La nature duale de l'IA : empowerment / asservissement

67- 04/09/2023 **L'usage de l'Intelligence Artificielle se répand massivement dans tous les secteurs, de la santé au divertissement, en passant par la politique, la défense, la sécurité, l'agriculture, l'environnement, la fourniture de biens et de services, les transports, l'information et la finance. S'immisçant dans notre intimité, l'IA améliore et voire dépasse nos capacités dans des proportions sans précédent… Au risque de nous rendre si dépendants que nous pourrions perdre le contrôle de notre avenir. L'IA est-elle un fantastique levier d'émancipation pour l'humanité ou va-telle nous réduire en servitude (volontaire ou inconsciente) ? Pour en savoir plus, Selfpower Community s'est tourné vers le professeur Joanna Bryson. La rencontre a eu lieu dans son bureau de la Hertie School (Berlin Center), quelques étages au-dessus de la très animée Friedrichstraße. Ses réflexions aident au discernement, une responsabilité qui incombe à tout scientifique qui se reconnaît comme un acteur engagé dans la société civile.**

Spécialiste de l'intelligence artificielle (IA), le Pr. Joanna Bryson étudie des questions cruciales concernant les potentiels de la technologie et la préservation des valeurs humaines. En tant qu'activiste, elle plaide pour une utilisation de l'IA respectueuse de la dignité humaine, des droits, des libertés fondamentales et des valeurs démocratiques.

L'IA est-elle un outil puissant d'émancipation individuelle et collective ou un cheval de Troie dissimulant une nouvelle forme d'asservissement ?

Joanna Bryson Je pense que c'est déjà une sorte de cheval de

Troie. Nous constatons déjà que les gens n'aiment pas les algorithmes de recommandation qui les incitent à passer trop de temps sur les médias sociaux. C'est donc déjà un peu le cas. Mais en même temps, c'est exagéré ! Les batailles que nous devons mener aujourd'hui consistent à faire comprendre aux gens ce qu'ils perdent avec l'affaiblissement de Twitter. Nous ne disposons plus de ce forum où les universitaires et les diplomates avaient ou suivaient de vraies conversations, où la population en général pouvait vérifier si quelqu'un était réel et pouvait être pris au sérieux.

Nous avons perdu tout cela du fait des actions délibérées des propriétaires actuels de Twitter. Pourtant, de nombreuses personnes affirment que le problème des médias sociaux ce sont les algorithmes de recommandation qui vous rendent dépendant et vous renvoient une image négative de vous-même. Il peut être légitime de s'en inquiéter et de s'en protéger. Mais encore une fois, avant que Musk ne l'achète, Twitter avait investi massivement dans un système de modération qui réduisait ces effets. Je ne dis pas que ces effets ne sont pas réels, mais ils étaient atténués par le fait que les gens apprenaient beaucoup. Les utilisateurs de Twitter, mais pas ceux de Facebook, étaient en fait bien informés sur des sujets tels que, par exemple, Covid, ou sur la manière de gérer les décrets gouvernementaux ou ce qui se passait dans le domaine des soins Les utilisateurs de Twitter ont donc tendance à être beaucoup mieux informés.

Nous ne créons pas l'IA pour qu'elle soit bonne ou mauvaise, c'est ce que nous en faisons qui peut être bon ou mauvais. Et parfois, de bons outils peuvent soudainement être retournés et rendus mauvais juste parce que quelqu'un a eu une intuition et vice versa. Il n'y a aucune raison de penser que l'on ne puisse pas faire l'inverse.

Il n'y a rien d'intrinsèque à l'IA qui permette de savoir si elle contribue ou non à la dignité. C'est à nous de décider. C'est nous qui prenons ces décisions, n'est-ce pas ?

Définition
L'IA, par convention, décrit des artefacts non vivants qui ont la capacités de percevoir des contextes d'action, d'agir et d'associer contextes et actions… de manière à agir intelligemment, c'est-à-dire de "faire la bonne chose au bon moment".
(Bryson, & Winfield, 2017 ; Byson, 2019)

Oui, mais il est plus difficile de décrypter la logique de son modèle, alors que, par exemple, il est facile de comprendre les prises de position d'un magazine d'opinion, ou le message d'une publicité : on voit vite où elle veut nous emmener…

Joanna Bryson : Je suis d'accord avec vous pour dire que la technologie numérique peut être utilisée pour créer des systèmes très obscurs, mais en même temps, nous devenons beaucoup plus puissants pour pénétrer et exprimer la complexité que nous ne l'avons jamais fait auparavant. C'est pourquoi les prévisions météorologiques s'améliorent. C'est pourquoi nous comprenons mieux l'univers. Nombreux sont ceux qui affirment à tort que l'IA est nécessairement opaque. Non, ce n'est pas vrai. C'est comme n'importe quel autre sujet sur lequel nous pourrions forcer les gens à être plus clairs. Comment l'ont-ils construite ? Comment l'ont-ils testée ? Quand se sont-ils assurés de pouvoir la diffuser ? Comment ont-ils pris ces décisions ?

Il n'est pas possible de connaître chaque nœud d'un réseau neuronal qui en compte des billions. Mais vous savez, c'est également vrai pour les personnes. Nous avons des billions de connexions dans notre cerveau et nous ne sommes pas obligés de savoir ce que fait chacune d'entre elles ? Mais nous sommes obligés d'être capables de nous réguler les uns les autres. Les gens déterminent ce qui fonctionne et ce qui ne fonctionne pas, puis tout le monde détermine les meilleures pratiques à suivre. Par exemple, à ce stade, tout le monde doit comprendre que l'apprentissage automatique reflète les préjugés de la société sur laquelle les données ont été construites. C'est

pourquoi tout le monde vérifie désormais ce point avant de lancer un produit.

En ce qui concerne le contre-pouvoir, de nouveaux travaux très intéressants sont en cours, comme ces sociétés d'investissement qui possèdent des portefeuilles entiers d'entreprises et qui tentent de garantir une certaine stabilité à l'avenir. Il existe donc de nouveaux types de réglementations et de pressions.

Etablir des règles de transparence et de traçabilité donne à la société le pouvoir de contrôler. Mais est-ce suffisant ?

Joanna Bryson : Bien sûr que non. Je m'efforce à présent de faire en sorte que l'IA soit utilisée pour accroître la valeur des travailleurs afin qu'ils reçoivent des salaires plus élevés, plutôt que de l'utiliser pour rendre les travailleurs plus semblables, ce qui aura pour conséquence de faire baisser les salaires parce qu'ils deviendront plus facilement échangeables. Par exemple, des chercheurs du MIT ont publié un excellent article sur le ChatGPT. Il s'agit d'une étude dans laquelle ils ont examiné différents groupes de personnes qui écrivaient des textes pour gagner leur vie, par exemple dans le domaine de la publicité ou autre. Ils leur ont appris à utiliser ChatGPT et ils ont tous adoré. Ils ont dit que cela les rendait plus rapides. Ils ont constaté que pour les trois quarts d'entre eux, l'utilisation de ChatGPT leur permettait d'écrire plus vite. Lorsque les chercheurs sont revenus quelques semaines plus tard, 70 ou 80 % des personnes utilisaient toujours le logiciel parce qu'elles le trouvaient utile, mais ce n'était pas le cas des meilleurs rédacteurs.

Mais l'inverse est également vrai. Vous savez, que les salaires chinois ont augmenté trop vite. La Chine considère cela comme un problème et le gouvernement veut donc utiliser l'IA pour simplifier les postes et réduire les salaires.

J’aimerais que nous trouvions des moyens d’utiliser cette technologie pour aider tout le monde à être plus productif. Il s’agit davantage d’une expression de ce que nous sommes en tant qu’individus ; pas seulement d’une expression de ce qu’est le travail. Mais les deux approches vont probablement coexister.

L’influence de l’IA repose entre les mains de ceux qui la manient, ce qui souligne notre responsabilité collective dans l’élaboration de sa trajectoire. Le discours sur l’avenir de l’IA doit se poursuivre, encadré par la sensibilisation, la réglementation et un engagement envers son potentiel bénéfique.

**Comment expliquer la façon dont ChatGPT présente ses réponses sur un ton affirmatif voire péremptoire,
sans donner ses sources ou en donner de fausses ?**
Joanna Bryson : L'intérêt du fonctionnement de ChatGPT est qu'il exploite des données. Vous aurez peut-être de la chance si vous cherchez quelque chose qui fait déjà l'objet d'un consensus. Reste qu'il est important de savoir que l'on obtiendra toujours quelque chose qui ressemble à une langue, qu'il y ait ou non un consensus ; parce que c'est comme ça que ChatGPT a été entraîné. Il a été formé à s'exprimer dans un langage compétent et assuré, par exemple le style employé les journalistes. Les machines sont des outils que nous achetons et la plupart des machines avec lesquelles nous travaillons sont des outils de travail conçus par nos entreprises. Quel que soit l'algorithme utilisé, le corpus livré et les réponses présentées pourront être différentes, mais la tonalité restera la même.

Decryptage

Alors que le rôle de l’IA reste ambigu, le professeur Bryson partage une perspective nuancée. Toute la question est de

savoir comment utiliser l'IA pour nous aider. En tant que consommateurs européens, nous devons nous assurer que le droit de la consommation est appliqué afin de prévenir ou d'atténuer le risque de surveillance, de manipulation, de coercition, d'exploitation, d'autocratie…

Idéalement, nous devrions être conscients des potentiels et des limites inhérents à la conception des systèmes intelligents. Cela suppose que l'industrie technologique investisse dans la transparence et donne accès à la logique et aux paramètres de construction des algorithmes, permettant une compréhension globale de leurs tenants et aboutissants. Le libre accès à ces explications est le seul moyen de garder le contrôle sur les résultats – une question de démocratie.

Propos recueillis par Marie-Georges Fayn

Joanna Bryson

est professeur d'éthique et de technologie à la Hertie School de Berlin, en Allemagne. Elle a été l'une des premières expertes membres du Partenariat mondial sur l'IA (GPAI), une initiative internationale visant à soutenir le développement et l'utilisation responsables de l'IA. Elle a été consultée par divers gouvernements, organisations et médias sur des questions liées à l'IA, comme la Commission européenne, le Parlement britannique, l'OCDE, l'ONU, la BBC, le Guardian, etc.

Bibliographie

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, *356*(6334), 183-186.

Bryson, J. J. (2022) One Day, AI Will Seem as Human as Anyone.

What Then? Wired – Ideas, https://www.wired.com/story/lamda-sentience-psychology-ethics-policy/

Bryson, J. J. (2019). The past decade and future of AI's impact on society. *Towards a new enlightenment*, 150-185.

Bryson, J., & Winfield, A. (2017). Standardizing ethical design for artificial intelligence and autonomous systems. *Computer*, *50*(5), 116-119.

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Available at SSRN 4375283*.

Theocharis, Y., Cardenal, A., Jin, S., Aalberg, T., Hopmann, D. N., Strömbäck, J., … & Štětka, V. (2021). Does the platform matter? Social media and COVID-19 conspiracy theory beliefs in 17 countries. *new media & society*, 14614448211045666.