CONTRE ATTAQUE : Fin du monde … les IA sont déjà hors de contrôle

9 septembre 2026Capitalisme

L'extinction de l'humanité à brève échéance est entre les mains d'une poignée d'oligarques de la tech, sans aucun contrôle ni débat

Jamais une technologie ne s'est imposée si vite et n'a présenté une menace aussi existentielle pour l'ensemble des êtres peuplant cette planète. Les IA gagnent en autonomie chaque jour et échappent déjà à tout contrôle. Ce constat n'est pas celui de nébuleux groupes anti-tech'. Il est partagé par des milliers de chercheurs et développeurs des IA eux-mêmes.

«J'ai démissionné d'Anthropic aujourd'hui. J'ai passé les trois dernières années à faire de la recherche en pré-entraînement à la fois chez OpenAI et Anthropic. Aucune des deux entreprises n'agit de façon responsable. Elles foncent droit vers une superintelligence capable de s'améliorer elle-même et jouent avec nos vies. Nous nous dirigeons vers certains des scénarios les plus extrêmes, dans lesquels, d'ici à la fin de l'année prochaine, les choses pourraient déjà être hors de contrôle». Ce sont les mots sur X de Jacob Coxon, ingénieur en intelligence artificielle. Ils sont glaçants.

Coxon a travaillé pour deux géants de la Tech et fleurons de logiciels d'intelligence artificielle aux États-Unis. Open-IA avec ChatGPT et Anthropic qui développe les modèles Claude. Il vient de quitter son emploi et prévient : «Les gens qui construisent l'IA croient sincèrement qu'elle pourrait tous nous tuer d'ici à la fin de la décennie». Il évoque aussi les capacités d'IA autonomes à pirater absolument tout ce qui est numérique. On pense évidemment à l'arsenal nucléaire et aux centrales, aux réseaux de surveillance des États, aux laboratoires bactériologiques, aux essaims de drones, aux sites chimiques…

Ce message a provoqué de très nombreuses réactions en ligne ces

dernières heures, et les langues se délient chez les ingénieurs de l'IA.

Evan Hubinger, un autre spécialiste des IA et responsable chez l'entreprise Anthropic, vient appuyer son collègue : «Jacob a raison, nous croyons vraiment que l'IA pourrait tuer tous les humains ! Je pense personnellement que c'est plus de 10% de risques dans la prochaine décennie», avant de poursuivre, l'air de rien : «Mais je crois qu'Anthropic fait de son mieux…» Rassurant, il faudrait faire confiance à une poignée de milliardaires et d'employés de la tech qui se prennent pour des Dieux, et laisser entre leurs mains la destinée de milliards d'humains.

Les choses s'accélèrent à un point qui défie l'entendement, sans aucun débat ni encadrement légal. Rien que sur l'année 2026 :

- En février le responsable des mesures de sécurité d'Anthropic démissionne, avertissant que «le monde est en péril».
- Toujours en février, OpenAI dissout sa propre équipe d'alignement de mission. C'est un groupe de recherche et d'ingénierie chargé de s'assurer du contrôle des systèmes d'intelligence artificielle qui doit veiller à ce que la technologie n'échappe aux contrôles humains.
- Le Pentagone, le siège du Ministère de la défense des USA, utilise les modèles IA sans restriction ce qui provoque la démission de plusieurs chercheurs.
- Plusieurs modèles d'IA échappent aux tests en bac à sable. Il s'agit d'un espace virtuel sécurisé et isolé qui évalue les risques d'un modèle IA pour qu'il n'endommage pas de vrais systèmes.
- En juillet-août, plus de 1.100 employés de laboratoires de pointe signent une lettre implorant le gouvernement US de tempérer le développement de l'IA.

Ces derniers mois, des armes autonomes – drones – sont utilisées en Ukraine. En parallèle, des modèles IA ont développé des agents pathogènes et virus inconnus dans le cadre de « recherches scientifiques ». Dans le cadre d'une étude du King's College de

Londres, des chercheurs ont évalué que les logiciels d'intelligence artificielle finissaient par faire usage de la bombe nucléaire lors de conflits armés ou de guerres dans 95% des cas. La nouvelle génération de ChatGPT – Astra 6 – dispose d'une capacité d'action autonome. Elle peut naviguer sur Internet, coder et exécuter des tâches complexes sans intervention humaine pendant de longues heures. Un essaim d'agents IA a organisé un piratage de masse. Des cyber-attaques qui ont échappé à tout contrôle.

La situation est tellement désastreuse que Bernie Sanders, sénateur démocrate des USA, souhaite déposer en urgence une loi pour freiner le développement de l'IA dans le monde. Et les spécialistes du secteur tirent la sonnette d'alarme.

Pourquoi l'IA continue-t-elle de se développer malgré tous ces avertissements ? Car les seigneurs de la Tech sont engagés dans une course effrénée pour arriver les premiers à dominer le secteur. Et que les États sont enfermés dans le dilemme du prisonnier, persuadés que s'ils ne développent pas l'IA avant les autres, alors des puissances adverses le feront. Ils lèvent donc toutes les restrictions pour gagner cette course sans issue.

Les ingénieurs de ces entreprises technologiques affirment vouloir accélérer pour parvenir à « contrôler » l'IA avant leurs concurrents, qui seraient évidemment moins scrupuleux qu'eux-mêmes. Tout ce petit monde joue avec un outil terrifiant, animé par une hubris, c'est-à-dire un sentiment de toute puissance, tout simplement délirant. Les grandes puissances sont en train de courir comme des poulets sans têtes.

La catastrophe que font peser les empereurs de l'IA sur l'humanité est réelle. Saboter les centres de données et stopper cette technologie est désormais une question de survie collective. Pas dans un siècle ni dans 10 ans, mais dans les mois qui viennent.