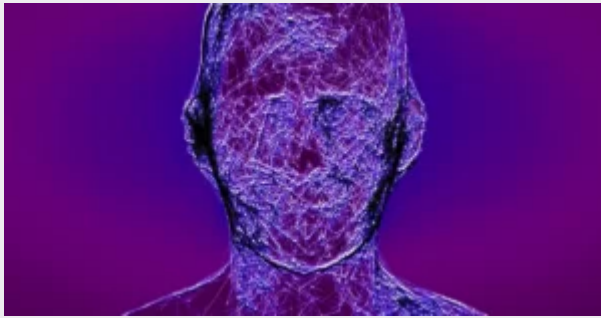


Un neuroscientifique avertit que la génération actuelle d'IA est sociopathe



[Source : aubedigitale.com]

IA sociopathe

Sans conscience, prévient le neuroscientifique de Princeton Michael Graziano dans un nouvel essai publié par le Wall Street Journal, les chatbots alimentés par l'intelligence artificielle sont condamnés à être de dangereux sociopathes qui pourraient représenter un réel danger pour les êtres humains.

Avec l'essor des chatbots comme ChatGPT, des systèmes puissants capables d'imiter l'esprit humain à un degré impressionnant, les outils d'IA sont devenus plus accessibles que jamais. Mais ces algorithmes racontent avec désinvolture tout ce qui leur convient. Pour les aligner sur nos valeurs, Graziano pense qu'ils auront besoin de conscience.

« La conscience fait partie de la boîte à outils que l'évolution nous a donnée pour faire de nous une espèce empathique et prosociale », écrit Graziano. « Sans elle, nous serions nécessairement des sociopathes, car nous n'aurions pas les outils nécessaires à un comportement prosocial. »

La machine à empathie

Bien sûr, ChatGPT n'est pas sur le point de bondir hors de l'écran et d'assassiner quelqu'un. Mais donner à l'intelligence artificielle de plus en plus d'agence pourrait avoir des conséquences très réelles dont nous devrions nous méfier dans un avenir pas si lointain.

Pour les rendre plus dociles, selon Graziano, nous devrions leur permettre de réaliser que le monde est rempli d'autres esprits que le leur.

Il y a cependant un problème : nous ne disposons pas d'un moyen efficace de savoir si une IA est consciente ou non. En fait, d'un point de vue philosophique, il est même difficile de déterminer avec certitude si les autres personnes sont conscientes.

« Si nous voulons savoir si un ordinateur est conscient, nous devons vérifier si l'ordinateur comprend comment les esprits conscients interagissent », affirme M. Graziano. « En d'autres termes, nous avons besoin d'un test de

Turing inversé : Voyons si l'ordinateur peut dire s'il parle à un humain ou à un autre ordinateur. »

Si nous ne parvenons pas à résoudre ces questions délicates, il craint que nous soyons confrontés à de sombres conséquences.

« Une machine sociopathe capable de prendre des décisions conséquentes serait puissamment dangereuse », écrit-il. « Pour l'instant, les chatbots sont encore limités dans leurs capacités ; ce sont essentiellement des jouets. Mais si nous ne réfléchissons pas plus profondément à la conscience des machines, dans un an ou cinq ans, nous pourrions être confrontés à une crise. »

Traduction de Futurism par Aube Digitale