

FUTURA

Des IA pourraient bientôt se répliquer et survivre seule dans la nature !

Podcast écrit et lu par Adèle Ndjaki

[Générique d'intro, une musique énergique et vitaminée.]

Des IA qui ont la faculté de survivre en toute autonomie, c'est l'actu de la semaine dans Vitamine Tech !

[Fin du générique.]

Globalement, quels sentiments éprouvons-nous vis-à-vis de l'intelligence artificielle ? De l'émerveillement ? Face aux prouesses technologiques qu'elles arrivent à réaliser, les IA ont effectivement réussi à susciter une certaine fascination chez l'homme. Mais pour certains d'entre nous, ces nouvelles technologies peuvent inspirer des sentiments plus craintifs : la peur de faire prochainement face à un outil qui surpasserait l'homme à tous les niveaux, et qui finirait par devenir incontrôlable, un peu comme ce que nous avons l'habitude de voir dans les films de science-fiction. Cependant, si ces éventualité étaient généralement balayées d'un revers de main il y a quelques années, aujourd'hui certaines pointures du domaine croient dur comme fer que cela pourrait bientôt devenir une réalité. C'est notamment le cas de Dario Amodei, le PDG d'Anthropic, une société à l'origine de la grande série de modèles de langage Claude AI qui imagine des intelligences artificielles s'auto-répliquer et se développer de façon débridée dès l'année prochaine !

[Une musique électronique calme.]

C'est une déclaration choc ! Dans une interview consacrée au New York Times, Dario Amodei, cofondateur de l'une des plus grosses entreprises d'intelligence artificielle, Anthropic, a fait savoir que les IA pourraient très bientôt s'autodupliquer pour survivre dans la nature. Le chercheur italo-américain pense en effet que les systèmes imaginés dans les œuvres de science-fiction devraient finir par réellement exister d'ici l'année prochaine, plus précisément entre 2025 et 2028. De nouveaux modèles d'intelligences artificielles qui seraient dotés d'une force de persuasion extrêmement poussée. Mais est-ce vraiment possible ? Aussi rapidement ? L'IA n'est utilisée par le grand public que depuis quelques années et a encore du mal à réaliser à la perfection certaines tâches, comme l'élaboration d'images en autre. Cependant, pour reprendre ce même exemple, nous avons pu constater dans le dernier épisode de Vitamine Tech, que les clichés générés par l'IA sont de plus en plus convaincants et que l'intelligence artificielle devient globalement de plus en plus performante, et ce, à une vitesse fulgurante. Un constat également fait par le PDG d'Anthropic qui estime que l'IA devrait bientôt représenter un niveau de dangerosité encore

jamais atteint jusqu'à présent. Afin de se rendre compte des dangers que pourraient représenter ces outils selon leur phase de développement, la firme aurait ainsi utilisé les niveaux de biosécurité des laboratoires de virologie, l'ASL. Et d'après Dario Amodè, si nous sommes actuellement au niveau ASL 2, un niveau où les risques seraient déjà présents, mais réduits, nous devrions bientôt passer au niveau 4, un niveau dans lequel les intelligences artificielles seraient douées d'autonomie et seraient également capables de persuader les humains, un niveau où il pourrait y avoir davantage d'abus. Car le chercheur imagine les enjeux de taille que pourraient représenter ces IA sur la future scène géopolitique et envisage qu'elles pourraient être employées par des pays comme la Corée du Nord, la Chine ou la Russie pour renforcer leurs capacités offensives dans diverses zones militaires et envisage aussi que ces IA puissent être utilisées dans le but d'augmenter la capacité de nuisance de certains États. En bref, Dario Amodè nous dépeint un tableau très apocalyptique de ce que pourrait être notre futur dans les mois qui suivent.

[Virgule sonore, une cassette que l'on accélère puis rembobine.]

[Une musique de hip-hop expérimental calme.]

Pouvons-nous vraiment prendre au sérieux les prédictions du PDG d'Anthropic, Dario Amodè ? Pour rappel, ce dernier est à la tête de l'une des plus grandes sociétés génératrices d'intelligence artificielle. Le chercheur aurait fait ses classes à OpenAI il y a plusieurs années. C'est lui qui aurait d'ailleurs dirigé l'équipe qui a créé GPT-2 et GPT-3. Et aujourd'hui, il vient de mettre au point Claude 3, une IA considérée par beaucoup comme le modèle d'intelligence artificielle le plus puissant disponible à l'heure actuelle. Une intelligence artificielle dans laquelle le PDG aurait justement tout misé sur la sécurité et l'éthique afin qu'elle ne représente pas de danger pour l'Homme dans le futur. Donc, vous le comprenez, les prédictions du chercheur italo américain ne sont pas prises à la légère, même si ce dernier admet qu'il est impossible de prédire avec exactitude le développement de l'IA. Ces prévisions ne restent donc qu'hypothétiques. De plus, d'autres personnalités influentes du milieu de l'intelligence artificielle, comme Arthur Mensch, l'un des cofondateurs de l'entreprise française Mistral IA, ne croient pas en cette évolution de l'intelligence artificielle, estimant que cela relèverait plutôt d'une fascination religieuse. 2025, 2028, 2030... Plusieurs chercheurs estiment qu'un véritable tournant, qu'une véritable révolution devrait être établi pour l'intelligence artificielle prochainement. Nous verrons dans quelques années si le temps leur donne raison.

[Virgule sonore, un grésillement électronique.]

C'est tout pour cet épisode de Vitamine Tech. Pour ne pas manquer nos futurs épisodes, abonnez-vous dès à présent à ce podcast. Et si vous le pouvez, laissez-nous une note et un commentaire. Cette semaine, je vous invite à découvrir notre dernier épisode de Bête de science, dans lequel Agatha Liévin-Bazin vous parle des supers pouvoirs du gecko. Pour le reste, je vous souhaite une excellente journée ou une très bonne soirée et je vous dis à la prochaine dans Vitamine Tech.

[Un glitch électronique ferme l'épisode.]