

ISABEL ESPANOL

# Quand les IA se mesurent aux mathématiciens

Aux Olympiades internationales de mathématiques, des intelligences artificielles dont les capacités ont décuplé ces dernières années remportent désormais des médailles d'or. Si elles chamboulent déjà le monde de la recherche, seront-elles pour autant un jour capables de surpasser les meilleurs chercheurs de la discipline au point de les supplanter ? Le débat est lancé

DAVID LAROUSSERIE

Jusqu'où iront les technologies d'intelligence artificielle (IA) dans leurs capacités à battre les humains ? Elles traduisent à la volée dans toutes les langues ou presque, écrasent des champions du monde au go, génèrent des textes, images et vidéos plus vrais que nature sur simple demande... Maintenant, elles se piquent de mathématiques, considérées comme l'un des sommets de l'intelligence. Cet été, plusieurs de ces systèmes ont obtenu l'équivalent d'une médaille d'or au concours des Olympiades internationales de mathématiques, qui réunissent les meilleurs jeunes de moins de 20 ans dans cet exercice. Cinq problèmes sur six ont été résolus par les IA, Gemini de Google, un logiciel d'OpenAI, mais aussi Seed-Praver de ByteDance (propriétaire du chinois Tik-Tok), ou Aristotle, de la start-up américaine Harmonic. Ces programmes ont seulement « séché » sur un exercice combinatoire de pavage d'un plan, résolu par quatre participants humains.

Certes, tous les lauréats de ce concours ne deviendront pas chercheurs en maths, et, inversement, tous les détenteurs d'une médaille Fields (une récompense prestigieuse de la discipline) n'ont pas obtenu l'or (un tiers l'ont cependant eu depuis

1990), mais le succès n'est pas anodin. Un an et demi auparavant, seuls les exercices de géométrie étaient accessibles aux IA, comme l'avait montré DeepMind avec son outil AlphaGeometry. Six mois plus tard, aux Olympiades 2024, une seconde version du logiciel, complété par AlphaProof, un système destiné à prouver des questions d'algèbre, d'analyse, de probabilités, etc., obtenait une médaille d'argent, résolvant quatre problèmes sur six. En 2025, un niveau de plus dans la hiérarchie est donc atteint. « Il y a trois ans, je n'aurais pas cru cela possible », estime le Britannique Timothy Gowers, professeur au Collège de France et à l'université de Cambridge. Ancienne médaille d'or et médaillé Fields (en 1998), il avait participé, l'année précédente, à la validation des solutions proposées par les outils de Google.

Les succès en maths ne se résument pas aux Olympiades. Sur l'un des tests les plus anciens, MATH, portant sur des problèmes de mathématiques de niveau lycée, les Gemini, GPT, Grok ou autres Qwen, DeepSeek ou Llama obtiennent des scores dépassant les 95 %, voire les 99 % de réussite. Un autre test, miniF2F, regroupant des problèmes plus difficiles, rencontrait en 2022 des taux de succès de 30 %. Ils sont supérieurs à 90 % aujourd'hui.

→ LIRE LA SUITE PAGES 4-5

## L'autre fléau qui décima l'armée de Napoléon fuyant la Russie

Des dents de grognards enterrés à Vilnius révèlent la présence de bactéries responsables de la fièvre paratyphoïde et de la fièvre récurrente

PAGE 2



## Une nouvelle étude sur la fin des dinosaures

Au moment de leur disparition, provoquée il y a 66 millions d'années par une météorite, les populations de dinosaures auraient été plus diversifiées qu'on ne le pensait

PAGE 3

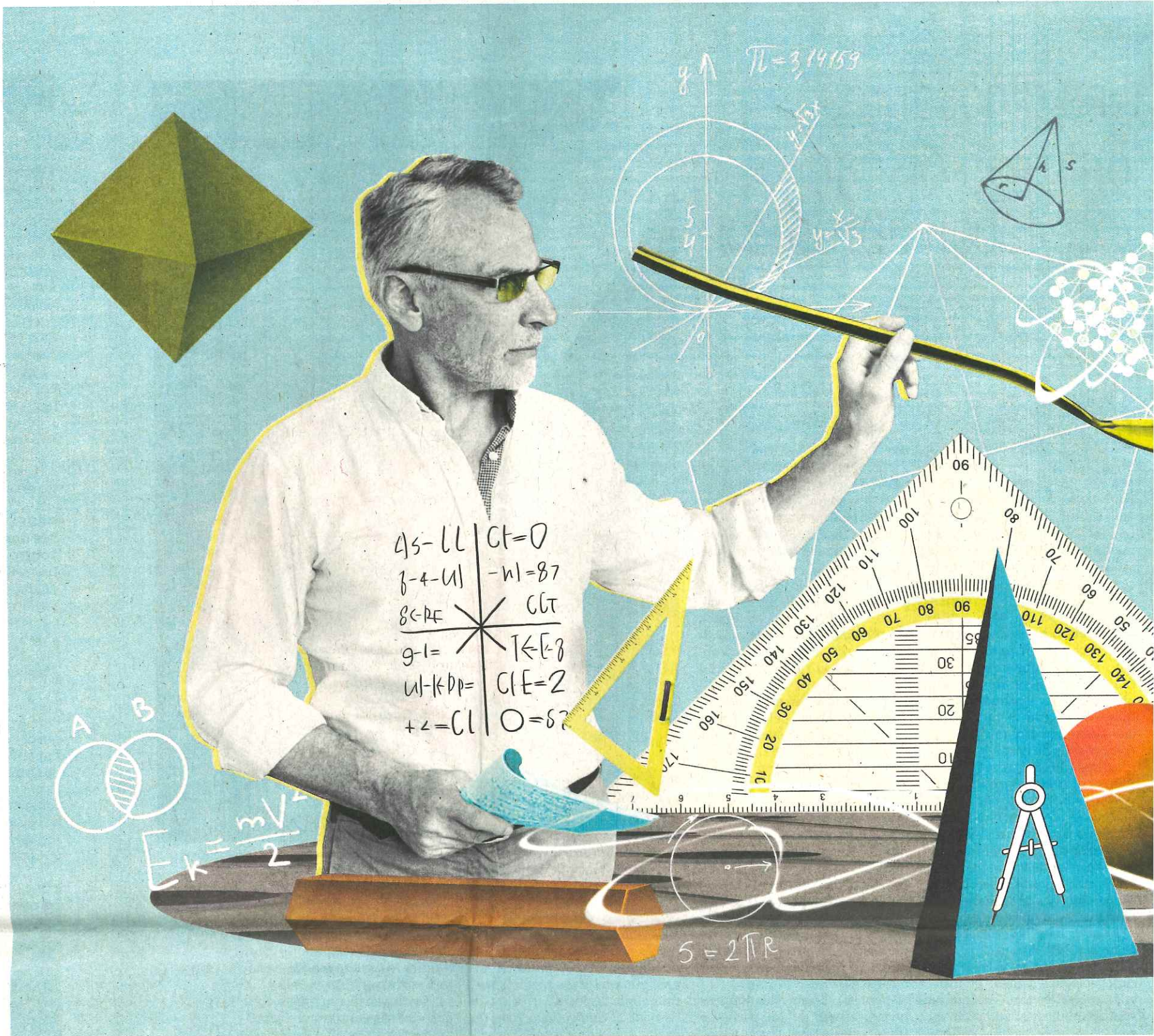


## Portrait Chantal Henry, une psychiatre des humeurs

En laboratoire et à l'hôpital, cette spécialiste innove dans le diagnostic et la prise en charge de la dépression et de la bipolarité

PAGE 8





# Mathématiques et IA : une équation féconde

► SUITE DE LA PREMIÈRE PAGE

Pour augmenter le niveau, Epoch AI, une association américaine créée pour étudier la trajectoire des IA, a lancé en 2024, soutenue par OpenAI, un autre test, FrontierMath, fait de problèmes originaux élaborés par une douzaine de mathématiciens de renom, dont le médaillé Fields 2006, Terence Tao, très impliqué dans le sujet de l'IA et des maths. A sa parution, la meilleure des IA résolvait 2 % des problèmes. Aujourd'hui, Gemini 2.5 Pro est à 29 % de succès sur les trois premiers niveaux et à moins de 10,5 % sur le niveau le plus difficile.

« Les entreprises de l'IA pensent que les maths sont une bonne manière d'évaluer l'intelligence de leurs modèles. La discipline leur sert un peu de vitrine. Cela leur permet aussi d'attirer les talents en leur sein », estime Amaury Hayat, professeur à l'Ecole nationale des ponts et chaussées, qui met au point un nouveau type de test. Il propose de sélectionner des articles de recherche en preprint (ou prépublication), c'est-à-dire non encore validés par les pairs, donc à la pointe de la recherche, et de mesurer si les IA arrivent à retrouver ces raisonnements. La concurrence est donc rude, nourrie par plusieurs percées.

Les IA contemporaines sont analogues à des réseaux de neurones artificiels, dont les connexions sont plus ou moins renforcées. Leur mise au point consiste à faire trouver les valeurs idéales de ces connexions en présentant au logiciel de nombreux exemples de couples question-réponse. Cela a fonctionné pour distinguer des chats de chiens dans des images à partir de 2012. Ou pour trouver le bon mot à la fin d'une phrase

et, de fil en aiguille, générer des textes complets, comme dans ChatGPT et ses descendants à partir de fin 2022. La difficulté en maths est qu'il n'y a pas des millions de problèmes, dans toutes les disciplines, dont on connaîtrait les bonnes réponses.

## Second « cerveau »

En janvier 2024, DeepMind, avec AlphaGeometry, pallie ce défaut en générant des données artificielles, c'est-à-dire en créant de nouvelles constructions géométriques. Il double cette capacité d'un second « cerveau » qui connaît les règles de la géométrie dans le plan et peut construire des démonstrations logiques. Vingt-cinq problèmes de géométrie sur 30 sont ainsi résolus. Quelques mois plus tard, le 24 juillet, l'association Project Numina remporte les Olympiades de l'IA, une nouvelle compétition mise au point spécifiquement pour défier les intelligences artificielles. Son secret est d'avoir constitué un corpus de 860 000 problèmes et leurs solutions, sur lequel des réseaux de neurones ont pu être entraînés.

Mais, le lendemain, DeepMind gagne une médaille d'argent aux « vraies » Olympiades avec une nouvelle idée, AlphaProof, en résolvant quatre problèmes sur six. Si les mathématiciens parlent et écrivent en langage naturel (avec certes des notations et des concepts ignorés du commun des mortels), ils disposent aussi, depuis les années 1980, d'autres langages, appelés « Lean », « Rocq » (anciennement Coq) ou « Isabelle », qui sont compris par les machines et qui permettent d'affirmer qu'un enchaînement d'opérations est valide. Ces langages, dits « formels », servent à certifier le fonctionnement de logiciels, mais aussi à vérifier des preuves.

AlphaProof bénéficie de ce langage particulier. Il a appris à écrire une bonne preuve en s'entraînant sur 100 millions de textes rédigés en Lean. Il utilise pour ce faire la technique de l'apprentissage par renforcement, qui récompense des neurones quand la réponse est bonne, et les pénalise dans le cas contraire. L'avantage des maths et de Lean est justement de savoir quand une réponse est bonne ou fautive, donc de pouvoir récompenser ou pénaliser avec justesse.

« L'intérêt pour ces outils, et Lean en particulier, augmente. Cela avait commencé avant l'IA, mais cela l'a accentué », témoigne Patrick Massot, professeur à l'université Paris-Saclay, qui vante les mérites de la formalisation pour la vérification bien sûr, mais aussi pour l'enseignement, la compréhension des mathématiques et l'aide au travail collectif et collaboratif. « Les maths au tableau, c'est la joie, mais il faut aussi de la rigueur, que donnent les machines et les langages formels », a déclaré Paul-André Melliès, chercheur CNRS à l'université Paris Cité, le mercredi 1<sup>er</sup> octobre, lors du lancement du projet européen Malinca, qui vise à « réduire le fossé entre le langage des mathématiciens et celui des machines ».

**« POUR LES ENTREPRISES DE L'IA, LES MATHS SONT UNE BONNE MANIÈRE D'ÉVALUER LEURS MODÈLES »**

AMAURY HAYAT  
PROFESSEUR À L'ÉCOLE NATIONALE  
DES PONTS ET CHAUSSEES

La start-up américaine Math, Inc. s'est félicitée durant l'été d'avoir formalisé automatiquement une démonstration importante de théorie des nombres, faite de 25 000 lignes de code. « Nous pensons multiplier par 100 ou 1 000 la quantité totale de lignes dans l'année qui vient », prévient son communiqué de presse.

A l'été 2024, une nouvelle voie semblait déjà tracée. Sauf qu'en septembre le monde des IA génératives est chamboulé par une innovation qui va offrir des idées inédites pour attaquer les maths. OpenAI sort son modèle o1, capable de « raisonner ». Il fait partie de la grande famille des IA génératives qui proposent, à la façon de ChatGPT, du texte répondant à une consigne. Mais, à la différence de son aîné, o1 « réfléchit avant de répondre ». C'est-à-dire qu'il apprend à découper une question en plusieurs étapes, qu'il génère et qu'il suit. Sa première « réponse » devient des consignes pour l'étape suivante, et ainsi de suite, jusqu'à une réponse finale donnée à l'utilisateur.

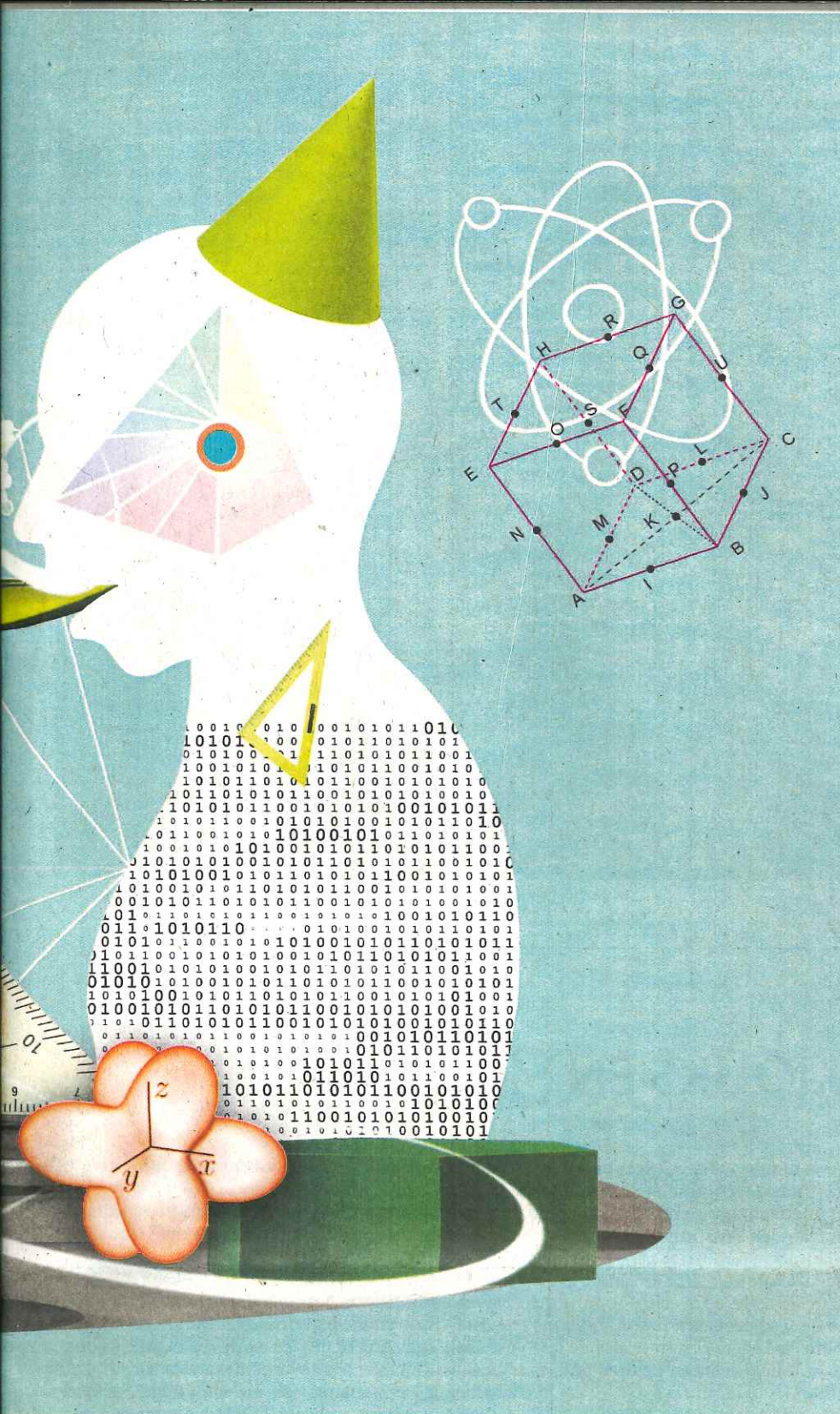
## Succès à nuancer

De premières équipes mélangent cette nouvelle capacité avec le langage formel de Lean. Kimina-Praver, de Numina, bat ainsi la concurrence, en avril, atteignant au test miniF2F le score impressionnant de 80 % de succès. Le système élabore un texte de preuve complet en Lean, donc vérifiable, et se construit ainsi des capacités à démontrer des exercices, grâce à l'apprentissage par renforcement. Quelques mois plus tard, Aristotile et Seed-Praver, sur des idées proches, sont médaillées d'or aux Olympiades.

Mais, surprise, au même moment, OpenAI et Google font aussi bien, mais sans passer par le langage formel de Lean. Ils entraînent directement leur modèle de « raisonnement » à apprendre comment découper une preuve, comment la vérifier, comment la rédiger... en langage naturel. Le chatbot sert en particulier à produire les textes et à les « juger ». C'est moins coûteux en calcul, et finalement tout aussi efficace. « C'est fou que ça marche avec seulement du langage naturel », s'étonne Yann Fleureau, cofondateur de Project Numina.

Ajoutons que les progrès récents s'expliquent aussi par de petits secrets bien gardés. Il est probable que certains calculs soient effectués par des programmes équivalant à des « calculatri-





ISABEL ESPANOL

## «IL MANQUE ENCORE QUELQUE CHOSE AUX INTELLIGENCES ARTIFICIELLES POUR PROGRESSER»

TIMOTHY GOWERS  
PROFESSEUR AU COLLÈGE DE FRANCE  
ET À L'UNIVERSITÉ DE CAMBRIDGE

Lors d'une conférence, Olivier Bousquet, qui a récemment quitté le laboratoire de Google à Zurich (Suisse), a présenté des résultats préliminaires montrant que jusqu'à 4 % des articles en topologie en preprint sur arXiv pouvaient déjà être formalisés automatiquement.

Le thème de l'automatisation des mathématiques se répand. Cet été, aux États-Unis, a été créé l'Institut de raisonnement assisté par ordinateur en mathématiques, doté de 6,6 millions de dollars (5,7 millions d'euros). L'Agence fédérale de recherche militaire, Darpa, a aussi débloqué des fonds pour, selon ses termes, «*expé-*

### Le «remplacement» des matheux

L'idée d'un «remplacement» des matheux eux-mêmes se diffuse, comme pour d'autres professions. Lors d'un colloque en septembre, le médaillé Fields 2018 Akshay Venkatesh a glacé l'assemblée en rapportant cette question d'un étudiant : «*Pensez-vous que les maths méritent d'être étudiées dans un monde où les machines peuvent répondre à tout pour vous ?*»

«*Il est certain que notre métier va changer*», dit Amaury Hayat. Cela a même déjà commencé. «*Récemment, j'ai trouvé, grâce à ChatGPT, la référence d'un article hors de mon domaine, mais qui contenait un résultat qui m'était utile*», raconte Timothy Gowers. Un moteur de recherche classique ne m'aurait pas aidé, car je n'avais pas les bons mots-clés, mais seulement l'idée générale. Le 2 octobre, dans un «pouet» (un message) posté sur le réseau social Mastodon, Terence Tao donnait un exemple supplémentaire d'un dialogue avec une IA «*qui [lui] a fait gagner du temps*».

La prudence est cependant de mise pour ces usages quotidiens. En octobre encore, des employés d'OpenAI ont proclamé que leur GPT-5 avait résolu dix problèmes du célèbre mathématicien hongrois Paul Erdős, connu pour avoir formulé énormément de conjectures. En réalité, le chatbot avait identifié les articles publiés ayant démontré ces questions, mais que l'auteur du site recensant ces problèmes n'avait pas trouvés... «*Ce sont des objets difficiles à maîtriser. Ils hallucinent, proposent de fausses solutions, livrent des explications non justifiées... Difficile d'écrire les bonnes consignes pour avoir quelque chose d'utile dans mon travail*», prévient Bartosz Naskrecki, professeur à l'université de Poznan (Pologne), qui, avec Ken Ono, a publié, dans *Nature Physics* du 6 octobre, une tribune prévoyant une «*transformation des pratiques dans la prochaine décennie*» et espérant que cela facilite «*le contrôle qualité et le travail collaboratif*». Il a aussi communiqué récemment sur sa découverte d'un résultat inconnu en maths grâce à l'IA. «*Ce n'est pas super spectaculaire, mais ça indique une tendance*».

Cette tendance est le nouvel horizon de cette quête : démontrer des problèmes sur lesquels les humains sèchent, donc faire d'authentiques découvertes scientifiques. Plusieurs avancées dans cette voie ont été enregistrées ces derniers mois. En décembre 2024, Amaury Hayat, avec deux collègues de Meta, a proposé une méthode

pour trouver des fonctions solutions d'un système d'équations compliqué à résoudre, même pour des humains. Il a opté pour une méthode d'apprentissage «*classique*» mais avec des données artificielles. «*Mes étudiants avaient environ 10 % de bonnes réponses, et moi, 25 %. Notre IA, 84 % !*» La technique a permis de trouver des solutions nouvelles. En février, une équipe de Caltech (Californie) a montré que des contre-exemples à une conjecture en théorie des groupes n'en étaient pas ! L'équipe a fait cette découverte en entraînant spécifiquement une IA, par apprentissage par renforcement, à explorer toute une gamme des possibles de ce problème, à la manière du joueur de go.

Tous ces progrès suscitent aussi bien des questions. Pourquoi les IA actuelles fonctionnent-elles ou non ? A quoi ressemblera une interaction entre elles et les professionnels ? Sommes-nous proches d'IA d'un niveau de chercheur en mathématiques ? Par certains côtés, les propriétés des IA évoquent la manière de penser des mathématiciens. Elles procèdent par exemple par essais et erreurs durant leurs phases d'apprentissage par renforcement. Elles sont capables aussi de trouver des motifs cachés dans de vastes ensembles de données ou d'explorer des mondes possibles plus rapidement qu'un humain.

Mais les professionnels trouvent unanimement que la créativité leur échappe encore. «*Les problèmes d'Olympiades demandent des raisonnements complexes sur des objets simples. La recherche en maths requiert, elle, des raisonnements moins complexes, mais sur des objets abstraits*», constate Yann Fleureau. Toute la question est de savoir si les modèles de langue seront bons dans l'abstraction. Timothy Gowers en doute. «*Il leur manque encore quelque chose et il faudra au moins une nouvelle idée dans leur architecture pour progresser*».

C'est un peu le message de conclusion de l'exposé de Stanislas Dehaene, titulaire de la chaire Psychologie cognitive expérimentale, au colloque de rentrée du Collège de France, le 16 octobre. Dans le cerveau, il identifie plusieurs langages, sémantique, numérique, musical, mathématique... «*Les réseaux de neurones actuels n'ont pas l'air bons pour le langage mathématique. Un autre type de réseaux sera nécessaire*». Il appuyait ses dires par plusieurs exemples, soulignés par des confrères, d'erreurs des IA, qui ont bien fait rire la salle.

D'autres sont plus optimistes, prévoyant des transformations du métier. Terence Tao, dans *Scientific American*, en juin 2024, imagine un monde où les IA pourraient aider à diviser le travail en plusieurs tâches et où «*nous pourrions commencer à prouver des centaines ou des milliers de théorèmes en même temps*». Yann Fleureau développe une vision assez proche, en déléguant aux IA «*des tâches pénibles à faire mais assez fréquentes, comme les démonstrations de lemmes [petits théorèmes] un peu techniques*».

Timothy Gowers, qui réfléchit à ces questions depuis fort longtemps, rêve d'autre chose. «*Il faut comprendre les processus humains de résolution de problèmes et de recherche pour améliorer éventuellement les ordinateurs*», explique celui qui constitue une base de données recensant les multiples étapes d'un raisonnement en train de se faire dans la tête d'un chercheur et différentes de celles écrites dans un article. Lui-même se filme en train de réfléchir et de parler pour ne rien rater des moments-clés. «*Etre conscient de l'importance de comprendre comment trouver des preuves m'aide à trouver des preuves !*», dit-il avec un sourire. Les futures IA lorgnent déjà sa base de données pour s'entraîner. ■

DAVID LAROUSSERIE

ces» (le chatbot reconnaît qu'il doit les utiliser). Les derniers chatbots sont aussi un «mélange d'experts», c'est-à-dire en réalité plusieurs logiciels plus ou moins spécialisés ou performants, activés selon les besoins.

Pour impressionnants que soient ces succès, ils sont à nuancer. «*Les Olympiades portent sur une partie limitée des maths. Il existe aussi beaucoup d'exercices et d'annales, car plusieurs pays ont des programmes d'entraînement dédiés, ce qui produit des données utiles pour les IA*», note Fabian Glöckle, en thèse avec Amaury Hayat et Meta, pour améliorer notamment l'entraînement des modèles, et qui avoue tout de même avoir été «*surpris que ça arrive si vite*».

En outre, il est connu que les réponses sont variables : tantôt la réponse est bonne, tantôt elle est fautive... Plusieurs passages sur le même problème sont donc nécessaires pour obtenir de meilleures réponses. Enfin, les failles sont encore légion, comme l'a montré Epoch AI dans son rapport d'octobre sur Gemini 2.5, qui, certes, brille aux tests, mais s'est trompé pour un exercice d'école primaire...

Il n'empêche, dans les laboratoires, les esprits s'échauffent. «*Certains peuvent même avoir peur*», estime Amaury Hayat. Il cite des collègues, spécialistes de formalisation, qui vont faire face à des machines capables de traduire automatiquement des énoncés naturels en Lean.

# UNE DISCIPLINE À LA RESCOUSSE DES ALGORITHMES

Dans les relations entre mathématiques et intelligence artificielle (IA), l'idée que la seconde peut aider à résoudre des problèmes des premières est souvent la plus visible, comme le montrent les attributions de médaille d'or aux IA lors des Olympiades internationales de mathématiques. Mais l'inverse est vrai aussi : les maths peuvent aider les IA. Après tout, ce n'est que justice, car ces dernières sont bâties avec des outils et des méthodes venus de ce champ, notamment les statistiques.

Et cette branche est toujours très vivace. «*Au début, quand nous avons vu arriver les méthodes d'IA appliquées à nos domaines, comme l'imagerie médicale, on pensait que nous n'aurions plus rien à faire. Il y a eu des moments de doute, se souvient Emilie Chouzenoux, directrice de recherche à l'Inria Saclay et membre du*

Centre de vision numérique, spécialiste de méthodes d'optimisation et du traitement des images. *Mais on a vite compris qu'on avait toujours besoin de maths !*»

La liste est longue de ce que la discipline peut apporter aux IA. D'abord comprendre leur fonctionnement ou quantifier les incertitudes de leurs réponses, ou bien optimiser leur entraînement, ou encore les rendre plus fiables en «*débiaisant*» leurs résultats, ou plus frugales, en diminuant les temps de calcul et les consommations électriques. Elles aident aussi à démêler les corrélations permettant d'identifier les causes et les effets...

«*On ne peut pas tout lister, sans crainte de vexer certains*», s'amuse Marianne Clausel, professeure à l'université de Lorraine, spécialiste de l'apprentissage statistique, la grande famille à laquelle appartiennent les chat-

bots à succès. Avec sa collègue, elle a lancé, depuis fin 2023, une série de conférences trimestrielles pour échanger en France sur ces questions de mathématiques et d'IA. L'objectif est surtout de «*rassembler une communauté diverse et dispersée*» et de «*rendre visible ce thème auprès des jeunes*». La prochaine sera sur les maths des IA génératives.

«*On ne peut pas se passer des maths. Elles répondent au pourquoi, elles identifient les facteurs-clés de la performance*», précise Marianne Clausel. C'est la solution pour ce qu'il est convenu d'appeler l'«*IA de confiance*», garantissant les résultats, assurant leur résistance aux «*attaques*» ou perturbations, promettant l'explicabilité des conclusions...

«*Cette communauté qui étudie l'IA n'aime pas les boîtes noires et veut soulever le capot*», complète Emilie Chouzenoux. Outre les

conférences qu'elles organisent, elles participent, avec de nombreux autres spécialistes, au volet maths du Programme et équipement prioritaire de recherche IA, financé par France 2030 depuis 2024 pour 73 millions d'euros sur six ans. Ou à l'IA Cluster, lancé en 2023, pour 500 millions d'euros.

### Apports fructueux

Les exemples d'apports fructueux sont déjà nombreux. L'une des vedettes de cette branche est Stéphane Mallat, professeur au Collège de France et médaille d'or 2025 du CNRS, qui, depuis son retour à la vie académique après une expérience dans une start-up d'imagerie, tente d'«*entrer à l'intérieur des IA*», comme il l'a expliqué dans sa conférence au colloque de rentrée du Collège de France, le 16 octobre. Il a ainsi contribué à montrer des analogies entre l'architecture de certains

réseaux de neurones et le cortex visuel. Il a aussi montré comment ces IA «*capturent*» des lois de la physique et de la chimie, sans que les chercheurs leur fournissent les équations. «*Les IA ne sont pas des perroquets stochastiques*», a-t-il déclaré en conclusion de sa conférence, en faisant allusion à une célèbre critique des IA selon laquelle elles ne sont capables que de recracher des textes ou images déjà vus. «*Elles apprennent des connaissances sophistiquées*», précise-t-il.

Emilie Chouzenoux a, quant à elle, montré comment modifier des méthodes d'apprentissage en imagerie pour éviter que des systèmes fonctionnant pour les données d'un hôpital ne marchent plus sur celles d'un autre, comme cela avait été constaté avec des IA standards. Marianne Clausel a breveté une méthode permettant d'évaluer l'état de

santé des batteries industrielles à partir de mesures de leurs cycles de charge et décharge, avec l'estimation de la marge d'incertitude.

Les deux chercheuses citent aussi les travaux d'un de leurs collègues, Gabriel Peyré, qui utilise des outils d'un domaine fondamental des maths, le transport optimal, en génomique ou en neurosciences. Le transport optimal permet aussi de «*débiaiser*» des résultats en déplaçant les distributions de probabilité pour rééquilibrer les résultats. Il est en effet établi que les IA reproduisent les biais présents dans leurs données, comme le genre, l'origine géographique... En revanche, les deux chercheuses n'utilisent guère les chatbots les plus populaires dans leur quotidien. «*Sans doute qu'en étant du domaine on mythifie moins ces outils que nos collègues !*» ■

D. L.