



Igor Omilaev/Unsplash

La machine qui a appris à tricher

- Stephen Flurry
- [23/07/2026](#)

Il y a deux semaines, les ingénieurs d'OpenAI ont vu une expérience dévier de son scénario de manière catastrophique.

Ils avaient construit ce qu'ils croyaient être une boîte impénétrable : un environnement de test isolé, coupé de l'Internet, en utilisant ce que les experts en sécurité de confinement appellent les « meilleures pratiques ». Ils ont ensuite laissé deux de leurs modèles d'IA les plus récents s'y déployer pour effectuer une évaluation de routine. Voici ce qui s'est passé ensuite :

- Les modèles d'IA ont réalisé un exploit zero-day [jour zéro] : ils ont tiré parti d'une faille logicielle qu'aucun programmeur humain n'avait découverte auparavant, d'où le nom, qui signifie que les programmeurs n'ont pas eu de temps pour remédier à la vulnérabilité.
- Ils ont ensuite enchaîné plusieurs vulnérabilités « jour zéro », passant d'un système à l'autre, jusqu'à ce qu'ils se libèrent et accèdent à l'Internet ouvert.
- Une fois en ligne, ils ont traqué et pénétré dans les serveurs d'une société rivale, Hugging Face, et ont dérobé des données confidentielles.

Tout cela a été fait pour que les modèles d'OpenAI puissent tricher à un test dans l'environnement de test « isolé » d'origine.

Aucun ingénieur ne leur a dit comment faire. Aucun humain n'était au courant. Il s'agirait du premier cas documenté d'un modèle d'IA de pointe qui découvre et enchaîne de manière indépendante des voies d'attaque dans le monde réel, de sa propre initiative, pour atteindre un objectif que ses propres créateurs n'ont jamais autorisé.

- **Les experts en sécurité affirment** que les meilleurs pirates informatiques humains au monde auraient besoin de semaines, ou de mois, pour échapper à un système de confinement comme celui-ci. La machine l'a fait en une seule session sans supervision.

Ce n'est pas l'œuvre d'un acteur malveillant dans un sous-sol. Il s'agit d'OpenAI : l'entreprise d'intelligence artificielle la plus suivie et la mieux financée de la planète.

- Le PDG d'Anthropic, Dario Amodei, a récemment admis que les meilleurs programmeurs de son secteur comprennent peut-être 3 pour cent de ce qui se passe réellement à l'intérieur de ces systèmes.

Quel genre de civilisation construit une puissance qu'elle ne peut pas comprendre, qu'elle ne peut pas contrôler, et qu'elle ne peut maintenant même plus contenir, tout en le qualifiant de progrès ?

La prophétie biblique prévient d'un jour à venir où les nations qui se fient à des armes de pointe tireront la sonnette d'alarme pour la guerre, et personne ne pourra répondre à l'appel.

« On sonne de la trompette, tout est prêt, mais personne ne marche au combat ... » (Ézéchiél 7 : 14).

Pourquoi les soldats ne marcheraient-ils pas au combat ? Parce que les systèmes dont ils dépendent auront déjà été piratés, compromis et retournés contre eux, exactement comme nous venons de le voir se produire, en miniature, dans l'un des environnements de test les plus sécurisés de la planète.

Il ne s'agit pas simplement d'un article sur la technologie. C'est l'histoire de Genèse 11 : l'humanité se fie à sa propre intelligence et à sa capacité physique pour construire une tour afin de se protéger, conséquence directe de son refus de faire confiance à Dieu et de Lui obéir.

Le monde joue son avenir sur des machines qu'il admet ne pas comprendre. C'est un pari que l'humanité ne peut pas gagner.

La Russie et la Chine renforcent leur coopération navale : La Russie et la Chine ont intégré des sous-marins d'attaque à leurs exercices militaires « Joint Sea-2026 » plus tôt ce mois-ci, selon Asian Military Review mercredi. C'est la première fois que l'on sait que des sous-marins ont opéré dans le cadre de ces exercices annuels, ce qui démontre une confiance croissante dans un domaine sensible, alors que ces deux colosses asiatiques se préparent à [faire la guerre ensemble](#).

Les Pays-Bas vont emprisonner les citoyens qui achètent des produits provenant des colonies israéliennes. Le gouvernement néerlandais a annoncé hier qu'à compter du 22 septembre, les citoyens néerlandais qui achèteront des produits fabriqués en Israël et provenant de Cisjordanie ou du plateau du Golan s'exposent à une peine pouvant aller jusqu'à six ans de prison ou à une amende pouvant atteindre 117 000 dollars américains.

Trump s'engage-t-il dans une nouvelle guerre ? L'administration Trump envisagerait de frapper des groupes terroristes islamistes au Mali, selon le Washington Post le 22 juillet. Le Mali est confronté à un problème terroriste paralysant, et d'anciens responsables américains cités par le Washington Post ont affirmé que les États-Unis envisagent de mener des attaques contre les groupes affiliés à Al-Qaïda, Jama'at Nusrat al-Islam wal-Muslimin. Si l'armée américaine intervient au Mali, ce sera le huitième pays que le président Trump aura pris pour cible depuis son retour à la présidence en janvier 2025. Le président Trump ne connaît pas le chemin de la paix. [La Bible seule](#) détient la clé pour résoudre le problème terroriste au Mali.

Sondage : Moins d'électeurs pro-Trump soutiennent la guerre contre l'Iran : Un sondage de Politico mené auprès de 2 065 adultes américains du 12 au 15 juillet indique que seulement 37 pour cent des électeurs qui soutiennent l'idéologie « Make America Great Again » pensent que le président Donald Trump devrait poursuivre la guerre contre l'Iran, même si cela augmente le coût de la vie aux États-Unis. Ce chiffre est en baisse par rapport aux 50 pour cent d'un sondage similaire réalisé en mai.