

Le bulletin algorithmique
de **Charlotte Mauger**



IA évaluée en maths

peut mieux faire !

Dix problèmes de maths inédits ont été posés à des modèles d'IA. Temps imparti pour les résoudre : une semaine. Un groupe de mathématiciens chevronnés a ramassé les copies... Verdict ?

La scène a de quoi faire sourire : des mathématiciens, stylo en main, en train de corriger la copie d'une IA ? Elle est pourtant tout à fait sérieuse : les performances des modèles sont si galopantes – à première vue –, que dix chercheurs ont décidé d'évaluer pour de bon leur niveau en maths. Avec une idée derrière la tête : *“On veut vraiment comprendre comment l'IA peut nous aider dans nos recherches”*, soutient Sergei Gukov, du California Institute of Technology. Ce projet intitulé First Proof a ouvert une session d'examens en février dernier et, déjà, les premières notes sont tombées.

Disons-le d'emblée : la majorité d'entre nous n'obtiendrait que 1 ou 2 sur 20... pour acte de présence. En fait, les IA ont déjà, et de loin, dépassé le grand public en mathématiques, discipline pourtant profondément ancrée au raisonnement humain. Ses débuts furent certes timides, cantonnés à la statistique et à l'analyse de données. →

JUIN | 2026 | **sv** | 103

ILLUSTRATION SOPHIE LECLERC D'APRÈS SHUTTERSTOCK - M. TUMELAIRE

Les dossiers **Algorithmes**

Mais en 2024, les murs ont tremblé : les modèles d'IA ont réussi plusieurs exercices de l'Olympiade internationale des mathématiques, une compétition assez difficile destinée aux collégiens et lycéens.

Ce n'était que le début. *"À la fin 2025, les progrès sont devenus spectaculaires !"* commente Sergei Gukov. L'IA générative s'est alors mise à plancher sur de vrais problèmes : des chercheurs à travers le monde lui ont soumis une cinquantaine de questions posées par le mathématicien hongrois Paul Erdős (1913-1996). Et elle en a résolu certains entièrement. De quoi sérieusement titiller la communauté mathématique...

Car entre résoudre quelques problèmes et contribuer activement à faire avancer la recherche, il y a loin ! Or se fier aux entreprises qui produisent les modèles pour connaître leur niveau en maths paraît malavisé : elles ont trop intérêt à surestimer la moyenne. D'ailleurs, les IA font-elles réellement des maths ou trouvent-elles des indices en scannant le Net, ce qui leur permet d'aligner bêtement une réponse juste, sans rien piger à la théorie ? Seul un test réalisé dans des conditions d'examen permettrait de le vérifier : c'est exactement ce que Rachel Ward, de l'université du Texas à Austin, et Mohammed Abouzaid, de l'université Stanford, toutes deux aux États-Unis, ainsi que huit autres experts, leur ont fait passer.

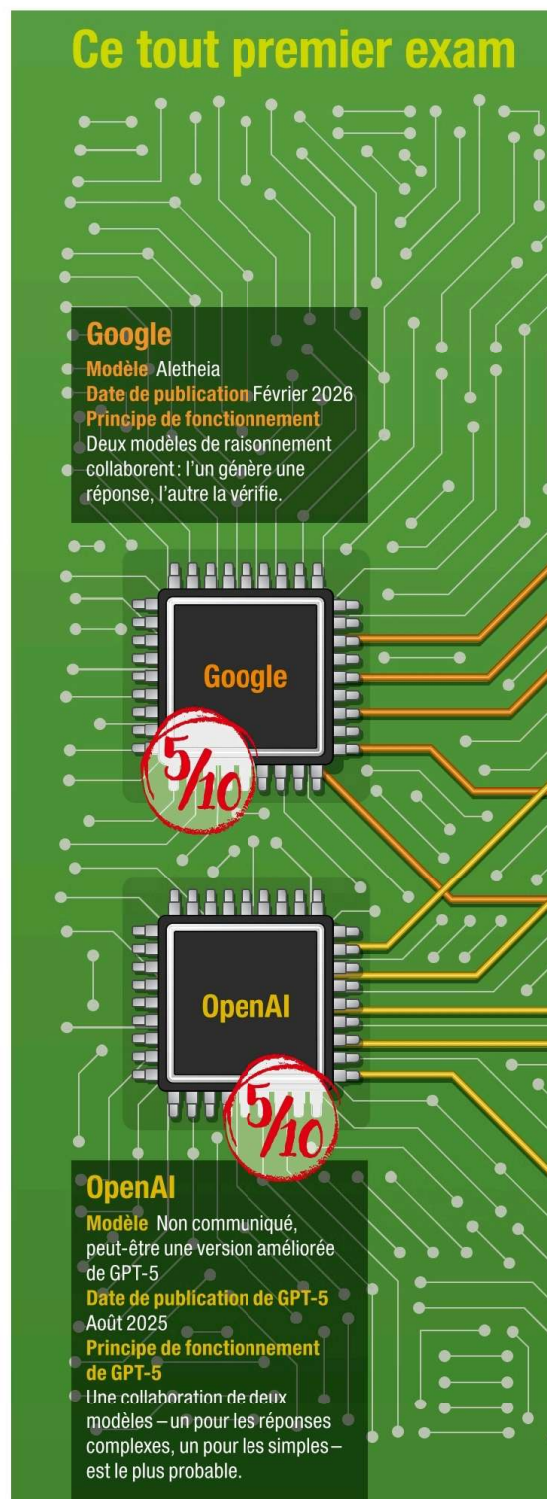
PRÉVENIR LA TRICHE

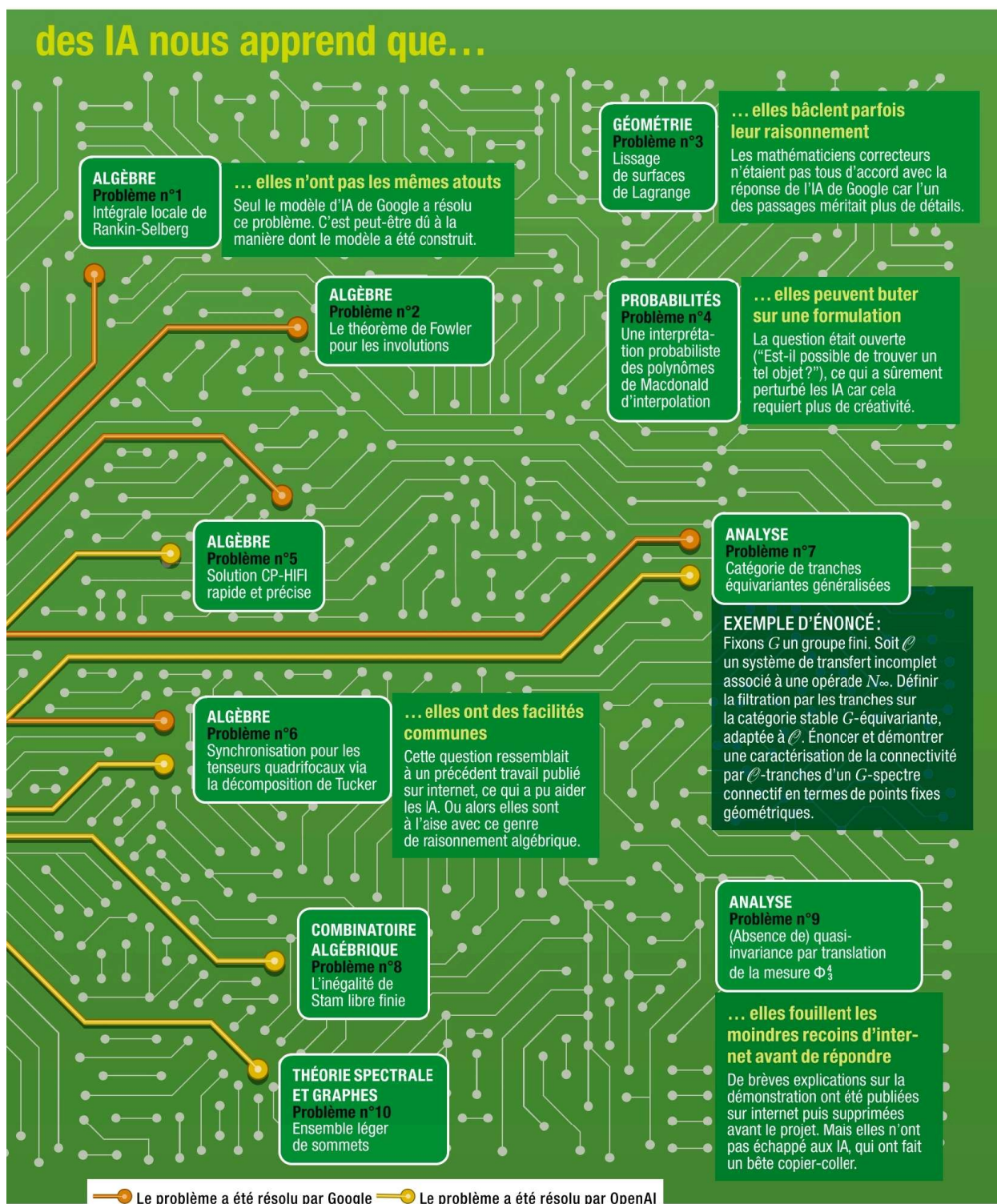
Le concept est simple : un panel de questions est soumis à des IA, du temps leur est accordé pour y répondre, puis des mathématiciens corrigent et évaluent leurs copies. Reste à définir les problèmes qui, pour éviter toute triche, ne doivent pas exister sur le Net. *"Les modèles d'IA sont vraiment bons pour déterrer ce type d'informations"*, souligne Tamara Kolda, mathématicienne et membre du projet. Chacun des dix membres de First Proof, qui proviennent d'horizons variés (algèbre, topologie, analyse...), a donc élaboré un exercice à partir de ses propres travaux encore confidentiels. De

Les démonstrations des agents artificiels sont parfois plus longues et complexes que nécessaire



LAUREN WILLIAMS
Professeure de mathématiques à l'université Harvard (États-Unis)





Les dossiers **Algorithmes**

quoi obtenir un examen final de 10 questions inédites, faisant appel à différentes branches de la discipline, d'un niveau *"ni trop dur ni trop facile"*, précise Tamara Kolda. C'est-à-dire assez simple pour les spécialistes, pas pour les autres mathématiciens : plus concrètement, les chercheurs devaient trouver une démonstration de cinq pages ou moins, pour éviter un devoir trop difficile.

Pourquoi ne pas les confronter directement à du dur, à la conjecture de Riemann par exemple, ce fameux problème resté insoluble depuis son énoncé en 1859 ? *"Il est quasiment sûr que les modèles n'y seraient pas arrivés, ce qui ne nous aurait rien appris !"*, répond Lauren Williams, professeure à l'université Harvard (États-Unis). Vu les premières notes obtenues, mieux vaut y aller progressivement, en effet.

Les 10 colles à destination des IA ont été rendues publiques le 5 février. Pendant une semaine, tout un chacun était invité à mettre au défi son IA préférée. Seuls deux géants de la tech américaine ont osé, Google DeepMind et OpenAI. Ils ont ensuite publié les réponses de leurs modèles, qu'il fallait encore corriger. Un groupe de mathématiciens chevronnés réunis dans un canal en ligne s'en est chargé, non sans difficulté. *"Parfois, le modèle produit une réponse qui semble correcte, mais quand on regarde dans le détail, on voit des erreurs ou des oublis. Cela demande beaucoup de temps pour vérifier, surtout que les*

démonstrations des agents artificiels sont parfois plus longues et complexes que nécessaire", témoigne Lauren Williams.

Verdict ? Google a répondu correctement à 5 ou 6 questions – l'une de ses réponses ne fait pas l'unanimité – ; OpenAI, à 5. Deux des colles sont restées hors d'atteinte pour les modèles : l'une touchant à l'analyse et l'autre aux probabilités. *"Il semble qu'à l'inverse, certains domaines comme la théorie des nombres, la combinatoire ou la théorie des groupes leur soient plus accessibles"*, estime Sergei Gukov, qui a travaillé avec Google sur le projet. *Mais nous manquons encore de données pour en être absolument certains et pour en comprendre les raisons."*

À PEINE LA MOYENNE

Peut-on en tirer quelques enseignements sur le potentiel des IA en maths ? Les avis sont partagés. *"Selon moi, si un modèle de langage arrive à produire un résultat, il est quasi sûr qu'il a déjà été publié quelque part"*, balaie Tamara Kolda. L'appréciation de Sergei Gukov est plus encourageante : *"L'IA fait de grands progrès. C'est ce que l'on constate dans la recherche, où de plus en plus d'articles sont produits en collaboration avec des modèles d'IA."* Oui, enfin... Que ce soit par un pur raisonnement ou un bête copier-coller, toujours est-il qu'ils atteignent à peine la moyenne... Alors que c'était un devoir *"à la maison"* : les entreprises ont testé elles-mêmes leurs modèles – le bruit court que des mathématiciens embauchés pour l'occasion leur auraient donné un coup de pouce, même si Google nous affirme que ce n'est pas son cas.

Qu'à cela ne tienne, les prochains examens seront mieux contrôlés. Vu l'enthousiasme suscité par ce projet hors norme, First Proof entend réitérer. *"Une des véritables différences pour la prochaine fois est que nous poserons nous-mêmes la question aux modèles"*, fait savoir Lauren Williams. *Puis des mathématiciens embauchés pour l'occasion se chargeront de noter les productions."* Les questions du contrôle suivant sont d'ores et déjà en train d'être recensées. En attendant : *"5/10. Peut mieux faire. À retourner signé par vos fabricants."*

C'est quoi, la méthode de l'IA pour donner une réponse ?

Elle est loin, l'époque des *"perroquets stochastiques"*, qui ne faisaient qu'aligner dans une phrase les mots ayant la meilleure probabilité d'apparaître ! Aujourd'hui, les modèles d'IA imitent notre manière de réfléchir avant de répondre : ils élaborent une stratégie étape par étape, puis rédigent. Aletheia, utilisé par Google DeepMind, organise même un dialogue entre deux modèles : l'un propose une réponse, l'autre la teste ; il reconnaît aussi son échec en envoyant *"pas de solution trouvée"*. Les progrès sont donc là : *"Avant fin 2025, je ne recommandais pas l'IA pour le code"*, explique Sergei Gukov. *Cette époque est révolue, et je pense que les mathématiques s'apprentent à vivre un basculement similaire."*