



LA QUESTION QUI FÂCHE

LA CHRONIQUE

de Nicolas Journet

L'INTELLIGENCE ARTIFICIELLE va-t-elle nous tuer?

Ce qu'il y a de bien avec ChatGPT ou d'autres *chatbots*, c'est qu'on peut leur poser des questions insolentes, comme «*quelles sont les capacités de nuisance des intelligences artificielles?*», et qu'on obtient des réponses très pertinentes. Par exemple, pertes d'emploi, inégalités économiques, catastrophes industrielles, crise énergétique, décisions injustes, atteintes à la vie privée, désinformation, pertes de compétences et, finalement, affrontement meurtrier entre l'humanité et ses robots favoris... Si la machine n'ignore rien des risques qu'elle peut représenter, il se trouve que le milieu qui l'a fait naître non plus, et que c'est aussi celui qui nourrit les plus sérieux doutes sur son innocuité. Et pour cause, dira-t-on, il faut avoir trempé dans la fabrication d'algorithmes pour être capable d'envisager leurs méfaits.

Bien des Don Quichotte qui, ces derniers temps, s'en prennent aux IA génératives ont commencé par fabriquer des moulins dans la Silicon Valley. Le 16 septembre 2025 paraissait aux États-Unis un livre vite devenu best-seller intitulé *Pourquoi l'intelligence artificielle surhumaine nous tuerait tous*. Son coauteur, Eliezer Yudkowsky, était encore, il y a quelque temps, un bon ami de Peter Thiel, le milliardaire libertarien et technophile fondateur de PayPal et actionnaire de Meta. Michael Trazzi et Denys Sheremet, deux militants cités dans un article récent du *Monde* (1), sont en grève de la faim devant les locaux de Deep Mind : ils sont diplômés en IA, mais réclament l'arrêt de son développement. Nick Bostrom, philosophe formé en neurosciences computationnelles, est un des premiers à avoir, en Angleterre, brandi la menace d'une superintelligence homicide. Ainsi, il expliquait en 2003 qu'une IA souveraine chargée de maximiser la production de trombones pourrait très bien décider

de liquider les humains, puisque leur corps contient des substances utiles à la fabrication des trombones. Or, jusqu'il y a peu, Elon Musk finançait le centre dirigé par Bostrom.

On le voit, l'intelligence artificielle est au centre d'une querelle de famille dont la polarisation ne fait que s'accroître, après avoir réuni tout ce beau monde autour de l'idée que, oui, le succès était à portée de main. La philosophe Anne Alombert note, dans un pamphlet récent (2), qu'en 2023 encore, Elon Musk et d'autres experts de la high tech réclamaient une pause dans la recherche et parlaient d'«*esprits numériques*» menaçant l'humanité. Était-ce juste pour faire le buzz? On est tenté de la croire, car en dépit des mises en garde qui s'accumulent, rien ne semble devoir faire obstacle à l'accélération de la recherche, pour l'instant dans le parfait désordre, mais tournée vers le Graal de l'IA généralisée. Le plus choquant, note lucidement Anne Alombert, est que la plupart des autres industries sont tenues à respecter des règles parfois très sévères de sécurité. Il faut au minimum sept ans d'essais garantissant l'efficacité et l'innocuité d'une molécule avant de mettre un nouveau médicament sur le marché. Étonnamment, les produits d'intelligence artificielle ne sont soumis à aucun principe de précaution, aucune garantie de qualité, aucune preuve d'innocuité. Pourquoi est-on réduit à redouter que, peut-être, ils rendront leurs usagers idiots et analphabètes? Pourquoi les débats à leur sujet ne reposent-ils que sur de vagues prophéties? Bonnes questions, non? ●

(1) Jules Darmanin, «*"Nous allons perdre le contrôle" : comment les catastrophistes de l'IA haussent le ton*», *Le Monde*, 18 octobre 2025.

(2) Anne Alombert, *De la bêtise artificielle*, Allia, 2025.