RÉSEAU THÉMATIQUE CNRS · INSMI

IA, Mathématiques et Formalisation

UMPA
ENS DE LYON

Aurélien Garivier UMPA, ENS de Lyon
avec Marc Lelarge (Inria / ENS) et Assia Mahboubi (Inria, LS2N)

Journée des réseaux thématiques de l'Insmi — Paris, 17 novembre 2026

En une page

Le réseau thématique IAMF

- création au **1er janvier 2027**, Insmi (principal) et INS2I
- 85 permanents de 35 laboratoires, à la date du dépôt (mars 2026), de mathématiques et d'informatique
- deux axes : *IA et découverte mathématique* ; *formalisation et structuration de la communauté*

Cet exposé

1. État des lieux : capacités et limites des outils
2. Production et validation en mutation
3. Métier, formation et choix collectifs
4. Compétences et réponses en France
5. Ce que le réseau propose, et ce qu'il attend de vous

Ce que le réseau veut être

Un lieu de dialogue pour toute la communauté : se rencontrer, se former, évaluer les outils, en indépendance.

État des connaissances au 16 novembre 2026. Plusieurs résultats cités sont des annonces non encore validées par la communauté.

1 / État des lieux

Un assistant de preuve, c'est quoi ?

Théorème (Euclide). Pour tout entier n, il existe un nombre premier $p \geq n$.

Preuve. Soit p le plus petit facteur premier de $n! + 1$. Si $p < n$, alors p divise $n!$, donc p divise 1 : absurde.

- énoncé et preuve sont des objets formels
- un petit programme, le noyau, vérifie chaque étape
- reste à l'humain : vérifier que l'énoncé formel dit ce qu'on voulait

Le même théorème dans Mathlib (bibliothèque de Lean) :

```
theorem exists_infinite_primes (n : ℕ) :
    ∃ p, n ≤ p ∧ Prime p :=
  let p := minFac (n ! + 1)
  have f1 : n ! + 1 ≠ 1 :=
    ne_of_gt ◁ succ_lt_succ ◁ factorial_pos _
  have pp : Prime p := minFac_prime f1
  have np : n ≤ p :=
    le_of_not_ge fun h =>
      have h₁ : p ∣ n ! := dvd_factorial (minFac_pos _) h
      have h₂ : p ∣ 1 :=
        (Nat.dvd_add_iff_right h₁).2 (minFac_dvd _)
      pp.not_dvd_one h₂
  ⟨p, np, pp⟩
```

Source : Mathlib, `Mathlib/Data/Nat/Prime/Infinite.lean` (branche principale, consultée le 16/11/2026)

Avant les agents : vingt ans de projets portés par des équipes

2005 **Quatre couleurs** en Coq (Gonthier, Werner ; Inria / Microsoft Research)
Une preuve par ordinateur depuis 1976, que personne ne pouvait relire à la main ; Coq vérifie dans un même cadre le raisonnement et le calcul.

2012 **Feit–Thompson** en Coq (Gonthier, Mahboubi, Cohen…) : 6 ans, 170 000 lignes
Un énoncé qui tient en une ligne, une preuve de 255 pages : six ans de travail d'équipe pour une version intégralement vérifiée par la machine.

2014 **Conjecture de Kepler** (Hales, projet Flyspeck) : environ 20 années-personnes, 500 000 lignes
Les relecteurs étaient convaincus sans pouvoir certifier les calculs ; Flyspeck, lancé en 2003, vérifie formellement les arguments et les calculs.

6 ans
pour formaliser Feit–Thompson (2006–2012)

≈ 20
années-personnes pour la conjecture de Kepler (2003–2014)

Une communauté petite, exigeante, où la France compte depuis l'origine (Coq, devenu Rocq, est né à l'Inria).

Sources : Gonthier, Notices AMS 2008 ; Gonthier et al., ITP 2013 ; phys.org 10/2012 ; Hales et al. 2015 (Flyspeck) ; Hales, blog de T. Tao, 9/10/2026

Avant les agents : Lean, une communauté qui s'organise

2022 **Liquid Tensor Experiment** en Lean (défi de Scholze, mené par Commelin) : 19 mois

Un médaillé Fields demande à la communauté Lean de vérifier un résultat dont il doute : doutes levés en six mois, défi achevé en dix-neuf.

2022 **Retournement de la sphère** en Lean (Massot, van Doorn, Nash)

Un théorème qu'on explique avec les mains, formalisé par une petite équipe en passant par un résultat plus général, le h-principe de Gromov.

2023 **Conjecture PFR** en Lean (Tao, Dillies et 25 à 30 contributeurs) : 3 semaines

Preuve déposée le 9 novembre, projet lancé le 12, achevé le 5 décembre, chacun sur un lemme sans maîtriser la preuve entière : la formalisation devient un mode d'organisation du travail.

3 semaines

pour PFR (2023), contre 6 ans pour Feit–Thompson ; périmètres non comparables

1 bibliothèque

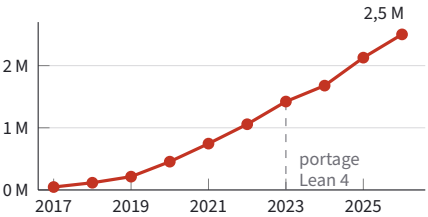
commune, Mathlib, et un plan de preuve (*blueprint*) qui découpe le travail

Ce qui accélère : la bibliothèque et l'organisation, avant toute IA.

Sources : blog de la communauté Lean (LTE, 2022) ; projet sphere-eversion (README) ; T. Tao, blog, 18/11/2023 ; Quanta 6/12/2023

Mathlib : dix ans de croissance, relue par des humains

Lignes de code (millions, au 31 décembre ; 2026 : au 16 novembre)



Croissance à peu près **linéaire** depuis 2020 : 300 à 450 000 lignes par an, environ 10 000 contributions fusionnées.

294 155

théorèmes, 140 046 définitions, 772 contributeurs

63

relecteurs dont 28 mainteneurs ; 2 800 PR en attente

Règles depuis janvier 2026 (« Use of AI »)

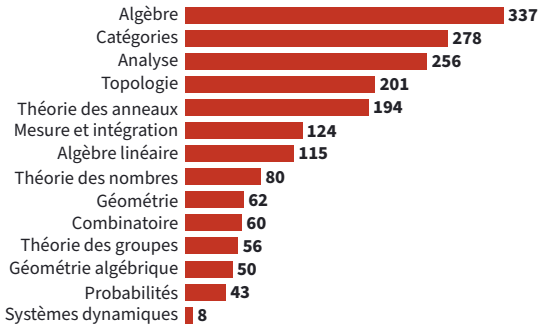
- tout usage d'IA **déclaré** (outil, manière) ; étiquette *LLM-generated*
- l'auteur doit **comprendre et justifier** chaque choix « *without the use of an AI* »
- commentaires rédigés par un LLM : interdits
- PR de basse qualité fermées sans discussion ; récidive : bannissement

Les bibliothèques écrites par des agents (Fermat 13 M de lignes, Tau Ceti 3 M, Axiom) restent à côté : « *Mathlib grows steadily, but more slowly than AI-generated libraries* » (Avigad, Buzzard, Commelin et al., 16/10/2026).

Sources : mathlib.org/mathlib_stats (fichiers gitstats, 16/11/2026) ; arXiv :1910.09336 ; leanprover-community.github.io/contribute ; xenaproject.wordpress.com, 16/10/2026

Mathlib : une bibliothèque inégale selon les domaines

Milliers de lignes par dossier de Mathlib (sélection)



Bien couverts : algèbre, catégories, topologie, analyse générale, mesure.

Plus récents ou plus minces : probabilités (TCL, martingales, mouvement brownien), géométrie algébrique, systèmes dynamiques.

Sans dossier propre : EDP, analyse numérique, optimisation, statistiques.

Ordres de grandeur : un dossier n'est pas une discipline, une ligne ne mesure pas la profondeur.

Sources : mathlib.org/mathlib_stats.html et décompte des lignes du dépôt mathlib4 (16/11/2026) ; blog de la communauté Lean (bilan 2021)

Un modèle de langage, c'est quoi ?

Une loi de probabilité sur le fragment de texte suivant (*token*) :

$$p_\theta\left(x_{t+1} \mid x_1, \ldots, x_t\right)$$

- θ : un très grand nombre de paramètres
- ajustés par descente de gradient pour prédire la suite de textes existants (dont arXiv, les manuels, les forums)
- on engendre un texte en tirant un fragment après l'autre

Conséquence

Engendrer du texte **ne garantit pas** sa correction.

Une sortie peut être juste ; ce qui manque, c'est la garantie. Elle doit venir d'une vérification adaptée.

Avant l'entraînement à raisonner (2024) : un « perroquet stochastique » (Bender, Gebru et al., 2021), très cultivé : la même aisance pour le vrai et pour le faux. D'où le niveau de 2021–2024, et le rôle que va jouer la vérification.

2021–2024 : un niveau faible, et mesuré

mars 2021 Banc d'essai MATH (niveau lycée et concours) : meilleur modèle 6,9 % ; un triple médaillé d'or aux Olympiades : environ 90 %

janv. 2023 ChatGPT : capacités « *significantly below those of an average mathematics graduate student* » (Frieder et al.)

mai 2023 GPT-4 : 59 % de réussite sur des multiplications de deux nombres à 3 chiffres (Dziri et al.)

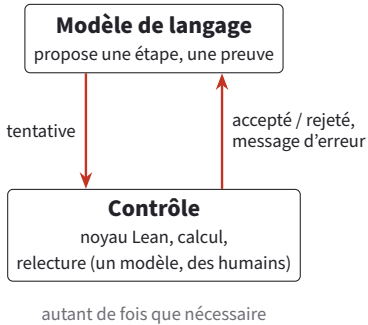
juil. 2024 PutnamBench (preuves formelles) : 6 problèmes sur 640 résolus, tous systèmes confondus

sept. 2024 T. Tao, sur le modèle o1 : un étudiant de thèse « *mediocre, but not completely incompetent* » ; les modèles précédents étaient « *actually incompetent* »

nov. 2024 FrontierMath (problèmes inédits de niveau recherche, Epoch AI) : les meilleurs modèles en résolvent moins de 2 %

Sources : arXiv :2103.03874 ; arXiv :2301.13867 ; arXiv :2305.18654 ; PutnamBench ; T. Tao (Mastodon, 9/2024, propos nuancés ensuite) ; arXiv :2411.04872

Ce qui a changé : proposer, vérifier, recommencer



Trois ingrédients

- **calcul au moment de la réponse** (sept. 2024) : le modèle essaie, se relit, recommence
- **renforcement sur des réponses vérifiables** : en langage naturel pour les modèles généralistes (o1, DeepSeek-R1) ; contre le noyau Lean pour les prouveurs spécialisés (AlphaProof, 2024 : 1 million de problèmes traduits)
- **agents** : le modèle agit, observe (document, Lean, calcul), recommence ; parfois à plusieurs, pendant des jours

Deux filières : prouveurs spécialisés en Lean (AlphaProof, Aristotle, Axiom) ; généralistes en langage naturel (or aux Olympiades 2025, distances unité, 722 manuscrits). **Environ un tiers des résultats de 2026 ont un certificat formel**, souvent après coup (Navier–Stokes : 88 h de preuve, 17 h de Lean).

Sources : DeepMind 25/7/2024 et 21/7/2025 ; arXiv :2501.12948 ; whataifound.org, 12/11/2026 (113 résultats 2026, 34 % avec preuve formelle)

2024–2025 : des olympiades aux problèmes ouverts

janv. 2024	AlphaGeometry (DeepMind) : 25 problèmes de géométrie d'olympiades sur 30
juil. 2024	AlphaProof et AlphaGeometry 2 : niveau médaille d'argent aux Olympiades (28 points sur 42), jusqu'à 3 jours de calcul par problème ; AlphaProof résout trois problèmes en Lean, AlphaGeometry 2 un quatrième
déc. 2024	o3 (OpenAI) : 25 % revendiqués sur FrontierMath ; environ 10 % mesurés sur le modèle public
juil. 2025	Performance de niveau médaille d'or aux Olympiades (Google, OpenAI) : 35 sur 42, en langage naturel, dans le temps de l'épreuve
sept. 2025	Théorème des nombres premiers fort formalisé par un agent (Gauss, Math Inc.) : 25 000 lignes en 3 semaines
fin 2025	Premiers problèmes d'Erdős résolus avec une contribution documentée d'une IA

Sources : DeepMind 17/1/2024 et 25/7/2024 ; TechCrunch 20/4/2025 ; S. Willison 21/7/2025 ; math-inc.github.io/strongpnt ; erdosproblems.com

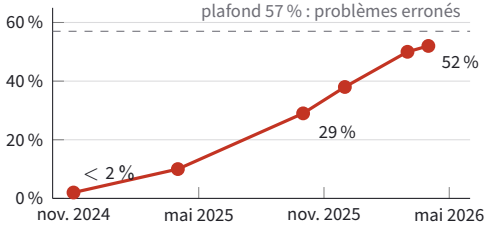
Janvier–août 2026 : la recherche

janv.–fév.	Problèmes d'Erdős : des dizaines passés à « résolu » avec l'aide d'IA depuis octobre 2025 ; entre Noël et le 14 janvier, 15 résolus, dont 11 créditant une IA (erdosproblems.com, TechCrunch)
20 mai	Conjecture des distances unité d'Erdős (1946) réfutée par un modèle d'OpenAI ; note de 9 mathématiciens. T. Gowers l'accepterait aux *Annals* « *without hesitation* »
21 mai	AlphaProof Nexus : 9 problèmes d'Erdős ouverts résolus sur 353 tentés, quelques centaines de dollars chacun, tout vérifié en Lean
20 juil.	Contre-exemple annoncé à la conjecture jacobienne (1939) (L. Alpöge avec un modèle d'Anthropic), vérifié en Lean et Isabelle en quelques heures
1er août	OpenAI : 10 problèmes ouverts depuis au moins dix ans, dont un groupe non sofique explicite ; environ 2 000 dollars de calcul
18 août	Fermat entièrement formalisé en Lean par des agents (Anthropic) : 13 millions de lignes en 11 jours

Sources : TechCrunch 14/1/2026 ; note Alon et al., 5/2026 ; arXiv :2605.22763 ; Fortune 21/7/2026 ; The Next Web 8/2026 ; anthropic.com 4/9/2026. **Statut :** annonces vérifiées formellement, relecture humaine inégale.

Deux mesures de la vitesse

FrontierMath, version 1 (part des problèmes résolus, réponses finales)



Version 1 close en juin 2026 (« *approaching saturation* », Epoch) ; la version 2, corrigée, non comparable, est à 95 % en nov. 2026 ; Epoch passe aux problèmes ouverts.

First Proof (10 questions de recherche inédites par lot, évaluées par des experts ; lots et critères différents)

juin 2026 — meilleur système — **6 / 10**

oct. 2026 — meilleur système — **9 / 10**

oct. 2026 — seul système à poids ouverts testé — **0 / 10**

Sources : Epoch AI (arXiv :2411.04872 ; données du Benchmarking Hub, 16/11/2026 : meilleur score mesuré par date d'évaluation) ; arXiv :2606.18119 ; 1stproof.org (troisième lot, octobre 2026)

8 septembre 2026 : Navier–Stokes

L'annonce d'OpenAI. Explosion en temps fini pour Navier–Stokes en dimension 3 avec forçage lisse (énoncés C et D de la formulation de l'institut Clay). 166 pages et une preuve Lean acceptée par le noyau.

Le statut, au 16 novembre.

- institut Clay, 11 sept. : « *has apparently been settled* » ; processus « *deliberately unhurried* » ; le problème reste affiché « Active »
- validation en cours : aucune vérification humaine complète rapportée dans les sources consultées
- des écarts entre preuve Lean et preuve rédigée, sans conclusion sur la correction (Bastounis et al., 6 oct.)
- controverse de priorité (T. Buckmaster, L. Alpöge)

88 heures
de calcul, du 1er au 5 septembre

≥ 10 000 agents
simultanés, 2,7 millions de messages

130 milliards
de *tokens* produits

Coût : aucun chiffre officiel ; la presse l'estime à plusieurs millions de dollars.

Sources : Wikipedia (*Navier–Stokes priority controversy*, 8/11/2026) ; unite.ai ; Scientific American 8/9/2026 ; Hales 9/10/2026 ; arXiv :2610.08144 ; claymath.org (16/11/2026)

6 octobre 2026 : 722 manuscrits

Ce qu'OpenAI met en ligne (modèle interne non public)

- le 6 octobre : 722 manuscrits, 372 familles de résultats, présentés comme « *raw research data* »

Ce qui suit

- 7 oct. : 3 manuscrits retirés, 14 révisés ; la moitié des résultats « non confirmés » selon l'entreprise
- zéros non triviaux de ζ : partie réelle entre $1/8$ et $7/8$. 199 pages vérifiées en Lean, jugées correctes pour l'essentiel, réécrites par des humains en 7 à 13 pages
- certaines formalisations « *prove weaker statements than what's claimed* » (J. Commelin)

722

manuscrits en un jour

$\approx$ 42 %

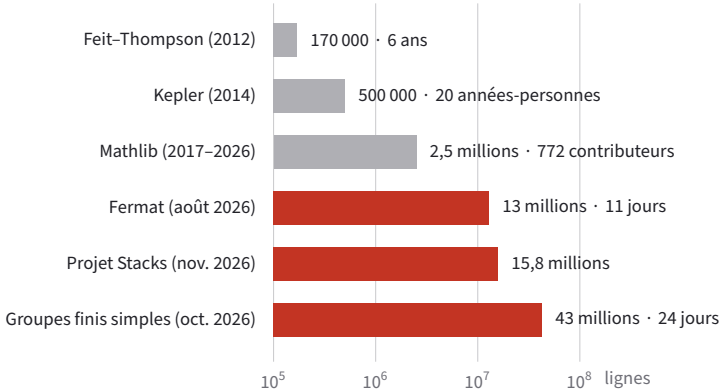
des résultats principaux accompagnés d'une preuve Lean

« horribly written »

— M. Radziwiłł, Quanta, 19/10/2026

Sources : github.com/openai/math ; Retraction Watch 8/10/2026 ; Quanta 9/10, 19/10 et 26/10/2026. Quanta et le New York Times comptent 377 familles.

La formalisation change d'échelle



écrit à la main
produit par des agents

11 j → 30 h
pour Fermat, entre août et novembre 2026

Le projet humain de formalisation de Fermat (K. Buzzard) était prévu sur **5 ans**.

Sources : Hales 9/10/2026 ; anthropic.com 4/9/2026 ; Quanta 11/11/2026 ; annonces d'Axiom Math (26/10 et 4/11/2026), sans certification indépendante

. . . et montre ses fragilités

- **Le noyau peut avoir des bugs.** 25–28 juillet 2026 : une fausse « réfutation de Collatz » est acceptée par le noyau de Lean *et* par un vérificateur indépendant. Corrigé en une heure ; d'autres bugs trouvés pendant l'été, par des IA.
- **L'énoncé formel peut ne pas dire ce qu'on croit.** Navier–Stokes : certaines hypothèses intermédiaires diffèrent entre Lean et le texte (5 dérivées au lieu de 4).
- **Le volume dépasse la relecture.** 26 manuels formalisés automatiquement en mai 2026 : résultat « *riddled with errors* » selon Quanta.
- **Le code produit reste hors de Mathlib**, sans ses conventions.

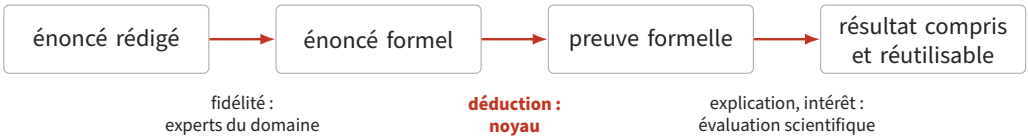
Sources : Hales, blog de T. Tao, 9/10/2026 ; gigazine.net 3/8/2026 ; arXiv :2610.08144 ; Quanta 11/11/2026

> *« A proof in Lean should not be accepted until a human audit is performed to ensure statement fidelity. »*
>
> — T. Hales, 9/10/2026

Ce que cela demande

Des spécialistes des fondements, des noyaux, des bibliothèques.
Cette compétence existe en France.

Ce qu'une preuve formelle garantit, et ce qu'il reste à faire

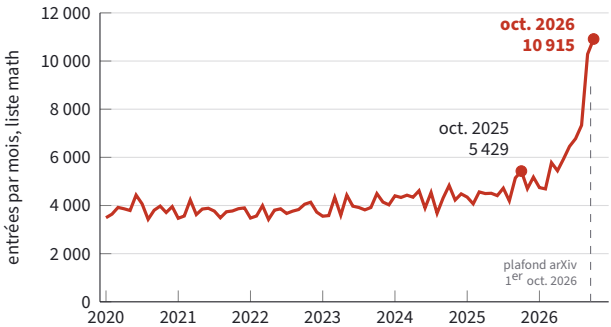


- Le **noyau** contrôle la preuve relativement aux définitions et aux axiomes déclarés, et rien d'autre.
- Les **experts du domaine** contrôlent que l'énoncé formel dit bien ce que dit l'énoncé rédigé.
- L'**explication**, la nouveauté et l'intérêt relèvent d'une évaluation scientifique.

Ces trois contrôles demandent des compétences différentes : celles d'IAMF et celles de vos réseaux.

2 / Production et validation en mutation

arXiv : les dépôts de mathématiques ont doublé en un an



× 2

entre octobre 2025 et octobre 2026

+ 6 % puis + 49 %

cumul de janvier à octobre : 2025, puis 2026

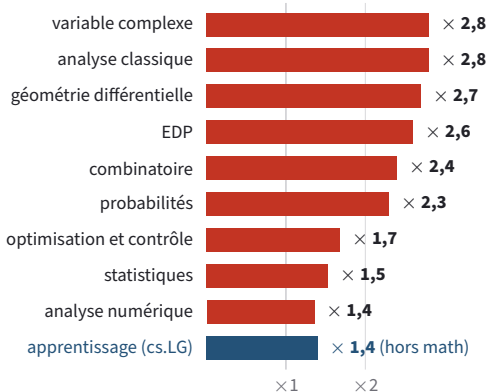
42 152

dépôts en octobre 2026 sur tout arXiv, toutes disciplines : record (27 692 un an plus tôt)

Sources : arxiv.org (listes mensuelles math, listes croisées incluses ; statistiques officielles), relevé du 16/11/2026 ; blog d'arXiv, 1/10/2026

Tous les domaines, inégalement

Octobre 2026 rapporté à octobre 2025 (sélection)



EDP, un comptage détaillé (M. Lewin)
du 15 sept. au 14 oct. : 531 articles en 2025, 1 302 en 2026 ; 493 déclarent un usage d'IA, dont 173 pour les preuves.

Probabilités
une dizaine d'articles par jour en 2016 ; 31 à 49 à la mi-octobre 2026.

Usage déclaré de l'IA (Epoch AI)
part des prépublications de mathématiques qui le mentionnent : 0,35 % en janvier 2025, 3,9 % en avril 2026, 25 % en août 2026.

Sources : arxiv.org, relevé du 16/11/2026 ; M. Lewin, proofsandprompts.com, 2/11/2026 ; epoch.ai/data-insights/math-preprints-disclosed-ai-use (déclarations volontaires)

Les digues : arXiv et les revues

1er octobre : arXiv plafonne les dépôts

- « *up to two submissions per calendar month, with a limit of three total active submissions* »
- toutes catégories ; mesure dite provisoire (« *stopgap* »)
- la hausse se poursuit en octobre

19 octobre – 8 novembre : cinq revues ferment leurs soumissions

- *Annales de l'IHP B*, *Bernoulli*, *EJP*, *ECP*, *ESAIM : P&S* ; *SPA* diffère le traitement
- à la réouverture, *Bernoulli* exige une déclaration d'usage de l'IA et autorise ses rapporteurs à s'en servir

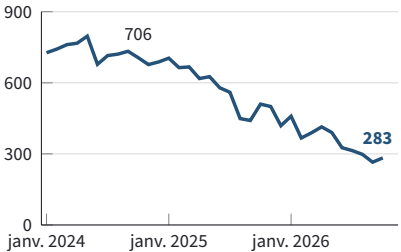
> *« At the current pace, we cannot give every manuscript the care and attention we believe it deserves. »*
>
> — communiqué des rédactions, 19/10/2026

Signé par les rédacteurs en chef :
N. Curien (*AIHP-B*), K. Kato (*Bernoulli*), C. Toninelli (*EJP*), P. Gonçalves (*ECP*), B. de Saporta et C. Duval (*ESAIM : P&S*), A. Guillin et E. Löcherbach (*SPA*).

Sources : blog.arxiv.org, 1/10/2026 ; communiqué des rédactions, smf.emath.fr, 19/10/2026 ; EDP Sciences, 19/10/2026 ; bernoullisociety.org, 11/2026

De nouveaux espaces de circulation

MathOverflow (questions posées par mois)



—60 % entre octobre 2024 et octobre 2026 ; les causes ne sont pas établies.

Des dépôts dédiés aux productions d'IA

- *Hexagon* : 1 061 travaux en dix semaines ; 51 % sans auteur humain revendiqué ; 18 % que le déposant déclare ne comprendre « en aucune partie »
- *Palomar* : 1 146 résultats vérifiés en Lean
- *MathHelm* (13 nov.) : forum de discussion des articles engendrés par IA

Sources : API Stack Exchange ; hexagonmath.org et palomar-registry.org (16/11/2026) ; blog de T. Tao, 13/11/2026 ; proofsandprompts.com, 14/11/2026

3 / Métier, formation et choix collectifs

Des prises de position collectives

2 juin 2026	**Déclaration de Leiden**, soutenue par l'Union mathématique internationale, 4 447 signataires : « *Mathematicians have a choice about whether and how to adopt artificial intelligence* »
11 sept.	**Lettre de 25 médaillés Fields** (28 aujourd'hui), 8 458 soutiens : « *The goals of the AI companies and the goals of the mathematical community are severely misaligned.* »
sept.	**Association for Human Mathematics** : 1 165 membres, dont 752 s'engagent à ne pas utiliser l'IA en recherche
21 sept.	**AGMAI**, groupe consultatif indépendant de 9 mathématiciens : demande aux entreprises de ne plus traiter les résultats mathématiques comme des « *marketing vehicles* »
fin sept.	Lettre **« Mathematics and AI »** : « *we encourage mathematicians to use them liberally but thoughtfully* »

Sources : leidendeclaration.ai ; mathandai.org ; ahmath.org/members ; agmai.org ; sites.google.com/view/math-and-ai-together (compteurs relevés le 16/11/2026 ; ce ne sont pas des mesures d'opinion)

Un éventail de positions

Refuser

« *Mathematicians do not need to use AI, nor should we use AI.* »

M. Weinreich, août 2026

Alerter

« *If we insist on focusing our entire profession solely on the tackling of conjectures [. . .] we may be out of business quite fast, not because of the end of mathematics but because of the end of our imagination.* »

M. Marcolli, sept. 2026

S'adapter

« *Deep theorems were scarce and difficult and so became an effective mechanism to identify deep thought. AI has broken this system.* »

B. Kra, sept. 2026

Accueillir

« *Je ressentis un vrai soulagement, et une sensation de liberté retrouvée.* »

S. Coste, oct. 2026

« *Mathematics has been struck by an asteroid. [. . .] We need to face it.* »

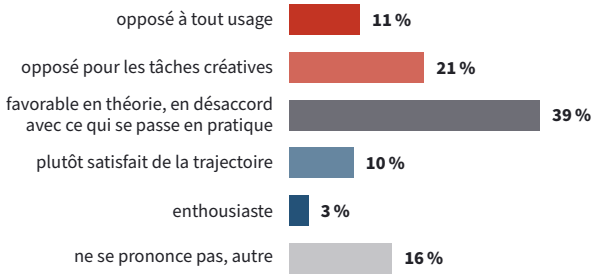
F. Breuer, nov. 2026

Une crainte partagée : « *that the mathematical community might become bitterly divided* » (T. Gowers, 17/9/2026).

Sources : arXiv :2608.02859 ; Marcolli, *Duet for the end of math* ; blog de T. Tao (13/9 et 17/9/2026) ; Le Grand Continent 25/10/2026 ; proofsandprompts.com 13/11/2026

Les jeunes chercheurs : une enquête

« Quelle est votre position générale sur l'IA en mathématiques ? »



268 doctorants et post-doctorants, surtout aux États-Unis ; 17 sept.–8 oct. 2026. Enquête exploratoire, non représentative.

70 %
jugent moins probable de rester dans la recherche académique

8 %
ont décidé de changer de voie ; 56 % y songent

55 %
des utilisateurs n'aiment pas s'en servir pour leur recherche ; 31 % aiment

62 %
des abonnés paient eux-mêmes (plusieurs réponses possibles)

Sources : Brandenberger, Mularczyk, Woodruff, arXiv :2610.40584 (questions 7.4, 3.2, 3.1, 4.2, 4.5)

Qu'est-ce qu'un bon travail de recherche, demain ?

Indicateurs moins fiables du travail humain

- être le premier à prouver (Harvard/CMSA, rec. 8)
- la preuve de routine, le lemme « gentil » (T. Gowers)
- la synthèse d'un domaine : produite vite, elle ne prouve plus la maîtrise
- le volume (arXiv $\times$ 2)
- le théorème comme signal de profondeur (B. Kra)

À mieux reconnaître et évaluer

- formuler et choisir les questions
- vérifier, auditer, relire
- expliquer, transmettre, enseigner
- relier des domaines, des communautés
- le goût et l'imagination : sentir ce qui est fécond, douter, recommencer (Marcolli ; Duminil-Copin : « *search, doubt, experiment, fail, and restart* »)

Une question ouverte, une orientation

Compléter le texte par des preuves directes de compréhension : discussion, explication, suivi du travail. Déjà engagé dans les recrutements : moins d'articles demandés, plus d'entretiens (CNRS, Cambridge, ETH Zurich, IAS, MIT, MPIM Bonn). T. Koberda : « *a polished dissertation can no longer, by itself, certify the intellectual formation of its nominal author* »

Sources : cmsa.fas.harvard.edu, 17–18/9/2026 ; gowers.wordpress.com, 8/5/2026 ; B. Kra, 13/9/2026 ; M. Marcolli, 9/2026 ; H. Duminil-Copin, 30/8/2026 ; arXiv :2609.28140 ; blog de T. Tao 13/11/2026 ; proofsandprompts.com 14/11/2026

Apprendre les maths avec l'IA : quelle autonomie à l'arrivée ?

Des risques, même si la réponse est juste

- comprendre une solution à la lecture sans savoir la reconstruire
- déléguer trop tôt la recherche, le calcul ou la preuve, et moins les pratiquer
- accepter une explication erronée faute de connaissances pour la discuter

Essai randomisé au lycée (Bastani et al., PNAS 2025, n ≈ 1 000) : +48 % aux exercices assistés, −17 % au test sans IA ; pas d'effet significatif avec un tutorat encadré.

Des bénéfices possibles, selon l'usage

- obtenir un indice ou une explication supplémentaire au moment du blocage
- s'entraîner sur des exemples variés, à son rythme
- faire discuter une tentative personnelle et expliciter ses erreurs

La conception de l'aide compte : un retour métacognitif réduit la délégation et améliore le test sans aide (Maier et al., 2026, n = 704).

Ce qu'il faut expérimenter et mesurer

chercher → aide graduée → reconstruire sans aide → réutiliser sur un problème nouveau, à distance. **Le critère : ce que l'étudiant sait ensuite expliquer et faire seul** (P. Etingof : « *would I still understand…?* »)

Sources : Bastani et al., PNAS 122(26), 2025 ; Maier et al., arXiv :2609.20143 ; cmsa.fas.harvard.edu 17–18/9/2026 ; P. Etingof, MIT, 14/9/2026. Pas d'étude contrôlée au doctorat (AHM, 27/10/2026).

Jeunes chercheurs : se former, rester, avec quelles règles ?

Le travail de thèse change

lire, se familiariser, premiers calculs, « *gentle problems* » : **déjà automatisables** (T. Gowers, 8/5/2026)

→ les **contributions attendues se déplacent** : formuler, choisir, vérifier, expliquer (Harvard/CMSA) ; les étapes de formation restent nécessaires

L'IA pendant la thèse : deux lignes

Expérimenter

Harvard/CMSA : usage déclaré ; Charles : « sous supervision »

Limiter, voire exclure

AHM : « *sparingly or not at all* » ; Villani : « surtout pas »

Évaluer la thèse autrement ?

« *A PhD should not be awarded primarily on the basis of the text of the dissertation* » (Harvard/CMSA, rec. 9)

Soutenance-discussion de plusieurs heures, menée par un **opposant extérieur** (modèle scandinave)

Leçon d'essai de 45 min, sujet hors thèse fixé par le jury dix jours avant, validée avant la soutenance (Norvège)

Protéger les débutants

« *Established mathematicians must actively protect developing researchers* » (sommet Harvard/CMSA, 23/10/2026)

ce qu'on sait de l'apprentissage avec l'IA : lycée et licence seulement (slide précédente)

Sources : AFP 15/9/2026 ; cmsa.fas.harvard.edu (17–18/9 et 22–23/10/2026) ; ahmath.org 27/10/2026 ; Gowers 8/5/2026 ; L'Étudiant 9/2026

Accès, coûts, dépendance

- **Certaines annonces majeures viennent de modèles internes**, non accessibles au public.
- **Le seul système à poids ouverts testé est loin derrière** (First Proof, oct. 2026 ; un seul modèle ouvert testé).
- **L'accès se paie** : abonnements à 20, 100 ou 200 dollars par mois ; au CNRS, l'agent Emmy (modèles Mistral).
- **Le calcul a un coût** : 6 à 15 millions de dollars estimés pour Navier–Stokes.
- **Des liens existent** : collègues consultants ou recrutés par les entreprises, accès anticipés.

« The use of proprietary internal models by AI labs to do mathematical research risks creating a two-tier system where labs outrun the rest of the field. »

— AGMAI, 29/9/2026

54 %

des jeunes chercheurs interrogés s'inquiètent de l'inégalité d'accès au calcul

Sources : agmai.org/general-sep29 ; Univ. Buffalo 10/2026 ; arXiv :2610.40584 ; estimations de presse pour Navier–Stokes

L'empreinte écologique : ce qu'on sait, ce qu'on ignore

Une requête

0,3 Wh

classique (Google 0,24 ; Altman 0,34 ; Epoch 0,3)

2 à 35 Wh

avec raisonnement : $\times 10$ à $\times 100$ (Jegham 2025 ; Microsoft 2026)

un chercheur, usage intensif : **100 à 500 kWh/an** $\approx$ son ordinateur $\approx$ 1 à 10 % d'un vol transatlantique (estimation)

Une exécution

0,1 à 1 GWh

Navier–Stokes, 130 milliards de *tokens* ; **estimation** : énergie non publiée, reconstruite ($\times 10$)

25 à 200

allers-retours Paris–New York en émissions (mix américain), ou un an d'un laboratoire de maths ; **estimation**

vérifier la preuve (Lean) : **0,1 %** de la générer (estimation)

Ce qu'on ignore

kWh, heures-GPU, lieu : jamais publiés ; seulement des *tokens* et des « millions de dollars »

tokens d'entrée, échanges entre agents : non comptés ($\times 2$ à $\times 10$)

carbone : $\times 8$ selon le pays d'hébergement (France $\approx$ 50 g CO_2e/kWh, États-Unis $\approx$ 400) ; **eau** : 0,26 ou 45 mL par requête selon le périmètre

effet rebond : $\div$ 33 par requête en un an (Google)… et pourtant $+50$ % en un an pour les serveurs d'IA (AIE) : l'usage croît plus vite que l'efficacité

Sources : arXiv :2508.15734, 2505.09598, 2509.20241 ; OpenAI 8/9/2026 ; bilan GES du LJAD 2023 ; impactco2.fr ; IEA 2026. Exécutions : nos reconstructions (*tokens*, prix), incertitude $\times 10$, voir sources/ecologie.txt.

Des risques, des choix

Ce qui pourrait être fragilisé

- **Emploi, missions** : « productivité » → postes, encadrement, crédits (AcadMathSci, 28/10/2026)
- **Confidentialité, attribution, responsabilité** : idées confiées aux outils · priorité contestée · qui signe, qui répond ? (Harvard/CMSA, règle 6 ; controverses 2026)
- **Tutelles, dirigeants** : justifier notre utilité

Des choix collectifs

protéger : les parcours de formation

préciser : les responsabilités, l'attribution

préserver : l'autonomie de recherche

expliquer : l'utilité collective de la recherche, aux tutelles comme au public

La technique ne tranche pas : évaluer, former, financer sont des choix collectifs.

Sources : AcadMathSci 28/10/2026 ; Harvard/CMSA 17–18/9/2026 ; Wikipedia (controverses de priorité 2026). Mécanismes de risque, pas des effets déjà mesurés.

4 / Compétences et réponses en France

Des compétences françaises en preuve formelle

1984	Calcul des constructions, puis **Coq** à l'Inria ; rebaptisé **Rocq** en 2025
2013	Coq reçoit l'*ACM Software System Award*
2021	CompCert (X. Leroy), compilateur C vérifié, reçoit le même prix
2005–2022	Quatre couleurs, Feit–Thompson, retournement de la sphère (partie 1) : des jalons portés par des équipes françaises

Équipes Inria autour de Rocq : Gallinette (Nantes), Picube (Paris), Cash (Lyon), Stamp (Sophia), Deducteam (Saclay)…

7 sur 28

mainteneurs de Mathlib en poste en France

A. Dedecker (Paris Cité), F. Nuccio (Saint-Étienne), J. Riou et P. Massot (Orsay), R. Brasca (Paris Cité), R. Degenne (Inria Lille), S. Gouëzel (Rennes)

2 écoles

deux cultures, Rocq et Lean, présentes toutes les deux dans le réseau

Sources : leanprover-community.github.io/teams/maintainers.html (16/11/2026, comptage manuel) ; rocq-prover.org ; awards.acm.org

IA et formalisation : des travaux en cours

2024–2026	**Numina** (association, Paris) : prix AIMO 2024 avec Hugging Face, données et prouveurs ouverts ; agent Lean, 12 sur 12 au Putnam 2025 ; prouveur informel avec A. Hayat (ENPC), une conjecture réglée
nov. 2024	ERC Synergy **MALINCA** (C. Simpson, Ph. de Groote, H. Herbelin, P.-A. Melliès) : rapprocher la langue du mathématicien de celle de la machine
2025	**Leaning and Rocq'ing** (G. Baudart, C. Cohen, N. Tabareau, M. Lelarge ; AI for Math Fund) : traduire automatiquement entre Lean et Rocq
nov. 2025	**Mouvement brownien** formalisé en Lean (R. Degenne et coauteurs)
mars 2026	Problèmes du **Putnam 2025 en Rocq** avec un modèle de langage (Baudart, Lelarge, Stérin, Viennot) ; **Leanstral** (Mistral), agent ouvert pour Lean
mai 2026	**ATLAS** (Meta FAIR Paris, avec A. Hayat, ENPC) : 26 manuels, plus de 45 000 énoncés Lean, 500 000 lignes, code ouvert
2026	Chaires, thèses et équipes à l'ENPC, à l'ENS, à l'Inria, à l'IRIF, au LIP…

Sources : projectnumina.ai ; arXiv :2601.14027 ; NeurIPS 2026 (MATH-AI) ; ins2i.cnrs.fr (MALINCA) ; renaissancephilanthropy.org ; arXiv :2511.20118 ; arXiv :2603.20405 ; mistral.ai 16/3/2026 ; arXiv :2605.29955

Des usages, des témoignages

- **Percolation.** Avril 2026 : un modèle prouve « $p_c < 1$ », preuve « *short and particularly elegant* » (H. Duminil-Copin). 28 août : $\theta(p_c) = 0$ en dimensions 3 à 10, preuve engendrée par un modèle et vérifiée en Lean.
- **Usage quotidien.** F. Marbach, « *Struggles and Sparks : Daily Math with LLMs* » ; I. Nourdin, expérience sur la méthode de Malliavin–Stein (Lyon, juin 2026).
- **Publication.** M. Lewin : traiter les preuves d'IA comme des données brutes, déposées à part avec modèle, requêtes et parties vérifiées.
- **Débat.** Une quinzaine de collègues français ont écrit sur le blog *Proofs and Prompts*.

> *« In my own research, I have chosen not to rely on artificial intelligence to replace my creative process. »*
>
> — H. Duminil-Copin, 30/8/2026

> *« We should use these new AI tools to do new things, rather than ten times more of what we were already doing. »*
>
> — M. Lewin, 2/11/2026

Sources : proofsandprompts.com (30/8 et 2/11/2026, liste des contributeurs) ; Quanta 8/2026 ; deglise.perso.math.cnrs.fr/ai-days.html

2025–2026 : une structuration déjà engagée

18 nov. 2025	Journée **AI 4 Maths** à l'IHP, organisée par la SFdS, la SMAI et la SMF
30 janv. 2026	Lancement du séminaire **Automath** à Paris (Geshkovski, Hayat, Lelarge, Marbach, Peyré) ; prochaine séance le 18 décembre
16–17 juin	Journées **IA, preuve et formalisation** à l'ENS de Lyon : plus de 70 inscrits, exposés et travaux pratiques Rocq et Lean
juin	CANUM : conférence plénière d'A. Hayat et mini-symposium « IA pour les mathématiques »
31 août–4 sept.	École d'été **Assistants de preuve et applications** à Strasbourg
depuis le 8 oct.	Séminaire en ligne **MIAM** de la SMF (Duminil-Copin, Ligozat, Charles, Wang)
21 oct.	Demi-journée à l'IHP : « *AI and the future of mathematics and physics* »
12 nov.	Première séance du **webinaire de l'Insmi** ; le 26 nov. : les thèses ; le 3 déc. : recrutement et évaluation

Sources : automath.dma.ens.fr ; deglise.perso.math.cnrs.fr/ai-days.html ; canum2026.math.cnrs.fr ; indico.math.cnrs.fr/event/16687 ; smf.emath.fr ; ihp.fr ; insmi.cnrs.fr

Les institutions ont commencé à répondre

déc. 2025	CNRS : ouverture de l'agent conversationnel **Emmy** (modèles Mistral) à tous les agents
15 sept. 2026	Bureau de la **SMF** : « sans doute une avancée majeure », mais « absence totale de stratégie collective pour l'avenir »
7 oct.	**Insmi** : éditorial de Ch. Besse ; la question est de savoir « quelles mathématiques nous voulons faire avec elle »
10 nov.	**Section 01** du Comité national : consignes pour le concours 2027

Réseaux thématiques. 16 à l'Insmi : aucun ne portait jusqu'ici sur l'IA pour les mathématiques ni sur la formalisation. À l'INS2I : GDR IFM et RADIA.

Concours chargés de recherche 2027

- le jury n'utilise pas d'IA générative
- ni l'usage ni le non-usage n'est pénalisé
- déclaration d'usage demandée
- 7 productions au plus
- entretien d'environ 30 min, contre 10 en 2026

Sources : next.ink (Emmy) ; smf.emath.fr 15/9/2026 ; insmi.cnrs.fr 7/10/2026 ; cn.math.cnrs.fr/Doc/Conseils_CR2027.pdf ; insmi.cnrs.fr (liste des réseaux, 6/11/2026)

5 / Le réseau IAMF : propositions et appel

Le réseau en bref

- **direction** : A. Garivier, M. Lelarge, A. Mahboubi ; adossé à l'UMPA (ENS de Lyon)
- **mathématiciens et informaticiens** : Paris et Saclay, Lyon et Saint-Étienne, Nantes, Rennes, Grenoble, Lille, Nice et Sophia, Strasbourg, Nancy, Bordeaux, Caen, Tours, Lens…
- **deux cultures**, Rocq et Lean, et des compétences en IA, en logique et en éthique

85 permanents et 35 laboratoires recensés dans le dossier (mars 2026) ; la liste est ouverte.

Comité scientifique

J. Commelin (Utrecht), F. Déglise (Lyon), J. Gómez-Serrano (Brown), A. Hayat (ENPC), M. Jamnik (Cambridge), M. Johansson (Chalmers), L. Kovács (Vienne), M. Lackenby (Oxford), B. Naskręcki (Poznań), G. Peyré (CNRS, ENS), C. Prunkl (Inria), S. Smets (Amsterdam), C. Tessier (ONERA), F. van Doorn (Bonn), G. Williamson (Sydney), J. Urban (Prague)

Formalisation, IA, mathématiques, mais aussi éthique et philosophie des sciences.

Deux axes

A. IA, formalisation et découverte

Faire des mathématiques avec des modèles d'IA (chercher, conjecturer, prouver), les assistants de preuve servant de garde-fous.

- modèles pour le raisonnement informel et formel
- traduction entre mathématiques rédigées et formalisées
- **données** : corpus, licences, représentations
- **évaluation** : protocoles robustes, généralisation, vérification formelle
- **intégration** aux outils existants : assistants de preuve, calcul formel et certifié, simulation
- **fiabilité, contrôle, traçabilité, sécurité** : un usage scientifique et pédagogique responsable

B. Formalisation et communauté

Les assistants de preuve et leurs bibliothèques comme objets de recherche, avec ou sans IA, et la communauté qui les fait vivre.

- **théorie** : fondements, méthodes de preuve, structuration des bibliothèques
- **logiciel** : ergonomie, performance, interopérabilité (Lean, Rocq…), reproductibilité
- **mutualisation**, documentation, bonnes pratiques
- **communauté** : formation de nouveaux utilisateurs, ouverture vers l'enseignement et d'autres disciplines, visibilité et pérennité

Quatre principes

- **Un lieu de dialogue et de cohésion.** La communauté est fortement éprouvée depuis plusieurs mois. Organiser des rencontres, structurer, former : c'est une fonction première du réseau.
- **Ouvert à tous**, quel que soit l'usage que chacun fait de l'IA, y compris aucun. L'axe B existe indépendamment de l'IA.
- **Indépendant.** Des espaces de discussion académiques, indépendants des acteurs industriels, pour une démarche rigoureuse, avec déclaration des liens d'intérêts et gestion explicite des conflits éventuels. Les collaborations avec les entreprises restent possibles ; les moyens : protocoles publics, plusieurs fournisseurs, outils ouverts.
- **Interdisciplinaire et complémentaire.** Mathématiques et informatique (Insmi, INS2I), ouvert à d'autres instituts ; le réseau apporte expérimentations et retours de terrain à la réflexion conduite avec l'institut et les sociétés savantes.

Actions : des pistes soumises à discussion

Animer

- journées thématiques, 3 à 4 par an
- soutien aux séminaires (Automath…)
- mobilité des jeunes chercheurs
- trimestre thématique AISSAI, oct.–déc. 2027

Former et accompagner

- ateliers Lean et Rocq pour non-spécialistes, hackathons
- compétences à **préserver**, évaluées sans assistance ; compétences pour **examiner et corriger** ce que produisent les outils
- **attentes explicites** entre doctorants et encadrants ; une voie sans IA

Évaluer et outiller

- évaluation indépendante des modèles, sur des problèmes choisis par la communauté
- audit des preuves formelles : fidélité des énoncés, axiomes, noyaux
- modèles ouverts : comparaison, coûts, empreinte
- formats de dépôt pour les productions d'IA

Donner un cadre

- recueil de bonnes pratiques, en lien avec les instances
- état des lieux des usages dans les laboratoires

Rien n'est arrêté : priorités fixées après consultation. Le réseau fournit des expériences et des ressources ; les règles des diplômes et des recrutements relèvent des instances.

Cinq chantiers et le réseau

1	**Pratique des mathématiques**	distinguer preuve formelle et preuve comprise : journées, retours d'expérience, audit des énoncés
2	**Publication et validation**	vérification formelle des résultats annoncés ; formats de dépôt pour les productions d'IA
3	**Formation et évaluation**	ateliers, écoles, appui aux enseignants et aux encadrants de thèse
4	**Accès et indépendance**	comparaison des modèles ouverts et propriétaires, mesure des coûts, recensement des besoins
5	**Organisation collective**	rendre visibles les compétences et les initiatives locales, les relier

Source : message de Ch. Besse aux laboratoires, octobre 2026 (colonne de gauche) ; propositions du réseau (colonne de droite)

Un modèle coopératif

Ce que nous vous demandons

- **une question, un besoin** : un résultat à comprendre, un outil à évaluer, une formation à construire
- **un interlocuteur** dans votre réseau ou votre laboratoire
- **vos suggestions et vos désaccords** : priorités, points d'alerte, place des collègues qui n'utilisent pas l'IA

Trois formats possibles

- comprendre et auditer un résultat ;
- comparer des outils sur une tâche scientifique délimitée ;
- construire une formation, avec et sans assistance.

IAMF apporte des compétences en outils formels, en IA et en protocoles d'évaluation ; ensemble : un atelier et une restitution courte, réutilisable par d'autres.

Prochaine étape proposée : identifier quelques binômes entre réseaux pour 2027.

Pour nous écrire



Aurélien Garivier
UMPA, ENS de Lyon
aurelien.garivier@ens-lyon.fr



Marc Lelarge
Inria, DI ENS
Marc.Lelarge@inria.fr



Assia Mahboubi
Inria, LS2N, Nantes
assia.mahboubi@inria.fr

Le réseau n’a pas encore d’adresse propre : écrivez-nous directement.