## Traduction de la conférence de Geordie Williamson à l'ICM 2026

Bienvenue. C'est un grand plaisir pour moi de présenter la conférence publique d'aujourd'hui donnée par le professeur Geordie Williamson. Il est professeur de mathématiques à l'Université de Sydney et également cofondateur et directeur du *Sydney Mathematical Research Institute*. Le professeur Williamson est reconnu pour ses nombreuses contributions dans le domaine de la théorie des représentations, où ses travaux ont résolu des problèmes et des conjectures remontant aux années 1980, voire plus tôt. Il est également un pionnier dans l'utilisation de l'intelligence artificielle et des outils d'apprentissage automatique pour la recherche en mathématiques pures, en particulier à travers sa collaboration avec DeepMind. En 2016, il a reçu le prix *New Horizons* et le prix de recherche du *Clay Mathematics Institute*. En 2018, il a été élu membre de la *Royal Society* et a également prononcé une allocution plénière au Congrès international des mathématiciens (ICM), qui s'est tenu cette année-là à Rio de Janeiro. En 2024, il a reçu le prix de recherche Max-Planck-Humboldt. Je l'ai personnellement rencontré pour la première fois en 2009, et je ne peux sous-estimer l'impact que ses idées ont eu sur mes propres recherches. Veuillez vous joindre à moi pour accueillir le professeur Williamson sur scène.

Merci.

Alors, merci beaucoup. C'est un honneur et un plaisir de m'adresser à vous aujourd'hui.

Lorsque l'on m'a demandé de donner une conférence sur l'ia et la découverte mathématique, une citation m'est immédiatement venue à l'esprit.

> "*Il y a une intemporalité sous-jacente dans cette conversation fondamentale que sont les mathématiques.*"

J'adore le fait qu'une idée exprimée il y a 200 ans soit encore pertinente et importante pour nous aujourd'hui, ou même il y a 2 000 ans. Et j'adore les bouffées d'enthousiasme dans cette conversation lorsqu'une idée nouvelle arrive.

Et j'adore aussi les longues périodes de silence où nous ne savons que dire, où nous ne savons que faire, ou peut-être où nous digérons ce qui a été dit.

Et aujourd'hui, je parle de l'impact de l'ia sur cette longue conversation.

Mais je veux aussi essayer de souligner que les racines de l'ia font également partie de cette longue conversation.

Ainsi, Alan Turing a été l'un des plus grands mathématiciens et penseurs du XXe siècle. Vous savez, il a été fortement impliqué dans le projet Enigma pour décrypter la cryptographie allemande pendant la Seconde Guerre mondiale. Et immédiatement après la Seconde Guerre mondiale, il a eu accès à l'un des tout premiers ordinateurs électroniques, et très peu de temps après, il a commencé à réfléchir à la possibilité de l'intelligence artificielle. Il a rédigé le premier article sur l'intelligence artificielle, intitulé *Intelligent Machinery* (Machines intelligentes). Vous savez, j'ai été stupéfait

lorsque j'ai lu cet article pour la première fois, et je le suis à chaque fois que je le relis. Il y présente la première esquisse d'un réseau de neurones rudimentaire qu'il appelle une "machine non organisée", et il discute également de domaines qui pourraient être explorés utilement avec l'ia ou servir de cas tests pour l'ia, à savoir : les jeux, les langages, la cryptographie et les mathématiques. Je trouve toujours cette liste tout à fait extraordinaire.

Passons maintenant à 1999.

Eh bien, c'était l'année suivant l'arrivée dans notre famille d'une petite boîte bizarre qui faisait du bruit et nous donnait Internet.

Dans cet article, Gowers est invité à spéculer sur les 100 prochaines années des mathématiques, et il imagine une discussion entre un mathématicien et un ordinateur. Il ne précise pas quand cette discussion aura lieu, et il suggère également qu'elle pourrait se dérouler dans une sorte de langage formel, ressemblant davantage à quelque chose comme Lean. Voici à quoi ressemble cette discussion. Il n'est pas nécessaire que vous compreniez entièrement cet échange, mais le mathématicien dit par exemple : "Le résultat suivant est-il vrai : soit $\delta$ supérieur à zéro, bla, bla, bla ?" Et l'ordinateur répond quelque chose d'absurde. Le mathématicien dit : "Non, non." Le mathématicien formule autre chose qui n'est pas utile. Le mathématicien dit : "Cela ne m'aide en rien." Et puis l'ordinateur commence réellement à être utile au mathématicien. Et tous deux convergent vers une compréhension du théorème de Ross.

J'ai été stupéfait lorsque j'ai vu cela pour la première fois, car je pense que lorsque les agents conversationnels (*chatbots*) ont fait leur apparition, beaucoup de mathématiciens, moi y compris, ont en quelque sorte rêvé d'une telle interaction. Mais nos expériences initiales ont montré qu'ils n'étaient pas très utiles.

Pourtant, cette année même, je crois que nombre d'entre nous commencent à avoir exactement ces conversations utiles, tout comme Gowers l'avait prédit il y a 27 ans.

Il poursuit dans cette section en affirmant qu'il pourrait y avoir un âge d'or où les ordinateurs seraient bons pour les exercices mais pas encore doués pour avoir des intuitions profondes.

Et beaucoup d'entre nous, moi y compris, ont le sentiment que nous vivons précisément cet âge d'or en ce moment-même. Mais de manière quelque peu décourageante, il ajoute que cet âge d'or, s'il se produit, ne durera probablement pas longtemps.

Je pense que ces deux citations résument bien ce que je ressens actuellement.

C'est de cela que j'aimerais parler aujourd'hui. Il y a ce sentiment d'émerveillement : c'est extraordinaire que ces systèmes fonctionnent, et les idées qui les sous-tendent ne sont pas si compliquées.

De plus, je suis quelqu'un qui adore trifouiller des programmes informatiques et les utiliser pour expérimenter dans ma vie mathématique. Le fait de pouvoir parler à un ordinateur en anglais est, au sens littéral, un rêve devenu réalité.

Mais d'un autre côté, je suis très préoccupé par ce qui se passe actuellement pour la prochaine génération de mathématiciens. Des étudiants en thèse me demandent : "Dois-je faire un doctorat en mathématiques ?" Je pense que la réponse est oui, mais le simple fait que la question soit posée constitue un défi.

De surcroît, tout va si vite, et je tiens à discuter de ces inquiétudes car je crois aussi que nous avons une certaine capacité d'action en tant que communauté.

Mais d'abord, l'émerveillement. Je souhaite tout d'abord décrire les idées simples et magnifiques qui sous-tendent les agents conversationnels et leurs origines, avant d'aborder certaines applications intéressantes en mathématiques.

Remontons à 1948.

C'est la même année où Turing a écrit *Intelligent Machinery* et où Shannon a publié ce que beaucoup considèrent comme l'un des articles les plus influents du XXe siècle : *A Mathematical Theory of Communication* (Une théorie mathématique de la communication). Il s'agit de la section 3. Il s'y donne pour tâche d'approximer l'anglais, de construire des approximations de l'anglais.

Pour ce faire, il prend d'abord 27 symboles : les 26 lettres de l'alphabet plus le symbole d'espace, et il commence à faire des échantillonnages de différentes manières. La première chose qu'il fait est d'échantillonner de manière purement aléatoire, ce qu'il appelle une approximation d'ordre zéro. Ensuite, il échantillonne selon les fréquences d'apparition des symboles en anglais : c'est l'approximation de premier ordre. Puis il échantillonne chaque symbole suivant selon la distribution de probabilité conditionnelle sachant le symbole qui vient d'être échantillonné, ce qui correspond à une structure de bigrammes. Il fait de même en fonction des deux derniers symboles, ce sont les trigrammes. Enfin, il passe aux mots.

Il réalise d'abord un échantillonnage avec la même fréquence de mots qu'en anglais, puis une approximation par bigrammes. Si l'on observe ces résultats, ils ressemblent à une approximation graduelle de l'anglais.

Voici ce qu'il en dit : "La ressemblance avec un texte en anglais ordinaire augmente de manière tout à fait perceptible à chacune des étapes ci-dessus.". En langage moderne, ce qu'il fait est de la prédiction du prochain jeton (*next token prediction*).

Les jetons (*tokens*) sont grosso modo des mots, ou plus précisément des morceaux de mots, des signes de ponctuation, etc.

La tâche consiste à prédire le jeton suivant à partir d'une chaîne de jetons potentiellement longue.

Une façon de concevoir cela est d'imaginer que vous lisez un livre ou que vous naviguez sur Internet, et que votre univers se résume aux jetons que vous lisez. Ce que vous essayez alors de faire, c'est de prédire le futur.

Remarquez que si vous pouvez prédire le futur à un pas de distance avec une certaine fiabilité, vous pouvez continuer à essayer de prédire la suite. Vous pouvez, par exemple, deviner la fin d'une histoire ou quelque chose de semblable, et commencer à explorer des trajectoires possibles dans le futur.

Dans la prédiction du prochain jeton, nous essayons donc de prédire le futur. Je veux vous donner deux exemples de prédiction du prochain jeton. Voici le premier, un extrait de ce que je considère comme le début de livre de maths le plus charmant qui soit : *Philosophy and Fun of Algebra* de Mary Boole. C'est un livre absolument magnifique.

Voici comment il commence : "Mes chers enfants, un jeune singe nommé Génie cueillit une noix verte et, à travers sa coque amère, mordit dans le fruit dur. Il..." et votre tâche est de prédire le prochain jeton.

Si vous connaissez l'anglais, vous devinez probablement qu'un verbe vient ensuite.

Mais nous sommes au début du livre. Vous ne savez rien de ce singe nommé Génie, vous ne pouvez donc pas vraiment prédire ce qu'il va faire.

Je pense que c'est l'image que beaucoup de gens se font de la prédiction du prochain jeton : une sorte de jeu statistique.

Je veux vous donner un autre exemple pour vous convaincre qu'il y a beaucoup plus dans la prédiction du prochain jeton que cela.

Il est tiré d'un livre qui m'a fait beaucoup souffrir en tant qu'étudiant diplômé : le *Algebraic Geometry* de Robin Hartshorne, à la page 312.

Votre tâche consiste maintenant à prédire le jeton que j'ai omis.

Si vous êtes mathématicien, vous regardez cela et vous pensez - "mais pas à ..." - qu'il est très probable qu'un symbole manque. Ainsi, d'après le contexte de ce passage, vous pouvez deviner qu'un symbole est manquant.

Mais je soutiens que même si vous lisez tout le paragraphe, vous ne pouvez pas deviner quel symbole manque. En fait, pour deviner quel symbole il s'agit, vous devez comprendre la démonstration.

De plus, si vous comprenez la démonstration, vous pouvez être sûr à 99,9 % de l'identité de ce symbole, et vous aurez raison, à moins qu'il n'y ait une coquille chez Hartshorne.

Cet exemple vise à illustrer qu'une quantité arbitraire de compréhension peut être dissimulée dans la prédiction du prochain jeton.

La prédiction du prochain jeton a donc l'air, au premier abord, d'un petit jeu amusant, mais c'est en réalité un phénomène très, très profond.

Shannon s'intéressait à la prédiction du prochain jeton parce qu'il s'intéressait à la mesure dans laquelle l'anglais peut être compressé. Il existe donc un lien entre prédiction et compression. Et j'espère vous avoir convaincu sur la diapositive précédente qu'il existe également un lien entre compression et compréhension.

La prédiction pourrait donc être une composante majeure de la façon dont nous comprenons le monde.

Pour vous donner quelques exemples frappants : en neurosciences, il existe le codage prédictif (*predictive coding*), selon lequel une grande partie de ce que nous pensons percevoir est en fait prédite, et nous percevons beaucoup moins de choses que nous ne le pensons.

Il y a aussi ce phénomène fou qui se produit en permanence dans les processeurs de nos ordinateurs à notre insu : le processeur devine l'avenir en effectuant une prédiction de branchement (*branch prediction*).

Et l'on trouve bien d'autres situations où la prédiction devient critique.

À la fin de la section 3 de son article de 1948, Claude Shannon écrit : "Il serait intéressant de construire des approximations supplémentaires, mais le travail impliqué devient énorme à l'étape suivante."

Il s'intéressait beaucoup à ce problème et a même mené des expériences avec sa femme, lui demandant de deviner un mot profondément enfoui dans un livre afin de démontrer qu'elle était bien plus capable de le faire qu'on ne l'aurait naïvement cru. Il cherchait à comprendre où se situait l'anglais entre le hasard total et la prédictibilité, s'intéressant ainsi à l'entropie de l'anglais.

Or, ce problème de la construction d'approximations de l'anglais est précisément celui que résolvent les réseaux de neurones dits *transformers*. Ils le résolvent en ce sens qu'ils parviennent à construire des approximations arbitrairement bonnes - ou du moins extrêmement bonnes - de la distribution statistique de l'anglais.

Qu'est-ce qu'un *transformer* ?

C'est une boîte qui reçoit des séquences de jetons à l'entrée et crache des prédictions à la sortie, sous la forme d'une liste de probabilités pour le jeton suivant.

Cette boîte est contrôlée par des paramètres. Pour les non-mathématiciens, on peut imaginer qu'elle comporte une multitude de boutons que l'on tourne : un petit ajustement de ces paramètres modifie légèrement les probabilités de sortie.

Ce *transformer* est donc entraîné. Au tout début, nous l'initialisons d'une certaine manière et il ne fournit que des prédictions aléatoires. L'entraînement consiste pour l'essentiel à ajuster les boutons pour rendre les prédictions un peu meilleures. Cet ajustement ne se fait pas au hasard, mais par le

calcul, répété un très grand nombre de fois.

C'est un processus automatique appelé entraînement par descente de gradient, qui produit un réseau de neurones entraîné. Il est également utile de garder à l'esprit qu'au cours de cet entraînement, on peut échantillonner le modèle à chaque étape pour observer la distribution qu'il a apprise.

Je dois dire qu'il se passe énormément de choses à l'intérieur, mais mes tentatives passées pour l'expliquer à un public général m'ont convaincu de ne pas m'y risquer ici.

Voici un exemple d'échantillons obtenus pendant l'entraînement :

À l'étape 10 000 : "*The blue of and into river that for light is on with time.*" (Il est très difficile de lire à haute voix des assemblages aléatoires rapidement !) Une certaine structure de l'anglais commence à apparaître.

Vers l'étape 50 000 : "*The river was in a blue light and the people were to the old house.*" Dans de nombreux cas, on ne peut plus distinguer cela de l'anglais authentique.

Ces types d'échantillons sont ceux que l'on observait au début des *transformers*, vers 2016. Aujourd'hui, les modèles modernes sont capables de cohérence et produisent un texte riche de sens sur des millions de jetons.

Dès lors, une question très naturelle se pose : à quel point sommes-nous proches de la distribution de l'anglais ? Comment mesurer la qualité de cette approximation ?

On peut mesurer cette distance à l'aide d'une métrique appelée entropie croisée (*cross-entropy*), qui évalue l'écart par rapport à la distribution de l'anglais.

C'est ainsi qu'ont été découvertes les lois d'échelle (*scaling laws*), un phénomène remarquable. Vers la fin des années 2010, les chercheurs entraînant des *transformers* ont constaté que les performances s'amélioraient à mesure que les modèles grandissaient. Mais la question était : à quel rythme ? Depuis les travaux de Shannon, de nombreuses stratégies avaient été proposées pour approcher l'anglais, mais elles finissaient toujours par stagner.

Or, on observe ces graphiques logarithmiques de lois d'échelle remarquables.

L'axe vertical représente, très grossièrement, la distance à l'anglais (axes horizontal et vertical en échelles logarithmiques). L'axe horizontal correspond pour l'essentiel à la quantité de calcul d'entraînement, à la quantité de données (la proportion d'Internet injectée dans le modèle) et au nombre de paramètres.

On observe alors une magnifique loi de puissance.

Ces résultats proviennent de Kaplan et al. (*Scaling Laws for Neural Language Models*), qui soulignaient : "Nous n'avons aucune compréhension théorique solide de la raison pour laquelle il en va

ainsi." Je tiens à insister sur le fait qu'il s'agit d'observations purement empiriques, et qu'il existe un vif débat sur leur exactitude littérale. Néanmoins, c'est un fait avéré : pour se rapprocher un peu plus de l'anglais, il suffit d'augmenter la puissance de calcul, les données et les paramètres. S'il y a un seul graphique qui explique l'investissement massif dans les centres de données, ce sont ces lois d'échelle, car elles garantissent en quelque sorte un résultat proportionnel à l'effort de calcul.

Je tiens à souligner qu'il reste beaucoup de mystères à l'intérieur de ces objets. On pourrait me dire : "Geordie, les *transformers* ont une forme rectangulaire." Ce qu'il y a à l'intérieur est fascinant, et à ma connaissance, il n'existe aucune dérivation mathématique expliquant pourquoi il doit en être ainsi.

De plus, cela réécrit de manière fascinante certaines hypothèses fondamentales des statistiques. En statistique, il existe le compromis biais-variance (*bias-variance trade-off*), qui stipule que pour tout modèle statistique, il existe un nombre optimal de paramètres. Ce principe est, selon moi, gravement remis en cause par les réseaux de neurones profonds.

Que se passe-t-il mécanistiquement à l'intérieur ?

Certaines choses sont comprises, mais beaucoup restent à comprendre. Pourquoi ces lois d'échelle tiennent-elles ? De nombreuses personnes se penchent sur ces questions fascinantes. Ce n'est pas qu'une simple curiosité intellectuelle : il arrive aujourd'hui que les *transformers* n'agissent pas comme nous le souhaiterions, et une meilleure compréhension de leur machinerie interne pourrait les amener à mieux se comporter.

Pour vous donner un aperçu de l'émerveillement que l'on ressent face à ces modèles, imaginez que l'on puisse pénétrer à l'intérieur d'un *transformer*. Tout s'y déroule dans un certain espace vectoriel, où une multitude de jetons associés à diverses données sont étiquetés de manière particulière (par exemple, des débuts de mots ou des signes de ponctuation, représentés par les couleurs sur l'image). Le jeton que j'ai mis en surbrillance est celui de la multiplication, parce que j'aime la multiplication.

On observe ici une géométrie émergente issue de l'entraînement, avec de nombreux aspects frappants que l'on comprend encore très peu.

Il y a là quelque chose de prodigieux qui nous échappe. J'aimerais maintenant expliquer en quoi ces outils sont réellement utiles pour les mathématiciens purs.

Avant d'en arriver là, il faut comprendre pourquoi ils sont parfois difficiles à manier.

Si l'on entraîne un tel *transformer* uniquement sur Internet, c'est un objet fort étrange : il croira à une théorie du complot avec la même probabilité que l'Internet y croit, et il vous insultera avec la même probabilité.

Pour le rendre utile, il existe toute une "magie noire" appelée le réglage fin (*fine-tuning*), qui s'efforce de rendre le réseau poli, de lui faire donner des réponses de la longueur adéquate ou d'adopter un style de discours particulier.

Le résultat de ce réglage fin est infiniment plus utile, et ces modèles parviennent parfois à imiter ce que j'appelle la trace de pensée (*thinking trace*) d'un mathématicien.

Lorsque l'on collabore sur un problème mathématique, bien que nous ignorions ce qui se passe précisément dans notre cerveau, on peut imaginer qu'un dialogue interne s'y déroule : "As-tu essayé ceci ? Et si on substituait cela ?" C'est une expérimentation. On peut imaginer consigner ce dialogue par écrit : cela forme une trace de pensée, qui peut être efficacement simulée.

Voici deux exemples concrets.

Le premier est un problème amusant sur lequel nous avons progressé grâce à un grand modèle de langage (llm).

De quoi s'agit-il ? Nous avons un grand graphe hautement structuré (pour les non-initiés, un graphe est la même chose qu'un réseau). Si vous êtes mathématicien, je vous ai décrit exactement de quoi il s'agit. Les carrés et les cubes y abondent. Pourtant, pour diverses raisons, on ne s'attend pas à y trouver des structures plus vastes, comme des cubes de dimension supérieure. Et pourtant, il y en a un. Ce cube m'a toujours surpris.

La question que nous nous sommes posée était : pouvons-nous trouver de tels graphes de manière générale ? Pouvons-nous trouver ces grands hypercubes en général ? Sans entrer dans les motivations profondes, c'est un problème de recherche extrêmement difficile. Supposons que vous vous y intéressiez : il y a un nombre colossal de sous-graphes dans ce graphe très complexe, et nous cherchons cela pour des dimensions $n = 100$ ou plus. C'est un problème de recherche immense et ardue.

Comment utilise-t-on un llm pour ce problème ? Nous avons employé une méthode appelée *FunSearch* ou *AlphaEvolve*.

Imaginez qu'un graphe d'intérêt possède une description structurée claire, que l'on peut exprimer via un programme Python.

Si vous demandez simplement au llm : "Voici mon graphe, peux-tu me donner un programme Python qui produit un sous-graphe intéressant ?", la plupart du temps, la réponse est inutile (le programme ne compile même pas, etc.).

Mais vous avez sans doute remarqué que sur les agents conversationnels, il y a un bouton "régénérer". Ce sont des modèles probabilistes, alors on peut cliquer sur régénérer... je l'ai fait 10 fois, 50 fois, avant d'être tenté d'abandonner.

Et puis, soudain, au milieu d'un flot de résultats aberrants, surgit quelque chose de tout à fait fascinant. L'idée est d'automatiser ce processus de régénération pour filtrer les trouvailles intéressantes. Si une construction pertinente apparaît, on peut dire au llm : "Peux-tu poursuivre dans cette voie ? Voici quelques exemples qui fonctionnent bien jusqu'à présent."

Telle est l'idée derrière l'algorithme *FunSearch/AlphaEvolve*.

Le système est configuré comme un chatbot. On lui donne une consigne : "Agis en expert en algèbre computationnelle et en optimisation combinatoire." Nous avons travaillé sur ce sujet avec Adam Wagner, qui a mené un nombre considérable de ces expériences et croit fermement à l'utilité d'un "coaching psychologique". Il dit toujours au modèle : "Tu vas y arriver, je crois en toi !"

On peut ainsi lire la trace de pensée du modèle. C'est assez remarquable. Au milieu de la nuit, le système écrit : "Je m'apprête à proposer quelque chose de vraiment extravagant, une manœuvre d'Ivan le Fou" - une référence au film *À la poursuite d'Octobre Rouge* et aux tactiques de sous-marins. La solution actuelle est bonne... puis on réalise qu'il s'agit d'une mauvaise transcription du mot hindi signifiant "tonnerre". "Devenons fous !"

Des choses vraiment bizarres surgissent dans ces traces de pensée. Il m'est aussi arrivé qu'un modèle discute de grands vins tout en essayant de résoudre un problème de mathématiques !

Le matin venu, on se réveille et l'on découvre un programme qui accomplit manifestement quelque chose de très, très intéressant.

Je tiens à insister sur le fait que ce domaine est soumis à un biais de publication : je ne vous parle que de ce qui a fonctionné, après de nombreuses tentatives infructueuses. Mais ce matin-là, nous disposions d'un programme effectuant une percée remarquable. C'est une situation pour le moins curieuse. Le code s'affiche à droite : on peut le vérifier, il fait incontestablement quelque chose d'inédit, mais il est extrêmement complexe et il faut décrypter ce qu'il réalise.

Ce n'est peut-être pas sans rappeler la situation où l'on obtient une preuve (qui compile peut-être dans Lean ou un autre assistant de preuve) et où l'on doit comprendre ce qui s'y passe réellement. Après avoir scruté le code pendant quelques jours, nous avons réalisé qu'il exploitait une idée issue de la théorie des nombres pour construire un hypercube de taille considérable.

C'est exactement ce que j'espérais de l'interaction entre les mathématiques et l'intelligence artificielle : qu'elle mette en lumière des objets que nous n'aurions jamais imaginés, enrichissant ainsi notre univers.

Pour les mathématiciens, voici la taille de l'hypercube obtenu. C'est un petit exemple, fort distrayant, mais loin d'être isolé : de nombreux cas similaires se sont produits ces neuf derniers mois.

Ceci faisait partie de diapositives que je préparais début mai, lorsque j'ai commencé à structurer cette conférence, pensant devoir vous convaincre de l'utilité des llms pour les mathématiciens. C'était le 1er mai.

J'aurais dû attendre un mois de plus ! Peu après, on a assisté à l'annonce spectaculaire d'un contre-exemple à la conjecture des distances unitaires d'Erdős, un problème ouvert majeur de géométrie combinatoire sur lequel de nombreux chercheurs peinaient depuis 80 ans.

Plus choquant encore, juste après la finale de la Coupe du monde, nous avons découvert un tweet de Levent Alpöge (et Ralph Furman), chercheur chez Anthropic et brillant mathématicien, qui, en guidant le modèle *Fable*, a réussi à trouver un contre-exemple à la conjecture jacobienne !

Nous n'avons guère plus de détails pour l'instant, si ce n'est ce tweet, et nous supposons qu'il dit vrai. La situation est délicate car ces laboratoires demeurent extrêmement secrets. Mais l'afflux simultané de tels exemples suggère qu'un tournant remarquable est en train de se produire, et cet ICM arrive à point nommé.

Pour réitérer ce point concernant la trace de pensée, on peut lire des passages édités tels que : "Cette erreur apparente doit être localisée, et pas seulement notée" - le modèle cherche à comprendre son propre faux pas -, ou encore, vers la fin du document, de l'autosatisfaction : "Cela semble toujours fonctionner, rien ne réfute cet argument", ce qui ressemble trait pour trait aux notes que je griffonne sur mon bloc-notes lorsque je cherche un problème.

Voilà pour l'émerveillement. J'en viens maintenant aux inquiétudes.

J'ai le sentiment que des personnes comme moi, enthousiasmées par l'ia, obtiennent beaucoup d'espace médiatique, tandis que les voix réticentes s'entendent moins. Je souhaite donc citer deux personnes qui ne sont pas enthousiasmées, afin d'être honnête face à des préoccupations légitimes. L'un d'eux déclare : "En tant que mathématicien, je ne me soucie pas seulement du résultat final, mais aussi du voyage qui y mène. L'ia détruit mon voyage." Cela concerne l'impact de l'ia sur la vie professionnelle quotidienne d'un mathématicien.

Et Peter Scholze : "Je considère déjà l'influence de l'ia comme fortement négative pour l'humanité, pour la démocratie et pour la planète", ce qui vise les impacts sociétaux plus larges.

Je souhaite discuter un peu de l'avenir et de la manière dont nous, mathématiciens, devons réagir face à l'ia.

La première chose à souligner, incarnée par la première citation, est que la période sera difficile. C'est un temps de changement, mais comme beaucoup, je pense que les temps de changement sont aussi de formidables opportunités.

Ensuite, rien n'indique que ces systèmes ralentissent. Au contraire, tous les indices montrent qu'ils continuent de progresser. Il est crucial d'en avoir conscience et d'agir de manière délibérée, tant collectivement qu'individuellement, d'une façon qui nous est inhabituelle. En tant que mathématiciens, notre système fonctionnait depuis longtemps, et nous allons soudainement affronter une foule de questions inédites.

Je souhaite terminer par une conviction profonde, qui prend la forme d'un plaidoyer pour la compréhension.

Ces citations illustraient la difficulté de la transition. Venons-en à l'évaluation comparative (*bench-*

*marking*) : nous devons nous attendre à ce que ces systèmes deviennent encore plus puissants.

Il est fondamental de comprendre que la difficulté et l'intelligence sont des notions à très haute dimension. On ne peut pas les projeter sur un axe unidimensionnel tel que le qi. Nous constatons de plus en plus qu'il existe des problèmes relativement faciles pour l'ia et très ardus pour les humains, et espérons-le inversement. Nous n'en sommes qu'aux prémices de la cartographie de ce paysage.

Il est donc très difficile d'évaluer les capacités des modèles en mathématiques, d'autant que nous ne disposons d'aucune assise théorique solide, et que les laboratoires se livrent une concurrence féroce tout en pratiquant un grand secret.

L'ia et l'apprentissage automatique raffolent des bancs d'essai (*benchmarks*), où l'on teste un modèle sur des ensembles de données pour obtenir des scores de 80 ou 90 %. Mais il est très difficile d'évaluer ainsi les mathématiques.

L'initiative *First Proof*, dont le second cycle vient d'être annoncé, constitue à cet égard notre étalon-or pour évaluer les systèmes d'ia en mathématiques de recherche. Le protocole est rigoureux : des mathématiciens élaborent des lemmes et des preuves inédits tirés de travaux en cours non publiés, éliminant tout risque de contamination des données. Ils s'engagent au préalable sur un pronostic (ex. : "Je serais très impressionné si l'ia résolvait ceci"). Différents systèmes sont ensuite testés, disposant d'une seule tentative, et les solutions sont expertisées par un comité de lecture.

Lors de ce dernier cycle, quatre systèmes ont été testés sur 10 problèmes. Les meilleurs modèles ont résolu environ 5 problèmes sur 10 (le vert sur ce graphique indique une réussite complète, le noir un échec, le rouge des révisions majeures, et les nuances intermédiaires des révisions mineures dues à des arguments manquants ou des citations inadéquates).

C'est une image précieuse de notre état actuel, mais elle ne montre pas encore de gradient temporel clair.

Pour vous convaincre de la trajectoire ascendante vers des systèmes plus puissants, examinons des mesures certes moins pures, mais dotées d'un gradient.

Ce graphique très intéressant présente, en échelle logarithmique sur l'axe vertical, le temps nécessaire à un humain pour accomplir une tâche (4 minutes, 4 jours, etc.). On y voit une progression nette : les llm parviennent à accomplir des tâches exigeant de plus en plus de temps aux humains, traduisant une amélioration exponentielle au fil du temps.

Un autre banc d'essai curieux est *FrontierMath Tier 4*, où l'on demande à des mathématiciens de concevoir des problèmes de recherche d'une extrême difficulté admettant une réponse immédiate, telle qu'un nombre (vrai/faux, correct ou non). Lorsque ce banc d'essai a été publié, les modèles obtenaient zéro, et aujourd'hui ils résolvent potentiellement plus de 70 % de ces problèmes. La trajectoire est indéniable, et ces problèmes sont redoutables : face à dix d'entre eux, je n'aurais aucune chance sur neuf problèmes, et peut-être une infime chance sur le dernier en y consacrant des jours.

Un autre indicateur fascinant est la capacité des modèles à mener des cyberattaques en s'introduisant dans des systèmes confinés. Il est troublant de constater que les progrès en mathématiques ont coïncidé avec cette capacité à pirater des logiciels. Trouver un contre-exemple mathématique et briser un système informatique ne sont finalement pas des tâches aussi éloignées que je le pensais.

Nous devons donc agir avec méthode. Nous allons affronter des questions inédites : comment allouer les ressources de manière appropriée ?

Hier, lors d'une table ronde, quelqu'un a posé une question qui me préoccupe également : comment gérer le fait que certains peuvent payer pour ces modèles et d'autres non ? Cela perturbe la beauté des mathématiques, discipline où l'on a toujours eu besoin seulement d'un crayon, d'une feuille de papier et d'un interlocuteur.

Quelle doit être notre relation avec l'industrie ? C'est une question épineuse, car une grande partie de la recherche et des connaissances réside désormais dans le secteur privé, lequel ne s'intéresse pas aux mathématiques par simple curiosité intellectuelle. C'est un fait. Et face à cette déferlante de publications, comment devons-nous réagir ?

Pour illustrer la question des ressources, j'ai cherché un équivalent dans l'histoire des mathématiques. En physique ou dans d'autres sciences, il n'est pas rare de réclamer d'importantes ressources pour mener une expérience cruciale. En mathématiques, c'est rarissime. J'ai trouvé un bel exemple historique dans les années 1970 avec le groupe simple *Monster* (*le Monstre*), un objet magnifique et gigantesque dont l'existence était soupçonnée mais non démontrée.

John Sims, dans un article sur la construction d'objets similaires, discutait de l'application de sa technique au Monstre, avant de douter que l'existence ou la non-existence du Monstre présente une telle importance pour la communauté mathématique qu'elle justifie de mobiliser les ressources d'un grand centre informatique pendant près d'un an pour trancher la question. Fait amusant, le Monstre a résisté aux approches computationnelles jusqu'à une époque très récente, et les premières preuves de son existence relevaient de la pure pensée.

Par coïncidence, au moment-même de cette conférence, se tient une table ronde sur la Déclaration de Leiden, un document fondamental très important à lire concernant la manière dont nous, mathématiciens, devons répondre à l'essor de l'ia. Cela fait partie des discussions cruciales que nous devons mener.

Pour conclure, je souhaite affirmer que la compréhension et le sens humains doivent impérativement demeurer. Ce "doivent" exprime un plaidoyer : je tiens à continuer de penser le monde mathématique, et je veux que la compréhension reste primordiale. Dans une certaine mesure, c'est un choix.

Pour étayer ce propos, je formulerai trois thèses, suivies de quelques réflexions.

Première thèse : Les mathématiques ont toujours été une question de compréhension.

Nous ne cherchons pas nécessairement des démonstrations ; nous cherchons la compréhension. Et de manière plus controversée potentiellement, le processus de création mathématique est plus important que le résultat.

Pour illustrer cela, imaginez que vous donniez à un étudiant un problème à résoudre ou un théorème à prouver. Le résultat final du calcul ou la preuve elle-même vous importe peu ; ce qui compte, c'est la compréhension qui émerge de la démarche.

Autre exemple auquel j'ai beaucoup pensé en préparant cet exposé : lorsque nous devons nous justifier auprès d'agences de financement ou de donateurs, nous affirmons souvent que sans les mathématiques, il n'y aurait ni cryptographie à clé publique, ni navigation, ni GPS. C'est vrai, et c'est un argument recevable. Mais il est infiniment plus vrai que la compréhension mathématique qui imprègne le monde est un bien public inestimable, et que la valeur de cette compréhension surpasse de loin tout résultat concret qui en découle.

Pour appuyer cette idée, citons ce livre délicieux, *Euclid and his Modern Rivals* (Euclide et ses rivaux modernes) de Charles Lutwidge Dodgson (Lewis Carroll), une pièce de théâtre où Euclide défend ses manuels de géométrie contre des concurrents modernes. Euclide y déclare, par la plume de Dodgson : "C'est l'*ergon* [l'œuvre en train de se faire] plutôt que l'*ergon* achevé dont nous avons besoin ici" - autrement dit, le travail à l'œuvre prime sur le résultat du travail.

Les mathématiques sont donc synonyme de compréhension. C'est notre première thèse.

Deuxième thèse : Nous utilisons la démonstration comme une mesure de substitution pour la compréhension depuis des siècles.

Il est très difficile de savoir si quelqu'un comprend véritablement quelque chose. Nous comprenons les choses de multiples manières, toutes extrêmement précieuses. Mais en tant que mathématiciens, nous disposons d'un indicateur unique et magnifique - "savez-vous le prouver ?" - que nous utilisons comme substitut de la compréhension.

Il y a eu de tout temps une tension entre la preuve et la compréhension dans l'histoire des mathématiques. Par exemple, l'incident célèbre où Alexandre-Théophile Vandermonde a énoncé un théorème de théorie des nombres sans le prouver, tandis que Carl Friedrich Gauss l'a énoncé, prouvé, sans citer Vandermonde. Ce dernier en fut fort marri. Félix Klein, commentant cet épisode, expliquait que Gauss considérait qu'une assertion mathématique n'a de valeur que si elle s'accompagne d'une démonstration complète, un point de vue rigoureux que Klein jugeait profondément injuste. Il existe ainsi une valeur inhérente dans ce qui n'est pas encore prouvé.

Troisième thèse : La preuve et la compréhension sont en train de se découpler.

Le premier choc majeur pour les mathématiciens à cet égard a été la preuve du théorème des quatre couleurs. C'était un problème topologique ouvert célèbre pendant un siècle, résolu grâce à une ingéniosité considérable en le réduisant à la vérification d'une multitude de graphes par ordinateur. C'était une preuve très controversée, car elle reposait pour une part essentielle sur des calculs auto-

matisés qu'aucun humain ne pouvait suivre ligne par ligne. Nous avons vu de nombreux exemples de ce type ces cinquante dernières années, mais je doute que nous en ayons tiré toutes les leçons, et nous allons assister à un découplage radical dans les années à venir.

Bill Thurston a formulé une citation célèbre à propos de cette controverse : "J'interprète cette controverse comme ayant peu à voir avec les doutes sur la véracité du théorème ou l'exactitude de la preuve, mais plutôt comme le reflet d'un désir persistant de compréhension humaine d'une démonstration, en plus de la simple certitude que le théorème est vrai." On observe ici clairement ce découplage entre preuve et compréhension.

Quelles sont dès lors nos options en tant que mathématiciens ?

Premièrement, nous affirmons depuis trop longtemps à la société que notre métier se résume à prouver des théorèmes ; nous devons marteler que notre rôle est de fournir une compréhension humaine du monde mathématique.

Deuxièmement, j'adore utiliser les systèmes d'ia au quotidien, mais je sens qu'ils m'entraînent subtilement vers l'absence de compréhension, en incitant à accepter passivement un résultat sans se l'approprier. C'est une pente glissante à laquelle nous devons résister, et je le rappelle chaque jour à mes étudiants en thèse comme à moi-même.

Si nous sommes constamment poussés à nous éloigner de la compréhension, nous devons créer une force de rappel : des incitations qui valorisent la transmission et la compréhension au sein de notre communauté, comme une excellente exposition et une communication rigoureuse à tous les niveaux. Je ne sais pas encore très bien comment les encourager au mieux, mais la transition sera ardue car nous abandonnons un indicateur unique et commode au profit d'une plus grande complexité.

Néanmoins, si ces outils sont aussi puissants que nous le pensons, nous pouvons aspirer à explorer des mondes nouveaux et extraordinaires, et c'est en cela que je demeure foncièrement optimiste.

Je souhaite conclure par une autre citation de Bill Thurston, adressée sur le forum *MathOverflow* à un étudiant diplômé qui demandait comment contribuer à cette longue conversation mathématique :

"Les mathématiques n'existent que dans une communauté vivante de mathématiciens qui propage la compréhension et insuffle la vie aux idées, qu'elles soient anciennes ou nouvelles."

Avant de vous remercier, je tiens à exprimer ma gratitude envers les personnes suivantes qui ont grandement contribué à l'amélioration de cet exposé.

Merci infiniment.

*[Applaudissements]*