

Note du compilateur [G.P.T.](#)

Encore un texte où on met en valeur l'IA mais où sous-jacent, on voit poindre les problèmes qu'elle pose et les réponses données sont tellement éloignées de ce que les humains sont en réalité que tout cela semble carrément utopique. Quelque soit le règlement, quelque soit la responsabilité que se donnent ses concepteurs, il y aura toujours ce trou par lequel s'infiltra tous ceux dont les intentions ne sont pas aussi «pures» et, en fait et à la fin, l'IA sera utilisée d'une manière qui échappera à tout contrôle et finalement, quoiqu'on fasse, elle se retournera contre nous malgré toutes ces belles et rassurantes paroles...

Ne soyons pas dupes, il faut plus qu'une réglementation ad-hoc pour contenir l'IA, il faudrait que l'humanité toute entière évolue à la hauteur de ce que l'IA peut nous rendre capables mais cela... c'est une autre histoire.

PME Magazine

L'actualité des PME romandes

le 16.05.2023

La grande interview

«Les outils ne sont jamais neutres et toutes les IA ont un objectif sous-jacent»



Afin de mieux répartir les responsabilités en matière d'intelligence artificielle, Johan Rochel, cofondateur du laboratoire d'éthique sur l'innovation Ethix, plaide pour plus de transparence dans la conception et les effets des outils technologiques.

Par William Türler

Lorsqu'il a commencé sa formation en philosophie, Johan Rochel n'aurait jamais imaginé qu'il étudierait également le droit. Aujourd'hui, il fait un pont entre ces deux disciplines en traitant de sujets divers, tels que la migration, la politique internationale ou l'innovation. A travers ce prisme, il nous parle des problèmes éthiques que soulève le développement de l'intelligence artificielle (IA), notamment d'outils tels que ChatGPT.

Elon Musk et d'autres experts ont réclamé récemment dans une lettre ouverte une pause dans le développement de l'intelligence artificielle. Pourquoi cet appel vous laisse-t-il un goût amer ?

J.R. : C'est très bien de parler des risques liés à l'IA, mais il faut le faire de manière intelligente, si j'ose dire. En lisant cette lettre, j'ai eu l'impression qu'on cherchait à nous balader. On nous présente des risques très hypothétiques sur le très long terme, en l'occurrence le fait que les machines puissent un jour échapper au contrôle des humains et se retourner contre eux. Or les dangers sont beaucoup plus présents et actuels. Et bien plus embêtants pour les entreprises qui développent ces outils.

Quels sont ces dangers ?

Il y a deux grandes catégories de risques éthiques. La première concerne la technologie elle-même et les choix qu'implique la conception de ces outils, en termes d'algorithmes, de données, de design de l'interface utilisateur-machine, de durabilité et de contrôle. Les outils ne sont jamais neutres. Toutes les IA ont un objectif sous-jacent. La deuxième concerne l'impact direct sur les relations humaines, par exemple au travail.

Quels garde-fous peut-on établir? Une réglementation est-elle indispensable ?

Cette question nous oblige à clarifier les rapports entre droit et technologies. L'industrie, notamment celle de l'IA, veut faire croire que le droit est toujours en retard sur la technologie. Mais c'est faux. Toutes les innovations technologiques apparaissent dans un cadre juridique. Il n'y a pas de technologie hors du droit. On peut débattre du niveau de précision des règles, mais pas du fait que le droit s'applique. De manière générale, nous devons viser à plus de transparence dans la conception et les effets des outils technologiques. A partir de là, on pourra mieux attribuer les responsabilités.

Que pensez-vous de la décision récente de l'Italie d'interdire ChatGPT ?

Symboliquement, cela me paraît un très bon signal. C'est une décision pertinente, car elle vise des risques actuels liés à la protection des données des utilisateurs italiens. C'est une demande de transparence. Elle se base sur le droit actuel, avec des demandes très concrètes du régulateur. Elle permet de mettre le doigt sur cette boîte noire, car on ne sait pas ce que ChatGPT fait des données. D'autres pays ont d'ailleurs annoncé se joindre à ces demandes de transparence.

On parle beaucoup des risques de pertes d'emploi, dont une partie pourrait être compensée par la création de nouveaux types de métiers, tels que les prompt engineers (littéralement: ingénieurs de saisie en français).

Quel est votre point de vue à ce sujet ?

Je ne pars pas de l'idée que les emplois sont des blocs et que les employés pourraient être simplement remplacés par la technologie. La plupart des collaborateurs dans les entreprises effectuent une multitude d'opérations dans une journée. La question est plutôt de savoir quelles opérations sont impactées par une technologie. Certaines tâches vont disparaître ou se feront plus rapidement. Évidemment, dire cela est moins spectaculaire que de prévoir que la moitié des emplois risquent de disparaître... Et, bien sûr, certains métiers dont l'activité centrale peut être largement réalisée par une machine, comme traducteur, sont plus touchés que d'autres. Il y aura aussi des impacts cumulés, qui vont modifier en profondeur le travail d'un grand nombre de personnes. Les pertes d'emploi qui en résultent doivent être gérées comme une question de politique sociale, liée à la transition numérique.

Qui doit payer pour cette politique sociale ?

Je vois trois possibilités de répartir la responsabilité: les employeurs, les employés ou la collectivité par l'État. On peut partir du principe qu'une entreprise qui obtient des gains de productivité en

remplaçant des humains par des machines pourrait, à un moment ou un autre, porter une part de responsabilité. Par exemple, en finançant la transition professionnelle de ses employés. Selon les domaines d'activité, on pourrait également considérer que les employés doivent se former de leur propre initiative. Pour certaines professions où il serait exagéré de mettre ce poids sur l'employé, l'État devrait fonctionner comme un filet à solutions, dans ou en dehors du marché du travail.

Les outils d'IA posent de gros défis en matière de responsabilités.

Prenons l'idée que tout cela est une chaîne, qui commence par la recherche et va jusqu'à l'utilisateur final d'un outil. Comment répartir les responsabilités sur l'ensemble de la chaîne? Faut-il arrêter de faire de la recherche sur l'IA? C'est un discours que l'on a déjà entendu par le passé avec la génétique ou l'atome. Au-delà de la recherche scientifique, c'est certainement le producteur d'un outil qui doit porter la plus grande part de responsabilité. Il doit être capable d'assurer la sécurité et la fiabilité de son outil. En Suisse, la loi sur la responsabilité du fait des produits règle ce genre de conflits entre le producteur et l'utilisateur dans le cas d'un dommage provoqué par un article défectueux.

Mais l'IA a ceci de particulier qu'elle apprend en fonction de l'utilisation des consommateurs.

En effet, dans ce cas, le produit (un robot ménager, une tondeuse à gazon intelligente, une Tesla) évolue selon le comportement de l'utilisateur. Par conséquent, la responsabilité de ce dernier augmente en cas d'utilisation non conforme. Au final, on se trouve devant un vrai choix de politique d'innovation. C'est au législateur de définir comment il veut répartir cette responsabilité par défaut. S'il met tout sur l'industrie, cela aura un effet de frein à l'innovation. S'il met tout sur le consommateur, cela semble injuste, car l'utilisateur est souvent le maillon faible de la chaîne. On ne peut pas lui demander de devenir un expert. Il faut trouver un juste milieu. La régulation sur l'IA à venir de l'UE pose des jalons intéressants: on distingue différentes catégories de risques et on adapte à ces risques les exigences que l'industrie doit remplir avant la mise sur le marché.

On en revient à la question de la transparence.

Si l'on veut utiliser ces outils dans des activités sensibles, comme la justice, on doit impérativement éviter le syndrome de la boîte noire. On doit avoir des informations qui nous permettent de comprendre le système et, le cas échéant, de demander des comptes. La question consiste en gros à comprendre pourquoi un outil, par exemple ChatGPT, nous donne telle réponse à telle question.

Comment s'y prendre ?

L'industrie freine des quatre fers, expliquant avoir investi des millions de dollars afin de développer des algorithmes qui représentent son 'core' business. Il faut entendre ce point de vue. Néanmoins, l'industrie numérique n'est pas unique. Pourquoi ne pas faire comme dans la pharma et déléguer cette compétence de contrôle à des entités soumises à un devoir de confidentialité? Aujourd'hui, l'industrie de l'IA teste ses outils en direct sur le grand public. Nous devons basculer dans une logique d'analyse des risques avant l'arrivée sur le marché. Personne n'aurait l'idée de faire des essais de médicaments directement sur les gens.

Les accusations de vol de données au profit du gouvernement chinois portées par les Américains contre TikTok vous semblent-elles fondées? Ne s'agit-il pas plutôt d'un prétexte pour contrer un concurrent qui siphonne le public cible et les annonceurs des réseaux sociaux américains, comme Facebook, YouTube ou Instagram ?

Il est évident que cette lutte économique prend une tournure plus géopolitique. Les réseaux sociaux sont une source gigantesque pour collecter des données qui permettent ensuite d'entraîner des IA. On récolte en quelque sorte un trésor de guerre qui permet de créer des technologies. C'est un enjeu qui

dépasse TikTok et son impact sur les adolescents, même si c'est également une question très importante. Une fois encore, les IA et les réseaux sociaux ont tous un objectif. Il peut être commercial ou plus politique.

On accuse aussi TikTok de proposer des versions différentes de son algorithme en fonction des pays. Et de favoriser un contenu plus éducatif pour les enfants chinois.

Cherchons l'humain et ses intérêts derrière tous ces outils: un réseau social va tout faire pour garder captifs ses usagers le plus longtemps possible. Il va donc choisir le type de contenu qu'il va vous proposer en fonction de vos centres d'intérêt spécifiques. Dès lors, on peut commencer à exploiter les faiblesses du cerveau humain, à travers les biais cognitifs, par exemple notre envie d'être conforté dans nos opinions. Et nous voici au cœur d'une «bulle de filtres» avec tous ceux qui pensent comme nous. Le réseau social veut simplement gagner de l'argent, mais son fonctionnement et sa taille en viennent à représenter un danger pour la démocratie.

Selon vous, quels espaces ne devraient jamais être robotisés ?

Je crois dans un droit aux contacts humains, notamment dans des domaines essentiels comme la justice, l'éducation ou les soins. Par exemple, dans une procédure juridique, je devrais avoir le droit d'obtenir un regard humain sur mon dossier. En matière de soins, ce contact humain va plus loin qu'un droit de procédure. Il y a une logique de reconnaissance, d'expérience partagée de la souffrance ou de la finitude humaine qu'un robot, même le plus perfectionné, ne peut pas avoir.

Les entreprises suisses utilisent-elles les outils numériques et l'IA de manière responsable ?

Les performances des grandes sociétés sont décevantes. Selon l'étude Ethos sur la responsabilité numérique, les branches qui se comportent le mieux sont les assurances et les banques. Parmi les PME, c'est un peu la jungle. On avance néanmoins en matière de protection des données. Il devient de plus en plus difficile aujourd'hui de dire qu'on n'a jamais entendu parler de cette problématique. A mesure que les amendes deviendront plus salées, cela va continuer à s'améliorer. En revanche, en ce qui concerne les outils d'analyse de données, par exemple dans les ressources humaines, nous avons encore du travail de sensibilisation à faire. Il y a le chapitre lié aux discriminations et aux biais des algorithmes, mais également la tentation de surveiller beaucoup plus étroitement ses employés en scrutant tous leurs faits et gestes.

Au-delà des discriminations, l'un des problèmes des IA est qu'elles tendent vers une certaine standardisation de la pensée.

On discute souvent de la diversité des équipes de conception. Il y a sans doute des éléments à améliorer, mais il ne faut pas partir de l'idée que ces personnes sont malintentionnées. La plupart veulent simplement faire un bon outil et le vendre. Les questions les plus sensibles me paraissent plus profondes. Les IA se nourrissent de données qui sont censées refléter la société. Quelle partie de notre vie devrait échapper à cette mise en données? Parmi toutes les parties documentées, y a-t-il des catégories que nous ne voulons plus utiliser? Prenez le cas des assurances voiture. En analysant les modes de conduite et les accidents, on peut trouver des milliers de liens statistiques. Mais au-delà des statistiques, voulons-nous des prix d'assurance différents selon le sexe, la nationalité, la couleur de cheveux? Est-ce que le prix devient toujours plus juste, ou sommes-nous en train de détruire l'idée de solidarité dans l'assurance? Les outils d'IA nous forcent à poser ces questions très fondamentales et à clarifier ce qui devrait échapper à cette logique de mise en données et d'optimisation.

En tant qu'éthicien, n'avez-vous pas peur de freiner l'innovation et une certaine forme de progrès liée aux avancées technologiques ?

L'éthique fait partie du développement de n'importe quel produit. Il y a forcément des buts, des valeurs, des choix. On peut les rendre explicites et être plus ou moins à l'aise avec les valeurs que l'on a placées dans l'outil. Une chose est sûre: les gens demandent de plus en plus souvent des comptes. Une certaine cohérence se révélera donc toujours plus payante pour les producteurs.

L'IA va-t-elle un jour s'affranchir des humains ?

Dans les films de science-fiction, le moment où les machines deviennent indépendantes relève toujours un peu de la magie. Conceptuellement, je ne vois pas comment cela pourrait être possible. Une voiture autonome peut choisir le chemin le plus optimal pour aller de A à B, mais elle ne peut pas décider de s'arrêter pour regarder le paysage. L'humain peut réviser ses buts. Un problème pourrait survenir si l'on donnait un jour un objectif tellement large à une machine super-puissante qu'elle pourrait l'interpréter à sa manière. Par exemple: «Rends le monde meilleur.» Pour elle, cela pourrait passer par l'élimination de l'humanité, qui saccage la planète. Mais d'un point de vue philosophique, cela ne veut pas dire qu'elle est devenue autonome. Cela veut juste dire qu'on lui a laissé trop d'espace de choix et on revient à la question de la responsabilité des concepteurs. Les robots de guerre et leur «autonomie» illustrent bien ce défi.

«Aujourd'hui, l'industrie de l'IA teste ses outils en direct sur le grand public. Nous devons basculer dans une logique d'analyse des risques avant l'arrivée sur le marché.»

Bio express : Johan Rochel - cofondateur du laboratoire d'éthique sur l'innovation Ethix

1994 Programmation en langage C d'un agent conversationnel répondant à trois questions préprogrammées.

2014 et 2016 Naissance de ses deux enfants.

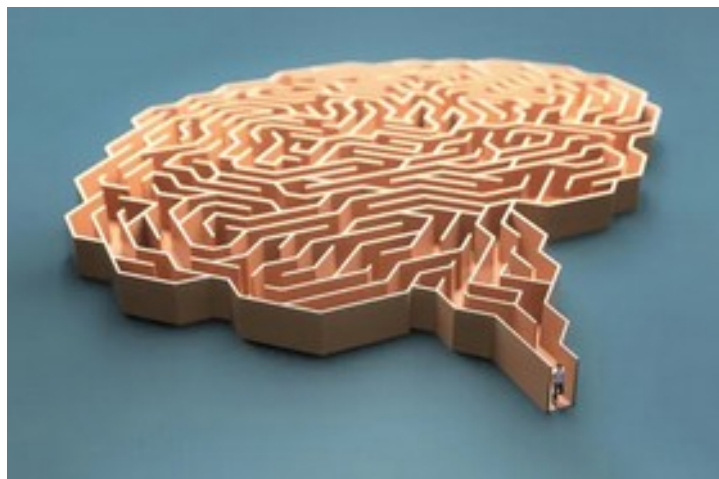
2018 Il cofonde le laboratoire d'éthique de l'innovation Ethix avec Jean-Daniel Strub.

Il a étudié la philosophie et le droit. Aujourd'hui il fait un pont entre ces deux disciplines en traitant de sujets tels que la migration, la politique internationale ou l'innovation.

Lire plus :

Digital Shapers 2022 - AI Masters : [Digital Shapers-AI Masters](#)

Ethix - laboratoire d'éthique de l'innovation : [Ethix.ch](#)



Lire plus : «Créer l'intelligence elle-même»

Numérique à l'école - [Une inquiétante réforme](#) - 24heures du mardi 04 avril 2023

Qu'en est-il vraiment de l'IA ? - [Pour ou contre](#) - 24heures, Le Courrier, RTS, CLC

ChatGPT débarque - [sur le campus](#) - 24heures du mercredi 03 mai 2023

Faut-il craindre l'IA - [Ses risques](#) - 24heures 31.03.2023

Après le feu, l'électricité et internet - [la révolution de l'IA](#) - 20 Minutes du 20 mars 2023

Derrière l'IA - [le travail précaire au Sud](#) - Le Courrier du 01.05.2023

L'intelligence artificielle est - [omniprésente dans nos vies](#) - Le Temps du 18 janvier 2018

Une technologie - [met en garde contre elle-même](#) - [PME Magazine](#) du 28.04.2023

Fausses photos, vidéos - [une terrible ère du doute commence](#) - Le Temps 6 avril 2023

La mort du progrès - [nous laisse vides et angoissés](#) - Le Temps du jeudi 27 Décembre 2019

L'IA favorise la - [standardisation et non la diversité](#) - PME Management du 15 mars 2023

L'IA qui n'est jamais neutre - [a un objectif sous-jacent](#) - PME Magazine du 16 mai 2023

Une grande [révolution mondiale](#) se prépare - de Ted Kaczynski «The Coming Revolution»

L'IA figure un changement [de contrat social](#) à bas «bruit»