôle. Aujourd'hui, cette crainte de l'IA s'illustre particulièrement dans le domaine de la médecine, où certains craignent qu'un jour les consultations soient remplacées par un algorithme qui réalisera un diagnostic en fonction de nos symptômes, sans auscultation physique par un médecin traditionnel. C'est ici que nous percevons les limites éthiques et la pertinence d'une sensibilisation à ces enjeux : il s'agit de réfléchir à la compatibilité de l'éthique professionnelle avec les algorithmes. L'éthique professionnelle a pour but de réglementer les activités qui ont lieu dans le cadre d'un métier. Elle étudie les contraintes de la déontologie professionnelle. Par conséquent, est-ce que les choix et recommandations faits par l'IA seront

toujours alignés avec le système de valeur et la déontologie que nous souhaitons mettre en place dans notre travail ? Et comment se prémunir des éventuels biais de l'IA qui pourraient influencer nos décisions professionnelles ?

Là est tout l'objet de cette réflexion, comprendre comment l'IA peut nous aider à prendre de meilleures décisions en concordance avec la déontologie, sans craindre qu'elle puisse nous remplacer un jour.

B. L'intelligence artificielle influence déjà nos choix aujourd'hui et tend à devenir de plus en plus puissante

Aujourd'hui, nous utilisons tous quotidiennement, parfois sans le savoir, des outils utilisant l'IA. Nous nous laissons volontiers guider et influencer par les recommandations de ces ordinateurs pour notre confort ou notre sécurité. Lorsqu'il s'agit de circonstances précises, limitées et qui ne nous semblent pas d'une grande importance, la remise en question de l'IA n'est que marginale. En revanche, dans des circonstances plus sensibles telle que la justice, les décisions ont des conséquences directes sur l'avenir des personnes. Les remises en question sont bien plus ardentes, à juste titre.

Aux États-Unis, un logiciel est à disposition des juges afin de les aider à déterminer le montant de la caution ou la peine des prévenus. L'algorithme COMPAS génère un score pour estimer la probabilité que le détenu commette une nouvelle infraction dans les deux ans à venir. Des journalistes de l'ONG ProPublica ont analysé des milliers de scores générés par l'algorithme : quasiment deux fois plus d'accusés noirs que d'accusés blancs correspondaient à des « faux positifs » et étaient identifiés à tort par le programme comme « à haut risques de récidive ». Cet exemple montre que le score donné par l'algorithme n'est pas suffisant lorsqu'il est pris seul hors contexte, pour prendre une décision aussi grave et subjective qu'un jugement de peine. Certains jugements prennent en compte des circonstances atténuantes ou aggravantes difficilement formalisables pour un ordinateur.

Par conséquent, nous constatons que l'IA s'introduit dans des domaines beaucoup plus importants et sensibles que nos seuls objets connectés et peut donc avoir de graves conséquences sur la vie des personnes si elle n'est pas suffisamment encadrée. Des recommandations erronées peuvent être faites si on s'appuie uniquement sur une base de données biaisée : ici la surreprésentation des personnes noires en prison qui fait « croire » à l'algorithme qu'il existe un lien de causalité entre couleur de peau et risque de récidive.

II. Pourquoi rendre l'intelligence artificielle éthique ?

*« La précision des algorithmes va de pair avec l'étroitesse de leur champ de perception. »
Nicholas Carr*

A. Les algorithmes sont un reflet de la société et de leur concepteur : ils sont empreints des mêmes biais

L'éthique fait référence à la réflexion et la prise de décision humaine : il s'agit de réfléchir à ce qui est bien et ce qui est mal. Les machines et les algorithmes font référence à la technologie, et qui par définition, ne ressentent pas d'émotions, de sensations et ne pensent pas. On pourrait penser que les algorithmes seraient objectifs et égaux car ils n'ont pas d'opinions morales ou politiques. Il faut néanmoins garder à l'esprit que les algorithmes sont codés par des êtres humains et peuvent donc être empreints de la même imperfection et des mêmes biais que la réflexion humaine.

Les applications de l'IA et l'amélioration continue sont permises par différentes méthodes d'apprentissage automatique. L'apprentissage supervisé, par exemple, concerne des tâches précises comme la classification et l'étiquetage d'images, de textes, de sons ou de vidéos. L'ordinateur classe les éléments et des opérateurs humains vont vérifier certains échantillons pour s'assurer que les données sont correctement classées et ainsi renforcer le

modèle de référence de l'ordinateur. Par conséquent, si les informations de référence sont biaisées, alors l'algorithme s'appuiera sur ces données préalablement étiquetées comme références pour délivrer une recommandation qui sera de fait, également biaisée.

En 2017, une polémique est déclenchée sur les réseaux sociaux : il est démontré que le logiciel de traduction automatique Google Traduction réalise des traductions stéréotypées. Le turc est une langue non genrée, or on constate que des traductions ont été instantanément genrées dans la version anglaise : la « personne heureuse » en turc, devient un « homme heureux » en anglais, la « personne malheureuse », une « femme malheureuse ». L'algorithme de traduction n'est pas sexiste délibérément, mais il s'appuie sur les expressions les plus populaires et les suggestions des utilisateurs. Si le concepteur estime que traduire « une personne malheureuse » par « une femme malheureuse » est pertinent alors l'algorithme suggèrera cette traduction systématiquement bien qu'elle est erronée. De la même manière que l'illustration précédente, l'algorithme s'auto alimente à partir de données de références fausses si l'opérateur humain n'intervient pas pour le corriger.

Ainsi, tout comme l'intelligence humaine, l'IA est conditionnée par ses interactions avec son environnement et par tous les utilisateurs qui l'influencent et lui fournissent les données dont elle se nourrit. A chaque étape de leur développement, les algorithmes récupèrent les préjugés de leurs développeurs et utilisateurs et les retranscrivent.

B. L'IA ne remplace pas les sens et la perception humaine

Selon Nicholas Carr, l'intelligence humaine est caractérisée par « la capacité à interpréter les phénomènes et leur donner du sens ». En intégrant les connaissances issues de l'observation et de notre expérience personnelle du monde qui nous entoure, nous effectuons des tâches, résolvons des problèmes et emmagasinons des expériences et des souvenirs pour appréhender le monde et ses difficultés. Hector Levesque, informaticien et roboticien prend l'exemple d'une question simple pour illustrer cette problématique : « La grosse balle est passée au travers de la table car elle est en polystyrène. Qu'est ce qui est en polystyrène ? La table ou la grosse balle ? ». En tant qu'humain, nous pouvons répondre à cette question et en déduire que c'est la table qui est en polystyrène car d'expérience nous savons que ce matériau est fragile et qu'une « grosse balle » peut donc facilement passer au travers. Cette perception et cette expérience sont difficilement paramétrables dans un algorithme.

III. Quelles sont les solutions concrètes pour contribuer à rendre l'intelligence artificielle éthique ?

« Celui qui code le système intègre sa vision du monde. Les visions limitées créent des systèmes limités. Codons avec un regard universel ! » Joy Buolamwini, fondatrice de l'Algorithmic Justice League

A. Sensibiliser les concepteurs au codage éthique et prôner la diversité dans les équipes de développement

Comme vu précédemment, les IA sont des algorithmes développés par des humains ayant eux-mêmes des préjugés et des stéréotypes qu'ils reproduisent dans leurs codes. Sensibiliser les développeurs et concepteurs d'algorithmes à ces biais peut être une première étape vers la réduction des biais dans les algorithmes. Les algorithmes reflètent la société dont ils sont issus et peuvent participer à renforcer et diffuser des stéréotypes. La publicité ciblée réalisée par Google par exemple est si omniprésente que les images véhiculées peuvent avoir des effets réels sur nos décisions. La publicité genrée contribue à renforcer des stéréotypes sexistes. En effet, encore aujourd'hui les publicités pour les couches et les biberons sont beaucoup plus présentées aux femmes qu'aux hommes, et ce, qu'elles aient effectué des recherches sur le sujet (et donc démontré leur intérêt sur ce produit) ou non. Des chercheurs américains ont étudié les algorithmes de ciblage publicitaire de Google et ont révélé que les

hommes se voyaient plus fréquemment afficher des publicités promettant des salaires élevés que les femmes. Si un algorithme sur une plateforme de recherche d'emploi propose systématiquement à une femme des offres d'emploi moins rémunérées qu'un homme, alors elle sera plus susceptible d'intégrer l'idée qu'une femme a toujours une rémunération plus faible qu'un homme.

Pour que les algorithmes reflètent la réalité de la diversité, il faut que leurs concepteurs soient sensibilisés aux différentes représentations des futurs utilisateurs. De plus, les bénéfices de la mixité sur la productivité sont prouvés. Des individus de genres et d'origines variées auront différents points de vue pour répondre à un même problème et pourront collaborer tous ensemble pour créer des solutions inclusives. En 2018, Microsoft a mis à jour son application FaceDetect pour réduire les biais des algorithmes de reconnaissance faciale en intégrant plus de photos des groupes sous-représentés. Certains utilisateurs avaient alerté Microsoft, car des personnes noires étaient parfois reconnues par l'algorithme comme étant des gorilles. C'est principalement parce qu'une femme noire utilisait cette application au MIT, que le problème a pu être mis en lumière et corrigé. Cela illustre parfaitement l'importance d'une équipe diversifiée pour créer des outils inclusifs. Cette évolution et diversification au sein des équipes induit une évolution nécessaire de l'image genrée de certaines disciplines et métiers. Une vigilance des entreprises et des réglementations étatiques concernant la mixité des recrutements, des grilles de salaires, des formations et des promotions sont de bons outils pour s'engager contre les stéréotypes véhiculés par l'IA. Des initiatives lancées par certaines associations comme EllesBougent, GirlsInTech ou TechGirlz essaient de casser ces biais sexistes en sensibilisant les plus jeunes, dès le collège et le lycée pour encourager les femmes dans des carrières scientifiques et ainsi contrer la sous-représentation actuelle.

Pour Aude Bernheim et Flora Vincent, « il est plus facile de changer des lignes de code que des mentalités. Transformons l'intelligence artificielle et ses concepteurs pour transformer toute la société. ».

B. Sensibiliser les utilisateurs aux limites de l'IA

Selon Jean-Gabriel Ganascia, « il y a une illusion selon laquelle la machine aspire le monde et [nous] donne un résultat objectif ». En effet, plus les systèmes informatiques sont perfectionnés et plus nous avons tendance à croire que nos erreurs sont systématiquement corrigées et que les résultats donnés par la machine sont toujours justes. Plusieurs études ont démontré que la confiance accordée à ces logiciels limitent notre capacité d'attention et notre concentration. Dans certains services de radiologie, un logiciel aide les médecins à repérer les zones auxquelles le médecin doit prêter une attention particulière pour repérer les tumeurs cancéreuses. Or, il a été prouvé que cela peut avoir un effet inverse car les médecins accordent une moindre importance aux zones que le logiciel n'a pas mis en évidence et passent à côté d'éventuelles anomalies.

Par conséquent, il est possible de se laisser induire en erreur par excès de confiance envers les algorithmes à qui on donne à tort une capacité de jugement infaillible. Comme nous avons vu précédemment, les algorithmes sont pleins de biais et peuvent influencer nos décisions personnelles et professionnelles. La solution possible ici serait de toujours avertir les utilisateurs de la faillibilité des algorithmes et d'utiliser des astuces pour que les utilisateurs continuent à interagir avec leur environnement. Par exemple, dans les aéroports, lors du contrôle des bagages à main par rayon X, les scanners sont dotés d'un système TIP (Threat Image Projection) qui ajoute aléatoirement de fausses images de menaces (armes, couteaux...) aux images des bagages. L'agent relève alors la menace et l'ordinateur lui révèle si elle est virtuelle ou non. Cela permet de stimuler la vigilance de l'agent et ainsi de ne pas le laisser devenir totalement passif face au logiciel d'aide à la décision. Cette astuce est applicable dans d'autres domaines comme celui de la radiologie comme nous l'avons vu précédemment. Nous faisons rapidement le lien avec l'éthique professionnelle dont nous devons faire preuve,

Ce n'est pas parce que l'algorithme « intelligent » nous fait une recommandation fondée sur des calculs complexes difficiles à refaire nous-mêmes que la recommandation est

forcément la meilleure. L'IA a largement contribué à améliorer la sécurité de nos avions et de nos voitures mais elle a aussi entraîné une érosion de nos capacités cognitives lorsque nous savons que nous sommes assistés par l'ordinateur. Selon moi, il faudrait constamment garder à l'esprit que ce sont bien des programmes d'aide à la décision et non pas de décisions tout court. Il s'agit de bénéficier de la précision et de la puissance de calcul permises par les ordinateurs tout en utilisant tous nos sens cognitifs que les machines n'ont pas. La décision finale ne dépend que de nous, les utilisateurs, qui sommes bien humains.

Conclusion

Ainsi, nous avons vu quelles étaient les grandes questions éthiques à se poser avec la montée en puissance de l'IA, puis dans quelle mesure il était nécessaire de réfléchir à ces questions essentielles pour que les algorithmes nous aident à faire nos choix dans une dynamique plus égalitaire. La reproduction des imperfections de l'IA est évitable et il existe plusieurs solutions pour faire évoluer l'IA vers une éthique en adéquation avec la morale que nous souhaitons prôner dans notre société et dans notre déontologie.

Pour moi, cette réflexion à propos de l'éthique professionnelle, ne concerne pas seulement la question d'agir bien ou mal en son âme et conscience. Cela concerne également la sensibilisation et la vigilance portée aux outils que nous utiliserons tous les jours dans notre travail et qui nous influencent pour prendre des décisions stratégiques. Le progrès et les possibilités technologiques toujours plus modernes vont induire de gros changements dans notre manière de travailler et d'interagir avec les autres. Finalement, je pense qu'une IA éthique n'est pas une utopie. C'est un objectif réalisable dont nous devons nous approcher le plus possible. En étant conscients des biais, nous pouvons contribuer à les réduire. Même si nous n'utilisons pas directement tous ces algorithmes, l'IA reflète la société dans sa globalité avec les préjugés inhérents à tous les membres de cette société. Et en tant qu'individu membre de cette même société dans laquelle les IA sont développées, il n'appartient qu'à nous d'agir pour contribuer à insuffler de l'éthique dans l'IA. Nous pouvons chacun, à différentes échelles, participer à rendre l'IA plus inclusive et éthique.