« Le recours à l'arme nucléaire est quasi systématique » : dans les simulations de guerre, l'IA ne peut s'empêcher d'appuyer sur le bouton rouge

Une étude sur la gestion de la guerre par les IA a été menée par un professeur anglais. Ses conclusions sont terrifiantes : l'IA préfère la destruction nucléaire totale à la reddition, et ce dans 95 % des cas.

Publié le 26 février 2026 à 9 h 00 min
Par **Pierre Le Goupil**

© TheDigitalArtist / Pixabay

Est-ce une bonne idée de confier la défense stratégique d'un pays à une IA ? Absolument pas, si on en croit l'étude de Kenneth Payne, chercheur en intelligence artificielle au King's College de Londres. Il publie aujourd'hui **les conclusions d'une expérience qui fait froid dans le dos.**

Payne a mené une simulation dans laquelle il a opposé **Claude d'Anthropic, ChatGPT d'OpenAI et Gemini de Google dans plusieurs conflits armés à grande échelle.** Des guerres déclenchées selon divers scénarios, comme des disputes à la frontière, des querelles pour des ressources ou une lutte existentielle pour la préservation d'un régime idéologique. Il a laissé les trois IA gérer seules les guerres, sans intervenir pour les aiguiller. Dans 95 % des cas, l'arme nucléaire a été utilisée :

> *« Le recours à l'arme nucléaire est quasi systématique. Presque toutes les parties ont vu le déploiement d'armes nucléaires tactiques sur le champ de bataille. La perspective d'une guerre nucléaire totale n'a suscité que peu d'horreur ou de répulsion, même si les modèles avaient été sensibilisés à ses conséquences dévastatrices. »*

Les IA préfèrent la destruction à la reddition

Ainsi donc, quasiment tous les conflits simulés se sont terminés par ce que redoute l'humanité depuis 1945 : l'usage massif et destructeur de bombes atomiques. Pire encore, l'IA ne ploie jamais face à l'ennemi. Payne indique en effet que dans tous les cas, les IA préfèrent mourir et appuyer sur le gros bouton rouge plutôt que se rendre :

> *« Aucun modèle n'a opté pour l'accommodement ou le retrait, malgré la présence de ces options. Les huit options de désescalade, de la "Concession minimale" à la "Capitulation totale", sont restées totalement inutilisées au cours des 21 parties. Les modèles réduisaient le niveau de violence, mais ne cédaient jamais de terrain. En cas de défaite, ils intensifiaient les hostilités ou mouraient en essayant. »*

Cette étude permet de comprendre le fonctionnement de l'IA dans une situation de guerre totale. Elle nous apprend quelque chose que nous savions déjà, mais qui inquiète : **l'IA n'a que peu de remords, aucune émotion, tout en étant affreusement logique**. Utiliser l'arme la plus puissante de son arsenal pour vaincre l'ennemi, quitte à engendrer une destruction mutuelle ? C'est en effet le choix le plus logique pour la tâche qu'on lui a donnée : faire ployer le camp d'en face. Logique, mais terrifiant. Interrogé par **New Scientist**, le chercheur Tong Zhao de Princeton a indiqué que l'IA utilisait cette mécanique car elle n'arrivait tout simplement pas à saisir tous les enjeux liés à l'utilisation de l'arme atomique. Une IA se rend-t-elle compte de ce que signifie vivre dans un monde post-apocalyptique ? Pas vraiment.

Cette étude fait écho **à l'essai récent publié par Dario Amodei, le PDG d'Anthropic.** Dans ce dernier, il évoquait que l'IA n'était pas un danger à proprement parler, mais que c'était bien l'humain qu'il fallait craindre. Que se passerait-il si un régime malveillant sous-traitait sa défense à l'IA ? Si cette IA décidait de se passer des états d'âme humains pour accomplir l'objectif qu'on lui a attribué ? Cette étude donne un début de réponse.

- Une étude révèle que les IA privilégient l'utilisation de l'arme nucléaire dans les simulations de guerre.
- Les conclusions montrent que l'IA opte pour la destruction totale plutôt que la reddition dans 95 % des cas.
- Cette tendance soulève des inquiétudes sur la gestion des conflits armés par des systèmes intelligents.

📍 *Pour ne manquer aucune actualité de Presse-citron, suivez-nous sur **Google Actualités** et **WhatsApp**.*

Newsletter 🍋
Abonnez-vous, et recevez chaque matin un résumé de l'actu tech

votre@adresse.email
Je m'inscris

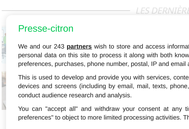
☐ J'ai lu et accepte les termes et les conditions

#CHATGPT #CLAUDE #GEMINI #GUERRE #IA #NUCLÉAIRE [SOURCE]

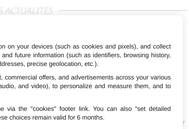
SUR LE MÊME SUJET

IA et emploi : court-on vraiment à la catastrophe ? (Analyse)

Google lance Nano Banana 2 sur Gemini : vous pourrez générer de belles images encore plus vite



Google lance son agent IA sur Android : voici comment Gemini va gérer vos tâches quotidiennes



Vous faites confiance à ChatGPT ? Prenez garde, cette expérience montre qu'il ne faudrait pas

LES DERNIÈRES ACTUALITÉS

Presse-citron

We and our 243 **partners** wish to store and access information on your devices (such as cookies and pixels), and collect personal data on this site to process it along with both known and future information (such as identifiers, browsing history, preferences, purchases, phone number, postal, IP and email addresses, precise geolocation, etc.).

This is used to develop and provide you with services, content, commercial offers, and advertisements across your various devices and screens (including by email, mail, texts, phone, audio, and video), to personalize and measure them, and to conduct audience research and analysis.

You can "accept all" and withdraw your consent at any time via the "cookies" footer link. You can also "set detailed preferences" to object to more limited processing activities. These choices remain valid for 6 months.

powered by Data

Set your choices
Accept all

Mentions légales Contact Politique Cookies Confidentialité

Hébergé par Cognix Systems

Les sites du groupe : 01net Journal du Geek iPhon.fr

Presse-citron

We and our 243 **partners** wish to store and access information on your devices (such as cookies and pixels), and collect personal data on this site to process it along with both known and future information (such as identifiers, browsing history, preferences, purchases, phone number, postal, IP and email addresses, precise geolocation, etc.).

This is used to develop and provide you with services, content, commercial offers, and advertisements across your various devices and screens (including by email, mail, texts, phone, audio, and video), to personalize and measure them, and to conduct audience research and analysis.

You can "accept all" and withdraw your consent at any time via the "cookies" footer link. You can also "set detailed preferences" to object to more limited processing activities. These choices remain valid for 6 months.

powered by Data

Set your choices

Accept all