

[Offrir un abonnement](#)[Mon compte](#)[Menu](#) [Société](#) [Politique](#) [Monde](#) [Ecologie](#) [Economie](#) [Idées](#) [Culture](#) [Tendances](#)[Sciences](#) • [Intelligence artificielle](#)

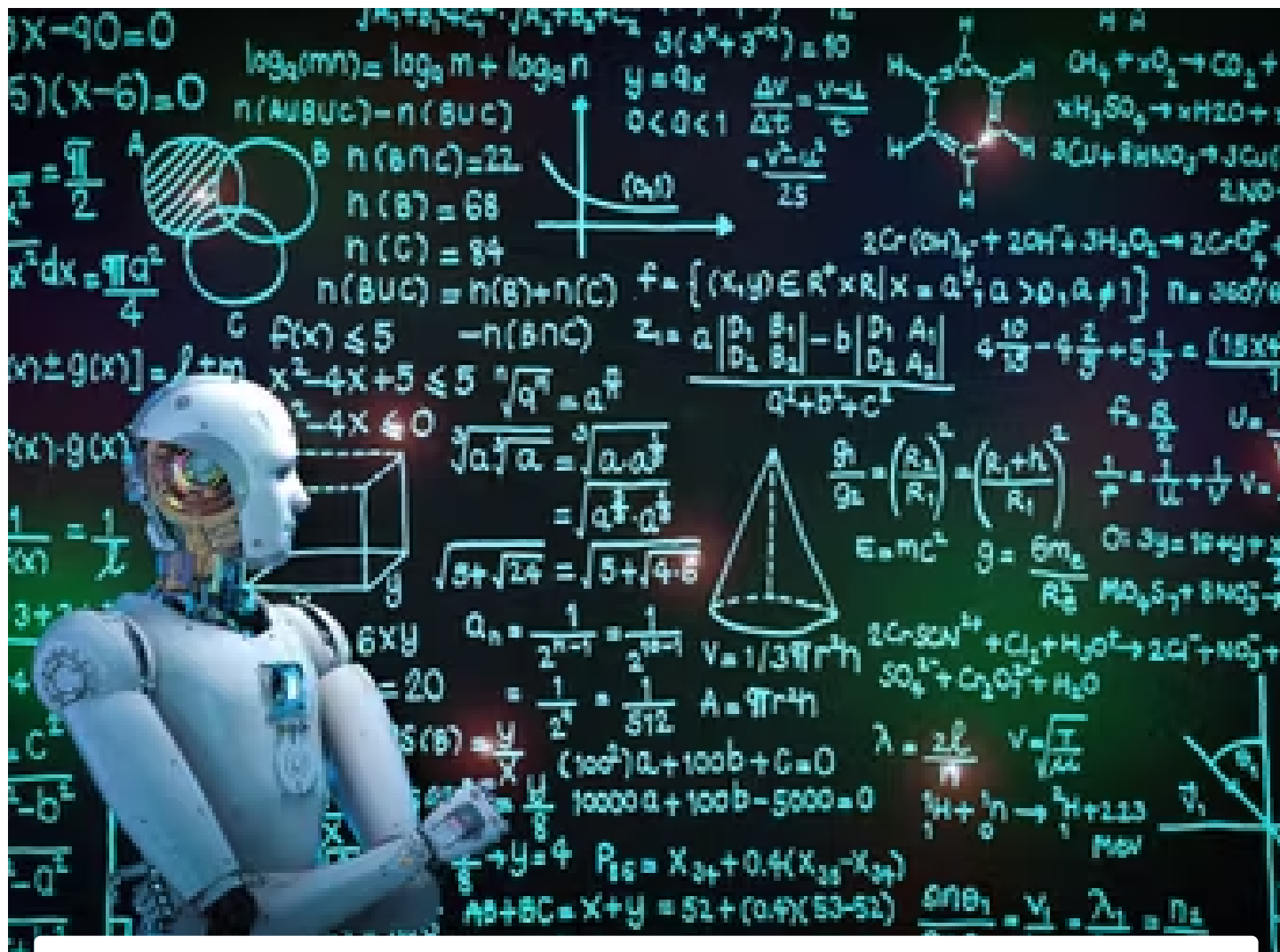
L'IA peut-elle rivaliser avec les meilleurs mathématiciens ?

Décryptage Le projet First Proof, qui vise à évaluer le niveau des IA en recherche mathématique, a rendu ses conclusions. Elles font encore beaucoup d'erreurs, sont parfois incompréhensibles, mais parviennent désormais à résoudre, toutes seules, des problèmes d'un très bon niveau.

Par **Oscar Leroy**

Publié le 5 juillet 2026 à 17h30

Lecture : 7 min. [Abonné](#)



Annuler

même si pour le moment, les très bons mathématiciens restent meilleurs. PHONLAIMAPHOTO/GETTY IMAGES/ISTOCKPHOTO

 Offrir cet article

 Commenter



Pour aller plus loin



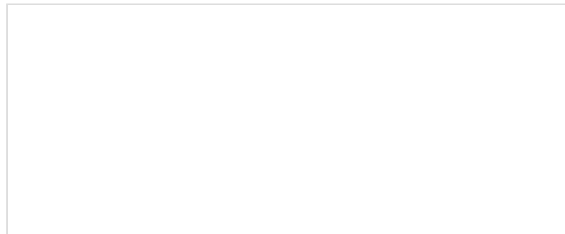
Dossier Notre sélection du week-end

Ç'aura pris plus de temps que dans d'autres secteurs, mais l'IA arrive maintenant à produire des résultats remarquables en mathématiques. On ne parle plus seulement de résoudre des problèmes d'Olympiades – le championnat international pour les élèves du secondaire – ou de fouiller Internet pour trouver une solution déjà existante. Le projet First Proof, lancé par des mathématiciens américains de haut rang afin d'évaluer le niveau actuel des IA dans la recherche, a rendu ses conclusions il y a quelques semaines. « Notre

but était de répondre au discours excessif des boîtes d'IA sur la capacité de leurs modèles et de sensibiliser la communauté mathématique à ces outils précieux dont beaucoup semblent ignorer l'utilité », nous explique Lauren Williams, mathématicienne à Harvard et coordinatrice du projet.

L'équipe a demandé à dix mathématiciens de sélectionner un problème qu'ils avaient résolu mais sans avoir encore publié le résultat, et dont aucune solution n'était disponible en ligne. « *Nous devons procéder avec une grande prudence, car les modèles d'IA sont extrêmement performants pour la recherche sur le web* », souligne Lauren Williams. Les chercheurs ont choisi des démonstrations assez courtes (huit pages au plus), qui demandent une intuition sortant des schémas classiques. « *Certains des problèmes avaient été résolus en quelques jours par les chercheurs quand d'autres ont nécessité plusieurs semaines* », explique Lauren Williams.

Publicité



Annuler



L'énoncé de chaque problème a ensuite été donné à quatre modèles. L'un d'entre eux est simplement la dernière version pro de ChatGPT. Les trois autres sont des programmes qui font appel à plusieurs grands modèles de langage (LLM) comme ChatGPT, en leur indiquant comment s'y prendre. Par exemple, le programme peut lui demander d'examiner le problème et de le résoudre. Un autre LLM relit la preuve et, si elle n'est pas suffisamment bonne, la soumet à un comité composé des meilleurs modèles d'OpenAI, de Google et de Perplexity, qui font des suggestions, et ainsi de suite. « *La qualité des preuves dépend vraiment de celles des modèles*, explique Amaury Hayat, professeur à l'Ecole des Ponts, et spécialiste de l'IA pour les mathématiques, qui n'a pas participé au projet First Proof. *Il ne faut pas compter sur la version gratuite de ChatGPT pour sortir une solution correcte.* » Pour chacun des problèmes, First Proof a lancé les quatre programmes, une seule fois, avec une durée limite de vingt-quatre heures, sans aller-retour avec l'utilisateur.

Des résultats prometteurs

Et donc, que valent ces IA ? Les très bons mathématiciens, spécialistes de leur domaine, restent meilleurs. Mais pour plus de la moitié des problèmes, au moins un des quatre programmes a fourni une preuve correcte. « *Nous avons vu toute une gamme de solutions, dont certaines sensiblement différentes de celles proposées par les humains* », raconte Lauren Williams. Pour l'un des problèmes, une IA a même établi « *un résultat intermédiaire légèrement plus solide que la solution humaine* ». La première qualité de ces modèles, c'est d'avoir une connaissance extensive de la masse des publications mathématiques. « *Ça leur permet de mettre en rapport un problème qu'on cherche à résoudre avec d'autres domaines des mathématiques que l'on connaît moins bien et dont on ne soupçonnait pas le lien.* »

A lire aussi



Entretien « Les entreprises de l'IA ressemblent à des joueurs de poker qui font tapis »

Abonné

« *Les capacités dont ont fait preuve les IA dans ce test sont cohérentes avec ce qu'on observe quotidiennement, il n'y a pas de surprise*, dit Amaury Hayat. *Le plus intéressant ici, ce sont les observations qualitatives des examinateurs sur les défauts des modèles.* » Celle qui vient la

[Annuler](#)

ChatGPT utilise un theoreme qui correspond exactement a celui qu'il doit prouver, et déduit trivialement le résultat attendu. « *Contrairement à nous, les modèles d'IA ont toujours un niveau de confiance très élevé, même lorsqu'ils avancent quelque chose de faux* », explique Lauren Williams. Cette certitude constante des LLM, couplée à leur amélioration, font que les erreurs de raisonnement sont de plus en plus dures à repérer. « *Avec un collègue, nous avons récupéré une preuve difficile, avec des idées très jolies, mais qui contenait un argument complètement faux, et on a failli se faire avoir*, explique Amaury Hayat. *Ça va être un de nos grands défis en mathématiques.* »

Les rendus des IA sont souvent assez peu « humains » : elles ont tendance à s'attarder sur les passages triviaux des preuves, tout en étant laconiques sur les parties les plus cruciales. « *Dans une démonstration, l'IA avait passé une demi-page sur un résultat qui mérite une demi-ligne* », raconte Amaury Hayat. Elles laissent parfois des détails inutiles et triviaux, tels que « $(d - 1) - b = d - b - 1$ », dans un des problèmes... Les solutions, même correctes, sont souvent rédigées avec une qualité d'explication médiocre. « *Certaines étaient extrêmement confuses et comportaient des notations tellement obscures qu'il était très difficile de les vérifier* », dit Lauren Williams. Pour l'une d'entre elles, « *le seul problème que j'ai remarqué, et qui n'est pas grave, est qu'au milieu de la solution se trouve une description générale de la stratégie pour compléter la démonstration*, indique un examinateur. Cette

description est au mieux trompeuse, au pire une description erronée de ce qui se passe réellement. » Une tendance des IA, cette fois plus humaine, est le plagiat : First Proof observe plusieurs cas où des travaux sont repris sans être cités.

Le plafond des IA

Tous ces défauts rendent la vérification ardue, font que dans un certain nombre de cas, la solution sera fausse, mais l'IA est-elle fondamentalement limitée, incapable de résoudre des problèmes de très haut niveau, se contentant de résultats intermédiaires ?

L'interrogation sous-jacente est sa faculté à faire preuve d'originalité, à trouver des méthodes novatrices, voire à inventer de nouveaux concepts, ce qui constitue le cœur de la recherche mathématique. *« Les problèmes où les LLM ont le moins réussi sont globalement ceux dont les solutions étaient les plus différentes des idées apparues dans des articles publiés auparavant »*, affirme Lauren Williams. A contrario, beaucoup de celles qui sont correctes s'inspirent de démonstrations analogues et les adaptent au problème. *« Pour l'instant, elles ont du mal à créer de nouveaux concepts et définitions, dit Amaury Hayat. Elles n'en sont pas encore à réaliser de longues preuves très innovantes »*. *« Mais je ne mettrais pas ma main à couper que ça restera le cas »*, avance Sylvia Serfaty, mathématicienne et professeure à Sorbonne Université, qui n'a pas participé au projet First Proof.

A lire aussi

Annuler



Abonné

« Désormais, la meilleure manière de mesurer leur niveau va être de regarder quels problèmes ouverts elles arrivent à démontrer », affirme Amaury Hayat. Car les résultats de First Proof ne donnent pas d'information sur le plafond des IA. Depuis plusieurs mois, Google DeepMind ou OpenAI ont annoncé, à plusieurs reprises, avoir résolu certains problèmes conjecturés par le mathématicien hongrois Paul Erdős. Jusque-là, il s'agissait de résultats mineurs, une étape a été franchie il y a un mois, lorsqu'OpenAI a trouvé la solution à un problème de « géométrie combinatoire », également posé par Erdős, mais bien plus difficile que les précédents. Le modèle a établi un lien avec un autre domaine des mathématiques – la théorie algébrique des nombres –, avec assez d'originalité pour que les mathématiciens n'aient pas envisagé cette solution. *« Je pense que l'on considérera cet événement comme la première fois qu'une IA a résolu un problème mathématique majeur (défini comme un problème sur lequel s'étaient penchés tous les experts d'un sous-domaine donné) »*, a écrit le médaillé Fields Timothy Gowers sur Bluesky. Bien sûr, ce n'est qu'un seul résultat et nous n'avons pas la liste de tous ceux que l'IA n'est pas parvenue à démontrer, mais il témoigne du niveau qu'elle peut, dès aujourd'hui, atteindre. *« Fatalement, d'autres résultats*

de ce type vont tomber », affirme Sylvia Serfaty. En un an, les IA ont progressé à toute vitesse : elles ont obtenu pour la première fois la médaille d'or aux Olympiades l'année dernière, et savent maintenant résoudre des problèmes de recherche d'un niveau croissant.

Les progrès s'accroissent

Des progrès récents pourraient renforcer cette dynamique. L'IA est de plus en plus forte pour traduire une preuve du langage naturel (comme l'écrirait un humain, avec des phrases et des formules) vers le langage formel, une succession de lignes de codes qui permet à l'ordinateur de vérifier la preuve mécaniquement. *« Ces dernières semaines, [la traduction automatique vers le langage formel] a atteint un stade où pratiquement toutes les tâches de formalisation que j'avais lancées pouvaient être accomplies en quelques heures »*, indique le médaillé Fields Terence Tao dans un blog. Les solutions fournies par les LLM pourraient alors être vérifiées automatiquement pour leur permettre de s'auto-entraîner sur de grandes bibliothèques de preuves. Mais à ce jour, la méthode est encore trop coûteuse pour être déployée à une telle échelle.

Quoi qu'il en soit, les mathématiciens prennent désormais l'IA au sérieux, elle qui était raillée, à cause de son faible niveau, dans les premières années qui ont suivi l'explosion de l'IA générative en 2022. *« Ça va démultiplier les possibilités de démontrer des choses difficiles, avance Sylvia Serfaty. On aura une idée de démonstration et on pourra en quelque sorte faire*

Annuler

Cette évolution technologique rappelle en un sens celle des ordinateurs, qui ont permis d'effectuer des calculs inaccessibles aux humains, de lancer des simulations pour émettre des conjectures, etc. *« Je pense que l'IA va devenir un outil de ce type, qui permettra aux chercheurs de devenir encore meilleurs »*, dit Amaury Hayat. Peut-être verrons-nous se généraliser des petits modèles, entraînés spécifiquement pour comprendre les structures de certains problèmes et proposer des conjectures, des pistes à suivre, sans toutefois fournir de preuve. *« Il y a un certain nombre de problèmes où avoir l'intuition de la bonne formule à utiliser résout essentiellement le problème »*, explique Amaury Hayat.

A lire aussi



Analyse Les jeunes face à l'IA, ou la difficulté de construire son identité au travail

Abonné

« Pendant un moment, il va rester un rôle important pour les humains, ne serait-ce que pour choisir les problèmes, piloter l'IA, orienter les directions de recherche ou interpréter les résultats »,

ajoute Sylvia Serfaty. Un groupe de mathématiciens a publié début juin la « [déclaration de Leyde](#) » pour réfléchir au futur du domaine. Ils y exposent plusieurs menaces que les IA font peser sur les mathématiques : bien sûr leurs erreurs et la difficulté de les repérer, le plagiat, mais aussi l'inégalité entre chercheurs dans l'accès à ces modèles qui coûtent cher – pour le meilleur des quatre utilisés par First Proof, le prix variait de 35 à 950 dollars selon les problèmes. « *Dans mon usage, ça coûte entre 20 et 100 euros par requête* », avance Amaury Hayat. Les chercheurs s'inquiètent aussi de l'implication croissante des boîtes d'IA, qui pourrait remodeler la recherche en mathématiques, avec la multiplication de communiqués de presse comme outil de partage des résultats ou le privilège donné aux domaines où les IA sont les plus performantes, au détriment des autres.

Par Oscar Leroy



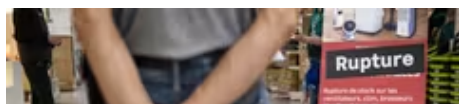
Offrir cet article



Commenter



Nos lecteurs ont lu ensuite



Annuler

Décrvptage Canicule : petites et grandes arnaques pour des

Abonné



Edward Carpenter et George Merrill : l'histoire d'amour inimaginable entre un intellectuel et un ouvrier dans l'Angleterre victorienne

Abonné



Reportage En meeting à Paris, Edouard Philippe se dévoile : « Je vais vous parler de moi »

Abonné



Billet Question pratique : ça va durer encore longtemps, cette Coupe du Monde de foot ?

Abonné



Rencontre Rima Hassan à la veille de son procès pour apologie du terrorisme : « Je me suis sentie trahie par mon pays »

Abonné



Récap Deux bateaux coulés, 150 secours mobilisés... Ce que l'on sait de l'incendie dans le Vieux-Port de Marseille

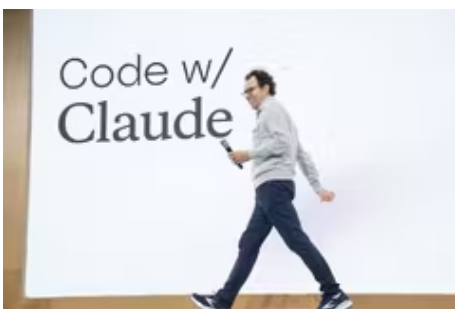
Annuler

Dans la rubrique Intelligence artificielle



Entretien « Le filet de sécurité social devra être renforcé pour absorber ce choc qu'est l'IA »

Abonné



Décryptage « Underclass » : pourquoi les patrons de l'IA parlent-ils de la constitution d'une « sous-classe » ?

Abonné



Analyse Les jeunes face à l'IA, ou la difficulté de construire son identité au travail

Abonné



Billet Et si l'histoire de la vanille nous aidait à anticiper les effets de l'IA sur nos écrits ?

Abonné



Entretien « Les entreprises de l'IA ressemblent à des joueurs de poker qui font tapis »

Abonné

Annuler



Récit Mélenchon, Le Pen, Philippe et Attal dans le bateau de l'IA à VivaTech, qui tombe à l'eau ?

Abonné

Dans la rubrique Sciences



Décryptage Quand aura lieu la prochaine canicule ? Pourquoi il n'est pas si simple de prévoir la météo au-delà de trois jours

Abonné



Décryptage Chez les enfants, le cerveau porte les marques du milieu social dans lequel ils grandissent

Abonné



Portrait Qui est Arnaud Prost, l'astronaute français qui devrait voler vers la première station spatiale commerciale ?

Abonné



Décryptage Envoyer des data centers dans l'espace ? Pourquoi l'idée d'Elon Musk est déraisonnable

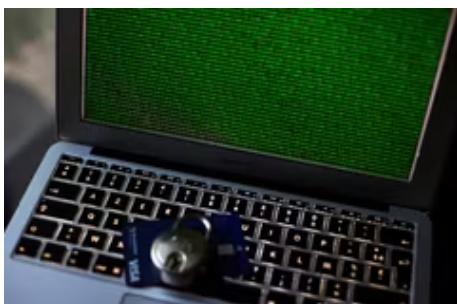
Abonné

Annuler



Décryptage La canicule précoce en mai préfigure-t-elle un été 2026 infernal ?

Abonné



Décryptage Cyberattaques en série : nos données personnelles valent-elles encore quelque chose si tout le monde les a piratées ?

Abonné

Sujets associés à l'article

Economie



Cours de langue

Service partenaire

Cours d'italien

Test gratuit

Cours de français

Cours d'allemand

Les univers

EcoloObs

Annuler

TéléObs

Tendances

Culture

Nouvelle ère

Les événements du Nouvel Obs

Dans l'actu

Canicule

Coupe du Monde 2026

Marine Le Pen

Accès rapide

Programme TV gratuit et complet

La conjugaison

Fonds pour l'indépendance de la presse

Les services

[Le magazine numérique](#)

[S'abonner / se désabonner](#)

[Avant-premières et jeux concours](#)

[Les newsletters](#)

[Les mots croisés](#)

[La boutique](#)

[Les archives](#)

[Le plan du site](#)

[Bilan retraite](#)

Les sites du groupe

[Le Monde](#)

[Télérama](#)

[Courrier international](#)

[Le HuffPost](#)

[La Vie](#)

[Le Monde diplomatique](#)

Le magazine

[Annuler](#)



L'application du Nouvel Obs



Liens utiles :

[Mentions légales](#)

|

[CGV](#)

|

Copyright

|

Publicité

|

Politique de confidentialité

|

Gestion des cookies

|

Aide / Contact

© Le Nouvel Obs - 2026

Les marques ou contenus du site [nouvelobs.com](https://www.nouvelobs.com) sont soumis à la protection de la propriété intellectuelle.

[Annuler](#)