

Intelligence artificielle

Par Ludivine Ponciau et Nidal Taibi

ChatGPT, baratineur sans cœur

Toujours plus proches de l'humain,
les chatbots seront-ils un jour
capables de simuler les émotions,
les sentiments ?



« Que signifierait pour les humains de vivre dans un monde où un grand pourcentage d'histoires, de mélodies, d'images, de lois, de politiques et d'outils sont façonnés par une intelligence non humaine, qui sait exploiter avec une efficacité surhumaine les faiblesses, les préjugés et les dépendances de l'esprit humain – tout en sachant nouer des relations intimes avec les êtres humains? Dans des jeux comme les échecs, aucun humain ne peut espérer battre un ordinateur. Que se passera-t-il lorsque la même chose se produira dans le domaine de l'art, de la politique ou de la religion? » En partageant cette réflexion dans les colonnes du *New York Times*, le 23 mars, l'historien et philosophe Yuval Noah Harari, auteur de plusieurs best-sellers dont *Sapiens, une brève histoire de l'humanité* (Albin Michel, 2015), met le doigt sur l'une des plus grandes préoccupations des scientifiques à propos des dangers associés à l'intelligence artificielle, à savoir sa capacité à influencer nos décisions, à orienter nos choix, à réfléchir à notre place.

Une mise en garde cosignée par Tristan Harris et Aza Raskin, les fondateurs du Center for Humane Technology, dont la vocation est de faire pression sur les entreprises technologiques pour qu'elles développent des logiciels « éthiques ». Tous trois figurent également sur la liste des centaines d'experts qui réclament d'urgence une pause de six mois dans le développement de l'intelligence artificielle. Un appel au moratoire qui cache probablement d'autres enjeux moins moraux, liés à la concurrence féroce que se livrent les start-up de la Silicon Valley et qui aimeraient sans doute rattraper leur retard sur OpenAI, la société qui a développé ChatGPT, mais qui n'en reste pas moins percutant. Les premiers remous ne se sont d'ailleurs pas fait attendre : deux jours après la publication du plaidoyer, l'Italie a bloqué le robot conversationnel, invoquant une violation de la législation sur les données personnelles et l'absence de contrôle de son utilisation par les mineurs. D'autres pays pourraient lui emboîter le pas.

Agent perturbateur

Une technologie qui fascine autant qu'elle inquiète. Son avènement rebat les cartes dans toute une série de secteurs : emploi, enseignement, médias, marketing... Mais pourrait-elle perturber notre rapport à l'humain, nos interactions sociales? ChatGPT pourrait-il devenir un jour notre mentor, notre psy, notre meilleur ami?

Professeur à la faculté de philosophie de Laval, au Québec, Jocelyn Maclure appelle dans *La Presse* à dépasser l'emballement et le battage médiatique autour de chaque percée de l'intelligence artificielle pour développer une appréhension juste et sobre de ses capacités et de son influence. « Même en écartant les scénarios fantaisistes souvent évoqués par ceux qui ont passé beaucoup de temps à écrire du code et à consommer de la science-fiction, on doit reconnaître que des technologies d'IA comme ChatGPT ou des générateurs d'images comme DALL-E ou Midjourney sont des technologies perturbatrices. Elles modifient substantiellement des pratiques sociales bien établies. »

Les prouesses de ces « perroquets stochastiques », comme les qualifie le chercheur canadien, sont en effet basées sur les généralisations statistiques que leur permet leur vertigineuse puissance de calcul. Mais, rappelle-t-il, ces baratineurs ont une fâcheuse tendance à inventer de toutes pièces des réponses en s'appuyant sur les corrélations qu'elles repèrent dans les données.

Observer ces émergences et les analyser permet de comprendre que ces systèmes ne présentent pas de rationalité telle que nous l'entendons, expose dans les médias français la professeure d'informatique à l'université Paris-Sorbonne et autrice de *Les Robots émotionnels* (L'Observatoire, 2020), Laurence Devillers. « Le cerveau humain, c'est avant tout une succession de cerveaux humains. [...] Nous sommes là parce que nos parents ont été là. On a un ADN, on a de l'inné et on va acquérir d'autres choses... Ce que fait notre cerveau dans l'inconscient est également énorme.

« Contrairement à nos amis, le robot nous évite la contradiction et la frustration. »

« Il y a un énorme faisceau d'informations qui explique qui je suis. La machine, elle, ne dispose pas de ça. »



Quand je raisonne, je raisonne à travers ma culture, mon éducation, ce que mes parents étaient, ce que la société nous renvoie. Il y a un énorme faisceau d'informations qui explique qui je suis. La machine, elle, ne dispose pas de ça. »

« Il va falloir s'habituer à interagir avec de nouvelles entités mais l'intelligence artificielle a cette particularité de ne pas avoir de devoirs moraux, alors que nous, nous en avons envers les humains et les animaux, pointe le philosophe et chercheur à l'université de Montréal, spécialiste de l'éthique de l'intelligence artificielle et auteur de *Faire la morale aux robots* (Flammarion, 2021), Martin Gibert. Il serait donc difficile de tenir pour responsable un système purement mécanique qui n'est pas un agent moral, qui n'a pas d'intention ni de conscience de nos agissements. Bien qu'on puisse néanmoins admettre qu'il est capable de nous stimuler. »

« Si on s'en réfère au test de Turing permettant de mesurer la capacité d'une intelligence artificielle à imiter une conversation humaine, poursuit le chercheur, on peut considérer qu'une machine est intelligente lorsqu'une personne qui discute avec elle depuis un certain temps ne s'est pas aperçue qu'il s'agit d'un robot. L'objectif, à terme, d'entreprises comme OpenAI, est de développer une forme d'intelligence artificielle générale (*NDLR : dotée de capacités cognitives humaines, voire selon certaines interprétations d'une conscience et de sentiments*). En ce sens, la technologie de ChatGPT est un pas important dans cette direction. Au point qu'il n'est pas exclu qu'une machine puisse un jour conseiller un humain sur la gestion de ses émotions, et que dans l'hypothèse où elle nous connaîtrait assez bien, qu'elle puisse fournir une réponse aussi cohérente qu'une personne à qui nous nous serions confié. »

Le philosophe québécois a eu vent de l'histoire de ce jeune Belge, devenu très anxieux, qui s'est donné la mort après avoir conversé intensivement avec Eliza, un chatbot utilisant la technologie de ...

... ChatGPT. Sans se prononcer sur ce cas spécifique, il valide la nécessité de mieux encadrer les futurs développements, notamment en permettant à l'utilisateur de pouvoir modifier les paramètres du système. Mais cette crainte de voir survenir d'autres « histoires malheureuses » ne doit pas occulter l'aide que l'IA pourrait apporter dans la gestion des émotions. « D'une certaine manière, elle pourrait aider certaines personnes à aller mieux. On peut même imaginer qu'elle puisse se montrer utile dans une intervention de première ligne en santé mentale. » Une start-up américaine a osé en lançant Koko, une plateforme en ligne aidant les personnes souffrant de problèmes de santé mentale. Ses bénévoles ont été temporairement remplacés par ChatGPT, sans que les patients ne soient mis au courant. Sous le feu des critiques, la société a finalement fermé la plateforme.

De Terminator à calinator

Pour Olivier Duris, psychologue français spécialisé dans l'usage des écrans et le numérique, un tel projet ne peut qu'être voué à l'échec. « J'ai moi-même fait le test et je ne pense pas que l'IA puisse venir en aide à des personnes souffrant de troubles. Il ne s'agit que de simulation. L'écoute, elle, n'est pas un dispositif protocolaire. L'important, c'est ce qui se rejoue en présence du thérapeute, ce qui est ressenti. Cela, une machine ne pourra jamais le remplacer. »

Ce sur quoi Olivier Duris et Martin Gibert sont par contre d'accord, c'est le risque que les chatbots nous rendent moins réceptifs à la contradiction. « L'homme a toujours eu une propension à l'anthropomorphisme en prêtant des sentiments aux objets. Les robots accentueront cette tendance anthropomorphique en créant l'illusion qu'ils se souviennent de nous, qu'ils peuvent éprouver des émotions. Ils fourniront toujours des réponses qui vont dans notre sens », décrit le premier. « Il répond à nos attentes. Trop, même. Contrairement



« L'important, c'est ce qui est ressenti. Cela, une machine ne pourra jamais le remplacer. »

à nos amis, notre entourage, il nous évite la contradiction et la frustration », appuie le second.

En 2020 déjà, le psychiatre et psychanalyste français et membre de l'Académie des technologies, Serge Tisseron, mettait en garde contre le « calinator », ce robot qui, à l'inverse de Terminator, a perdu toute agressivité pour s'adapter aux besoins de chacun et flatter son ego. Une confusion accrue par l'usage de la voix et des avatars. Un exemple parmi d'autres : début mars, une start-up basée à Tel-Aviv a annoncé le lancement d'une plateforme permettant des conversations « naturelles » et « en temps réel » avec un « humain numérique », la qualifiant de technologie complémentaire aux modèles de langage qui alimentent ChatGPT.

Dans une société très polarisée, peut-on imaginer que ces conversations plaisantes avec les logiciels conversationnels nous poussent à ne plus rechercher que la validation, à être confortés dans nos positions ? La comparaison avec le phénomène des hikikomoris est tentante... Apparue dans les années 1990 au Japon, ce syndrome se caractérise par un isolement extrême de jeunes qui ne trouvent plus de satisfaction dans les interactions sociales, la plupart d'entre eux basculant dans la cyberdépendance. « L'animisme est sans doute une attitude plus répandue au Japon, où l'on attache beaucoup d'importance aux machines. Je ne pense pas que cela puisse se produire chez nous. Par ailleurs, il n'est pas impossible que, paradoxalement, le fait d'être fréquemment en contact avec des robots nous pousse à côtoyer davantage les humains et à mieux apprécier ce qu'ils ont de si particulier », formule Martin Gibert.

Comment garder ses distances avec celles que l'on appelle Eliza, Amelia, Sunny, Replika... et qui semblent si bien nous comprendre ? Olivier Duris propose une charte qui fixerait les limites dans l'utilisation des données, qui obligerait à plus de transparence et qui promouvrait le développement de programmes de sociabilisation plutôt que la cyberdépendance. Et éduquer. Surtout éduquer. ● L.P.