

Vitalik Buterin prévient : l'IA, un “superprédateur” pour l'Humanité

Vitalik Buterin donne son avis sur l'intelligence artificielle

Vitalik Buterin donne son avis sur l'intelligence artificielle

Les capacités de l'intelligence artificielle, et ses dangers potentiels, sont au cœur de nombreuses discussions ces derniers temps. Entre opportunités inégalées pour l'humanité et menaces existentielles, le secteur est en pleine ébullition. Vitalik Buterin y est donc allé de son commentaire, par un des longs billets de blog dont il a le secret.

Pour le créateur d'Ethereum, l'IA est « fondamentalement différente » des autres inventions humaines, aussi importantes soient-elles. Il s'agit en effet de la première fois que l'Humanité crée quelque chose qui s'apparente à un esprit. Et un esprit dont les capacités mêmes peuvent surpasser celles des humains: « L'IA est un nouveau type d'esprit, qui gagne rapidement en intelligence. Et elle a une sérieuse chance de surpasser les capacités mentales des humains, et de devenir la nouvelle “superespèce” de la planète. »

Vitalik Buterin fait référence aux « superprédateurs », des animaux qui ne sont chassés par personne, et qui sont situés tout en haut de la chaîne alimentaire. Sur la planète, un seul animal peut chasser tous les autres: l'être humain. Mais il pourrait être remplacé si des intelligences artificielles prennent le contrôle.

Le créateur d'Ethereum estime que le danger des IA à qui on laisse libre cours de leurs actions, c'est qu'elles pourraient décider que les humains sont une menace. Une IA puissante qui se propagerait pourrait selon Vitalik Buterin causer « l'extinction des humains »: « Même Mars pourrait ne pas être un refuge. »

Ce scénario est extrême, mais il n'est pas uniquement présent dans la tête de Vitalik Buterin. En 2022, 4270 chercheurs en machine-learning avaient été interrogés, et ils estimaient les chances qu'une AI détruise l'humanité à 5 à 10%.

Une des solutions évoquées par Vitalik Buterin pour se protéger de ce danger est très transhumaniste. Il estime que des interfaces cerveau-ordinateur (BCI) pourraient être la solution, afin d'exercer un contrôle plus strict sur les intelligences artificielles. Le créateur d'Ethereum estime par ailleurs que l'humanité devra à un moment faire évoluer son cerveau: « Il existe des moyens variés d'améliorer nos cerveaux directement avec des innovations dans le secteur de la biologie. »

Il évoque aussi une éventuelle future étape qui permettrait de « télécharger » nos esprits afin qu'ils fonctionnent directement sur un ordinateur. Selon Vitalik Buterin, ces éventualités permettraient à l'humanité de se protéger : « Si nous voulons un futur qui soit à la fois superintelligent et "humain", un futur où les êtres humains ne seront pas simplement des animaux de compagnie [...], alors quelque chose comme ce que je viens d'évoquer semble être la meilleure option. »

De l'intelligence artificielle à Cyberpunk, il n'y a donc qu'un pas, selon Vitalik Buterin. Une chose est certaine: les capacités de l'IA sont au cœur des débats en ce moment...