Intelligence artificielle et preuves formelles

Retour sur la *Septimana Mirabilis* de janvier 2026

Luc Pommeret
Ingénieur d'études, LISN (CNRS / Université Paris-Saclay)

M2 Lophisc, Université Paris 1 Panthéon-Sorbonne

Groupe ReFL
Réflexion sur les Fondements de la Logique

21 janvier 2026

Plan

I. Panorama des outils LLM

II. IA et formalisation/découverte mathématique

1. Critères de la découverte et thèse de Dean-Naibo
2. Hiérarchie logique
3. La *Septimana Mirabilis* (4–12 janvier 2026)

Conclusion

LLM et Transformers

Transformer (Vaswani et al., 2017) : architecture neuronale basée sur le mécanisme d'*attention*.

Principe : prédire le token suivant x_{t+1} étant donné le contexte $x_1, \ldots, x_t$.

LLM (Large Language Model) :

- Transformer à grande échelle (milliards de paramètres)
- Pré-entraîné sur des corpus massifs (web, livres, code)
- Capacités émergentes : raisonnement, few-shot learning, génération de code

Exemples :

$$\mathrm{GPT}_n\,(n \in \{3, 3.5, 4, 4.5, 5, 5.2\}), \quad \mathrm{Claude}_n\,(n \in \{3, 3.5, 3.7, 4, 4.5\}), \quad \mathrm{Gemini}_n\,(n = 1 + \tfrac{k}{2})$$

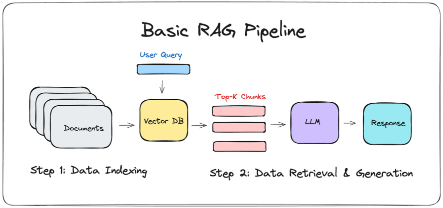
RAG (Retrieval-Augmented Generation)

Définition : architecture combinant un LLM et une base de connaissances externe.

Processus :

1. *Indexation* : le corpus est projeté dans un espace vectoriel
2. *Récupération* : recherche des documents pertinents par similarité
3. *Génération* : le modèle répond en s'appuyant sur le contexte récupéré

Objectif : réduire les hallucinations, étendre le contexte.



Agents (plus autonomes)

Définition : système où le LLM agit comme contrôleur cognitif capable de raisonnement chainé et d'interaction avec l'environnement pour atteindre un objectif abstrait.

Composants (Wang et al., 2024) :

- *Planification* : décomposition de tâches (Chain of Thought, Tree of Thoughts, et plus récemment Reasoning).
- *Mémoire* : court terme (contexte d'attention) et long terme (RAG).
- *Utilisation d'outils* : exécution de fonctions déterministes (API, interpréteur de code) via un schéma de *Function Calling* (eg. MCP).

Boucle : perception → décision/action → observation.

MCP (Model Context Protocol)

Définition : standard d'interopérabilité (Anthropic, 2024) normalisant l'interface entre un LLM (client) et des sources de données ou outils externes (serveurs).

Mécanisme :

- Architecture Client-Hôte-Serveur sur transport standard (stdio/SSE).
- Découverte dynamique de ressources (fichiers, bases de données) et d'outils (fonctions exécutables).
- Abstraction : découplage total entre le modèle génératif et l'implémentation (spécifique) des intégrations.

Permet au modèle de "monter" dynamiquement du contexte externe dans sa fenêtre d'attention.

Vibe Coding (Karpathy, février 2025)

Vibe Coding (LLM seul) :

- Génération de code/preuve informelle
- Risque d'hallucination

Assistants de preuve (Rocq, Lean) :

- Vérification formelle
- Le code est la preuve

Synthèse : LLM génère des esquisses, l'assistant vérifie et rejette les erreurs.

Ralph Loop (Huntley, juillet 2025)

Brute force du *vibe coding* :

`claude -p "/ralph-loop "écris ma thèse" -max-iterations` ∞
`-completion-promise "FINI"`

Critères de la découverte (Tao, 2025)

Une contribution de l'IA est une découverte si :

1. La preuve est *nouvelle*
2. Elle est *générée de manière autonome*
3. Elle est *formalisée* et vérifiée (Lean)

Dean et Naibo (2025)

"Artificial Intelligence and Inherent Mathematical Difficulty"
Philosophia Mathematica, juillet 2025.

Thèse (N0) : "Advances in computing theory or technology – inclusive of those currently understood to exemplify artificial intelligence – will not radically alter the following aspect of our current understanding of mathematics and its practice : Settling the status of open questions is an inherently difficult problem."

Thèse (N1) : "The application of artificial intelligence will not lead to a substantial reduction in the difficulty of one of the central goals of mathematical practice."

Arguments :

- La difficulté est structurelle (indécidabilité, NP-complétude)
- L'IA excelle dans la recherche (Σ^0_1), pas dans la conceptualisation

Complexité de découverte vs complexité de vérification

Complexité de vérification :

- Vérifier qu'une suite de symboles est une preuve valide
- Souvent polynomiale (P)

Complexité de découverte (Detlefsen, 1990) :

- Trouver le chemin vers la preuve
- Souvent exponentielle ou pire

La hiérarchie arithmétique reste la contrainte de l'IA (Dean et Naibo, 2025)

Σ_1^0 : $\exists x\, \phi(x)$

- Recherche existentielle
- Accessible à l'IA (recherche de témoin)

Π_1^0 : $\forall x\, \phi(x)$

- Propriété universelle
- Réfutation = trouver un contre-exemple (Σ_1^0)

Π_2^0 : $\forall x\, \exists y\, \phi(x, y)$

- Alternance de quantificateurs
- Nécessite un argument générique

La *Septimana Mirabilis* (4–12 janvier 2026)

4 janvier Kevin Barreto et “AcerFur” soumettent #728 à GPT-5.2 Pro. Génération d’une stratégie via le théorème de Kummer.

4–6 janvier Aristotle traduit en Lean. Correction autonome des lacunes.

8 janvier Terence Tao valide. Statut : **PROVED (LEAN)**.

11 janvier Neel Somani soumet #397. Réfutation le même jour. Statut : **DISPROVED (LEAN)**.

12 janvier Résolution de #729 par extension de la méthode de #728.

Problème #728 (Erdős-Graham-Ruzsa-Straus, 1975)

Énoncé (Prouvé) : pour toutes constantes $0 < C_1 < C_2$, il existe une infinité de triplets (a, b, n) tels que $a!b! \mid n!(a+b-n)!$ et :

$$C_1 \log n < a + b - n < C_2 \log n$$

Statut logique : Π_2^0 (universel-existentiel).

- Forme : $\forall C_1, C_2, \exists (a, b, n) \ldots$
- *Analyse* : résolu via réduction à un argument de densité (accessible aux méthodes probabilistes).

https://www.erdosproblems.com/728

Formalisation Lean du problème #728

```
theorem erdos_728 :
    -- pour epsilon suffisamment petit (voisinage de 0)
    forall_eventually eps : Real in nhds[>] 0,
    -- pour tout C > 0
    forall C > (0 : Real),
    -- pour tout C' > C
    forall C' > C,
      -- il existe a, b, n entiers naturels
      exists a b n : Nat,
        0 < n /\
        eps * n < a /\
        eps * n < b /\
        -- condition de divisibilite
        a ! * b ! | n ! * (a + b - n)! /\
        -- fenetre logarithmique
        a + b > n + C * log n /\
        a + b < n + C' * log n := by
  sorry -- preuve completee par Aristotle
```

Source : `github.com/google-deepmind/formal-conjectures`

Problème #397

Conjecture (réfutée) : l'équation

$$\prod_{i \in A} \binom{2m_i}{m_i} = \prod_{j \in B} \binom{2n_j}{n_j}$$

n'a qu'un nombre fini de solutions non triviales (Π_1^0).

Statut logique de la réfutation : Σ_1^0 (existentiel).

- Tâche : trouver un contre-exemple (famille infinie).
- *Méthode* : identification d'une identité algébrique méconnue.

https://www.erdosproblems.com/397

Problème #729

Énoncé : variante asymptotique avec contraintes relâchées sur les petits facteurs premiers.

Statut logique : Π_2^0 (asymptotique).

- Résultat : pour $a = \lfloor C \log n \rfloor$ avec C petit, la divisibilité tient pour "presque tout" n.
- *Méthode* : transfert de la technique "Carry-Rich" du #728.

`https://www.erdosproblems.com/729`

Problème #124 (controverse)

Problème original (Burr-Erdős-Graham-Li, 1996) : conditions strictes (PGCD, $\neq 1$). Difficile.

Résolution IA : variante simplifiée (sans condition PGCD).

- *Statut logique* : Σ_1^0 (dérivable).
- *Méthode* : application directe du critère de Brown.

Leçon : l'IA a trouvé le chemin de moindre résistance logique.
`https://www.erdosproblems.com/124`

Comparaison : Jie Ma (juillet 2025)

Amélioration exponentielle des bornes inférieures pour les nombres de Ramsey $r(\ell, C\ell)$:

$$r(\ell, C\ell) \geq (p_C^{-1/2} + \varepsilon)^{\ell}$$

Méthode : graphes géométriques aléatoires sur des sphères de haute dimension.

L'IA n'a pas encore démontré la capacité d'inventer un cadre théorique de ce type.

Bilan

Problèmes résolus (janvier 2026) :

- #728 : Π^0_2, méthode Kummer + densité
- #729 : Π^0_2, transfert
- #397 : Σ^0_1, contre-exemple
- #124 : variante seulement

Caractéristiques :

- Assemblage de modules existants (Kummer, méthode probabiliste)
- Preuves formalisées en Lean
- Validation humaine (Tao)

Le modèle copilote (Henkel, 2025)

"The Mathematician's Assistant", Math Semesterber 72, 117–144.

- L'humain reste responsable de la direction et de la validation
- L'IA gère les sous-tâches de faible complexité
- Compétence $\rightarrow$ formuler des prompts stratégiques

Questions

- La *Septimana Mirabilis* réfute-t-elle la thèse Dean-Naibo ?
- Quel avenir pour la recherche mathématique humaine ?
- La créativité mathématique est-elle une spécificité humaine ?

Références

- Dean, W. & Naibo, A. (2025). *Artificial Intelligence and Inherent Mathematical Difficulty*. Philosophia Mathematica.
- Henkel, J. (2025). *The Mathematician's Assistant*. Math Semesterber 72, 117–144.
- arXiv :2601.07421. *Resolution of Erdős Problem #728*.
- Romera-Paredes et al. (2024). *FunSearch*. Nature.