

Il n'y a jamais eu un tel fossé entre ce qu'un logiciel est capable de produire et nos capacités cognitives de le comprendre. Les autorités devraient s'emparer de l'IA pour la mettre au service du bien public.

ENTRETIEN

PHILIPPE LALOUX

ChatGPT est un élève médiocre avance le philosophe français, Raphaël Enthoven, après l'avoir sèchement battu à l'épreuve du Bac Philo. Pas si vite, a tenu à rétorquer Hugues Bersini, professeur d'informatique et d'intelligence artificielle à l'ULB. Les capacités des réseaux neuronaux, comme ChatGPT, sont époustouflantes et échappent à notre compréhension.

L'IA sera donc capable de « comprendre » le monde ?

Toute l'évolution de l'IA de ces 20 dernières années est très empirique. Les réseaux de neurones, par exemple, qui sont à la base de ChatGPT, on ne les comprend pas de manière théorique. C'est la raison pour laquelle je pense que Raphaël Enthoven passe à côté de ce qui est essentiel dans cette évolution. Il y a plein de choses que l'on peut questionner philosophiquement, mais ici, on vient avec des critiques que l'on aurait pu adresser il y a 40 ou 50 ans. Oui, l'IA n'a pas de conscience, ne ressent rien, ne comprend pas le contexte... Mais il s'est passé quelque chose avec ChatGPT qui devrait interroger le philosophe. Pourquoi ? Parce que l'on a court-circuité tout un courant de la linguistique théorique (portée par des chercheurs comme Chomsky) qui pensait qu'il fallait comprendre la science du langage pour pouvoir faire des systèmes qui produisent du texte. C'est le cas pour le correcteur orthographique de Word.

Nous, les informaticiens, les linguistes, n'avons jamais réussi à intégrer la compréhension du monde dans les logiciels de production de texte. Parce qu'en général ces logiciels se limitaient à la syntaxe, ce qui n'est pas de la compréhension, ni de la sémantique. Ce qui est extraordinaire avec ChatGPT, c'est que l'on amène un réseau de neurones à prédire le prochain mot. Pour y arriver, ils ont donc dû développer une véritable conceptualisation, je dis même une « compréhension ». En mettant des couches et des couches dans ces réseaux de neurones, ces réseaux apprennent beaucoup plus que les mots. Donc quand le philosophe dit « ChatGPT ne comprend rien », je dis tout le contraire.

La preuve avec Shakespeare ?

J'ai fait des exercices avec des étudiants en travaillant sur des textes de Shakespeare. On demande à des réseaux de neurones de prédire non pas le prochain mot, mais le caractère suivant. Donc, on travaille caractère par caractère et on se rend compte que si on met une ou deux couches, on ne produit même pas de l'anglais. Si on met quatre ou cinq couches, ça devient de l'anglais. Si on met 100 couches, cela devient du Shakespeare. Donc, si vous prenez les 10.000 caractères qui précèdent, en fait, pour prédire le prochain, vous devez comprendre plein de trucs. Et on ne sait pas comment ce réseau de neurones comprend. Mais c'est sûr qu'il a une conceptualisation de Shakespeare. Il se met à penser comme Shakespeare. C'est la première fois dans l'histoire des technologies qu'il y a un tel fossé entre ce que ce logiciel est capable de produire et nos capacités cognitives de comprendre comment ça marche.

Raphaël Enthoven qualifiait ChatGPT d'élève médiocre. Vous l'avez aussi mis à l'épreuve. Et le résultat est surprenant... J'ai pu apprécier les capacités de réflexion mathématique et logique étonnantes de GPT4 lorsque je lui ai soumis tous mes examens de Solvay de ces 30 dernières années. J'ai repris les énoncés tels quels. ChatGPT a répondu parfaitement. En informatique, ChatGPT programme de manière extraordinaire. On l'inscrit dans les compétitions de mathématique où il excelle. Donc, dire qu'il ne comprend pas, qu'il ne raisonne pas, ce n'est pas possible, tellement il est capable de résoudre des problèmes d'une rare complexité. Quand on dit qu'il y a une part irréductible de l'humain que la machine ne pourra jamais reproduire, on se demande laquelle.

Quelle serait donc encore la singularité de l'humain ?

J'étais avec un philosophe devant une œuvre de Magritte, *L'empire des lumières*. A juste titre, il me dit qu'une IA n'aurait jamais pu produire ça. Alors, bien sûr, ce que fait l'IA en matière de plagiat est fascinant. Mais je pense qu'une IA peut apprendre à être créative. Elle peut se rendre compte qu'il y a eu des sauts qualitatifs entre certaines pé-

riodes, que de nouveaux langages ont été inventés. Je suis parfaitement conscient que l'IA n'est pas consciente. Mais là n'est plus le problème. Il se passe malgré tout quelque chose dans ces systèmes qui montre qu'ils ont compris ce que c'est, en quelque sorte, de travailler avec l'émotion. Ils ont appris, notamment, grâce au fait qu'on les a entraînés en leur montrant des tableaux accompagnés de textes les décrivant. Ils sont donc capables de comprendre ce qui fait qu'il y a eu une révolution artistique, que l'on a transgressé des codes... Si vous l'avez nourrie avec toute l'histoire de la peinture et que, avec ce niveau de conceptualisation, elle a compris ce qui fait qu'il y a une révolution en art, elle pourrait être l'acteur de la prochaine révolution. Et il n'y a pas de raison que ça ne puisse pas se produire.

Personne ne s'attendait à ce que l'IA fasse aussi bien. Ses concepteurs sont tellement bluffés qu'ils ont avancé cette équation : plus de puissance, plus de paramètres, plus de mots. On va l'amener encore plus loin dans cette course effrénée à la complexification, parce que la complexification va donner encore plus d'intelligence. Et moins on la comprendra.

« L'IA a sa propre compréhension du monde »



Hugues Bersini est professeur d'informatique et d'intelligence artificielle à l'ULB. © DR.

Hugues Bersini

Membre de l'Académie royale de Belgique, Hugues Bersini enseigne l'informatique à Solvay et à l'ULB, dont il dirige le laboratoire d'IA. Il est le co-créateur de l'Institut Fari, installé à BeCentral, dédié à la réalisation d'algorithmes pour les biens publics. Il est l'auteur de nombreux ouvrages, dont *Algocratie* (Ed. DeBoeck). P.H.L.

Alors, finalement, faut-il avoir peur de l'IA. Ou pas ?

Même les trois chercheurs, lauréats du prix Turing, à l'origine des progrès sur l'IA, sont en discordance sur le risque existentiel de l'IA. J'aurais tendance à dire que l'on peut toujours enlever la prise. Oui, sauf qu'il y a de l'IA et de l'informatique partout et qu'on ne sait plus vraiment la débrancher. Je suis incapable d'avoir une réponse définitive là-dessus, mais, pour ma part, ma philosophie, c'est que l'on peut réellement mettre l'IA au service de l'humanité. On peut, par exemple, vraiment améliorer la mobilité (avec des solutions de covoiturage, d'intermodalité...), la consommation énergétique... Mais on a souvent une vision négative du numérique, alors que moi j'essaie justement de promouvoir une vision positive. Des tas de problèmes pourraient vraiment se résoudre grâce à l'IA. Mais ça effraie.

On a souvent une vision négative du numérique, alors que moi j'essaie justement de promouvoir une vision positive. Des tas de problèmes pourraient vraiment se résoudre grâce à l'IA. Mais ça effraie

”

C'est dans cet esprit que vous avez cofondé l'Institut Fari ?

Je m'étais un tout petit peu lassé de ces discours de l'IA, positifs ou négatifs, de fin de monde ou pas. La Commission se mettait à réguler à tous crins. Il y avait aussi, ici à l'ULB, tout un courant de pensée très anti gouvernance algorithmique. Je voulais prendre un peu le contre-pied en disant que l'on n'avait jamais vraiment mis de l'IA au service du bien public. Le covid a été un moment clé, car on s'est rendu compte qu'il y avait eu une invasion algorithmique à travers le mobile, avec le tracing, les pass vaccinaux, les prises de rendez-vous pour les vaccins, pour les tests... Tout était centralisé sous forme informatique et je trouvais que ce n'était pas bien fait. Il y avait moyen de faire mieux. Fari a été créé dans cet esprit. La première réalisation qu'on a voulu mettre en place, c'était la campagne de vaccination avec la Cocom. Cinq de mes chercheurs ont bossé jour et nuit pour créer une application.

Pourquoi cela a-t-il coïncé ?

Je me suis rendu compte que les administrations publiques ne sont pas prêtes à intégrer l'IA. Il y a beaucoup de conservatismes, de paranoïa. C'était trop compliqué pour les autorités qui ont préféré rester avec leurs tableaux Excel. Je les ai vus, ils étaient d'une complexité effrayante. Au final, c'était du gros bricolage. Il y a un monde entre ce qu'ils ont l'habitude de faire et des solutions qui leur permettent de se simplifier la vie. C'est d'autant plus invraisemblable que nous les faisons gratuitement.

Il y a souvent ce réflexe très ancré de déléguer au privé. On met un argent fou dans les boîtes de consultance, sans toujours comprendre ce qu'elles font. Alors parfois, elles font des choses de qualité, mais pas du tout dans la vision que moi j'ai de la démocratie. Il y a un risque réel de fracture numérique puisque le citoyen n'est même pas au courant. Avec l'Institut Fari, on tente justement de remettre les universités dans ces partenariats et de changer la manière de travailler. Lorsque l'on fait des algorithmes d'IA citoyens, il est important d'avoir son assentiment, de l'impliquer, de faire de l'ergonomie citoyenne (donc des choses qu'ils peuvent utiliser), pour légitimer ces systèmes. J'essaie de montrer qu'il y a des partenaires académiques qu'il ne faut pas court-circuiter. L'Institut Fari est là pour sensibiliser un peu nos administrations en disant qu'il y a une manière très positive de faire de l'IA.