
ATELIER

AI Scientist



Antoine Salazar



Yennipena Jérémie Coulibaly



Yannis Bordji

14 Octobre 2026

L'IA devance le médaille Fields 2022

SCIENCES • MATHÉMATIQUES

Nouveau coup de massue d'OpenAI sur les mathématiciens : « Ils ont annihilé mon domaine de recherche »

Après les dernières annonces de l'entreprise américaine sur la résolution de plusieurs centaines de problèmes de maths, le médaillé Fields Hugo Duminil-Copin évoque un « démantèlement » de sa discipline. Il craint en particulier les conséquences de l'intelligence artificielle sur l'avenir des jeunes.

Par David Larousserie

Publié le 10 octobre 2026 à 06h30, modifié à 11h30 ·  Lecture 3 min. · [Read in English](#)

L'IA : un assistant de l'Homme contrôlé... par l'Homme

L'IA assiste l'humain au quotidien (génération de texte, de code...) mais répond exclusivement aux ordres et instructions synthétisés par l'Homme (prompts) :

01

Finance

Recherche en finance quantitative, analyse des risques, détection des fraudes et modélisation algorithmique des marchés.

02

Défense

Recherche technologique et stratégique, surveillance, planification logistique et aide à la décision tactique.

03

Santé

Recherche médicale et pharmaceutique, assistance au diagnostic, analyse d'imagerie et accélération des traitements.

Les Limites Fondamentales de l'IA

01

Dépendance aux Données

L'intelligence artificielle repose entièrement sur les données d'entraînement fournies. Ses performances et analyses dépendent directement de la qualité et du volume de ces informations initiales.

02

Absence de Conscience

Elle ne possède aucune conscience de ses propres connaissances. Elle manipule des motifs statistiques sans comprendre la réalité ou le sens de ce qu'elle génère.

03

Incapacité d'Auto-Mise à Jour

L'IA est incapable de mettre à jour ses connaissances de manière autonome en temps réel. Sa vision du monde reste figée à la date de fin de son apprentissage.

Jusqu'où une IA peut-elle réellement participer à la production de connaissances scientifiques ?

PLAN

- Les 2 grandes approches de l'IA (connexionniste, symbolique)
- Vers une IA consciente ?
- Méthode CAIA de Jacques PITRAT
- Et après PITRAT ?
- L'IA dans la recherche

Les approches dominantes de l'IA

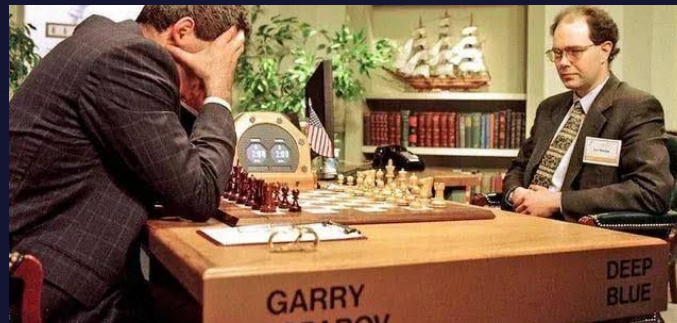
IA connexionniste

Approche de l'intelligence artificielle directement inspirée de l'architecture du cerveau biologique. Au lieu d'exécuter des règles logiques explicites dictées par un humain, elle repose sur des réseaux de neurones artificiels interconnectés.



IA symbolique

Approche historique de l'intelligence artificielle, fondée sur la logique pure, la modélisation des connaissances et la manipulation explicite de symboles.



Vers une IA consciente pour la recherche

Conscience

- KANT : « La conscience est une représentation qu'une autre représentation est en moi »
- Programme capable de s'examiner lui-même, d'analyser son propre code et d'observer son propre raisonnement en cours d'exécution.

Métaconnaissance

- La connaissance de la connaissance
- **Comment** utiliser, trier ou modifier cette connaissance de base.



Jacques PITRAT

1934 – 2019

Un héritage fondateur

Pionnier

Un des pionniers français de l'intelligence artificielle symbolique.

Métaconnaissance

Théoricien de la métaconnaissance en IA : représenter des connaissances sur les méthodes de raisonnement.

Ambition

Développer une IA capable de le remplacer intégralement dans ses activités de recherche.

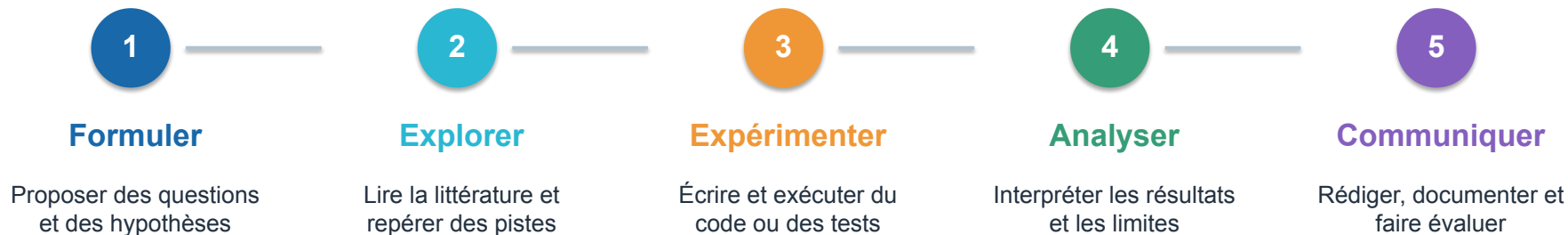
CAIA

Un projet de recherche emblématique, poursuivi sur plusieurs décennies.

Qu'appelle-t-on un « AI Scientist » ?

DÉFINITION

Selon Pitrat, il s'agit de tout système informatique autonome, capable de mener des recherches scientifiques et de s'améliorer par lui-même.



Problématique centrale

Une machine peut-elle faire de la science de manière autonome et avec quelles garanties de fiabilité, de transparence et de contrôle humain ?

Le précurseur : Jacques Pitrat et CAIA

Jacques Pitrat défendait une idée ambitieuse : déléguer à la machine non seulement l'exécution, mais aussi une partie de la conception de ses méthodes.

CAIA

Chercheur Artificiel en Intelligence Artificielle

Un projet de recherche développé sur plusieurs décennies, centré sur la résolution de problèmes logiques et combinatoires.

Une idée clé

La machine représente ses connaissances explicitement et peut raisonner sur les méthodes qu'elle utilise pour résoudre un problème.

Une ambition

Construire un système capable de générer des programmes, d'analyser ses échecs et d'améliorer certaines de ses procédures.

Les 4 piliers de l'approche CAIA

1 • Générer ses programmes

À partir des règles et du but, CAIA peut produire le code nécessaire, l'exécuter et corriger certaines erreurs.

2 • Métaconnaissances

Le système représente aussi des connaissances sur ses propres méthodes de résolution.

3 • Bootstrapping

Les procédures peuvent être améliorées et réutilisées : le système contribue à sa propre évolution.

4 • Déclaration

L'humain décrit les contraintes et l'objectif ; le système cherche une méthode pour parvenir à une solution.

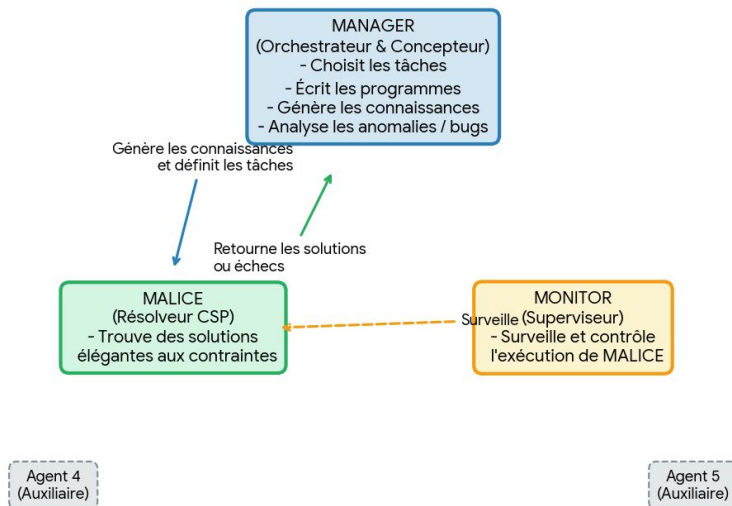
CAIA : Chercheur Artificiel en Intelligence Artificielle



Jean-Louis Laurière

<https://independent.academia.edu/LaurieuCAre>

Architecture Multi-Agents du Système CAIA (Jacques Pitrat)



Une organisation multi-agents pour répartir les rôles de décision, de mémoire et de résolution.

CAIA : apports réels et limites

Ce que le projet a illustré

- Génération automatique de programmes pour résoudre des problèmes formels.
- Repérage de symétries et adaptation de stratégies pour éviter des calculs inutiles.
- Expérimentations d'auto-amélioration sur certaines procédures.

Ce qui a freiné sa portée

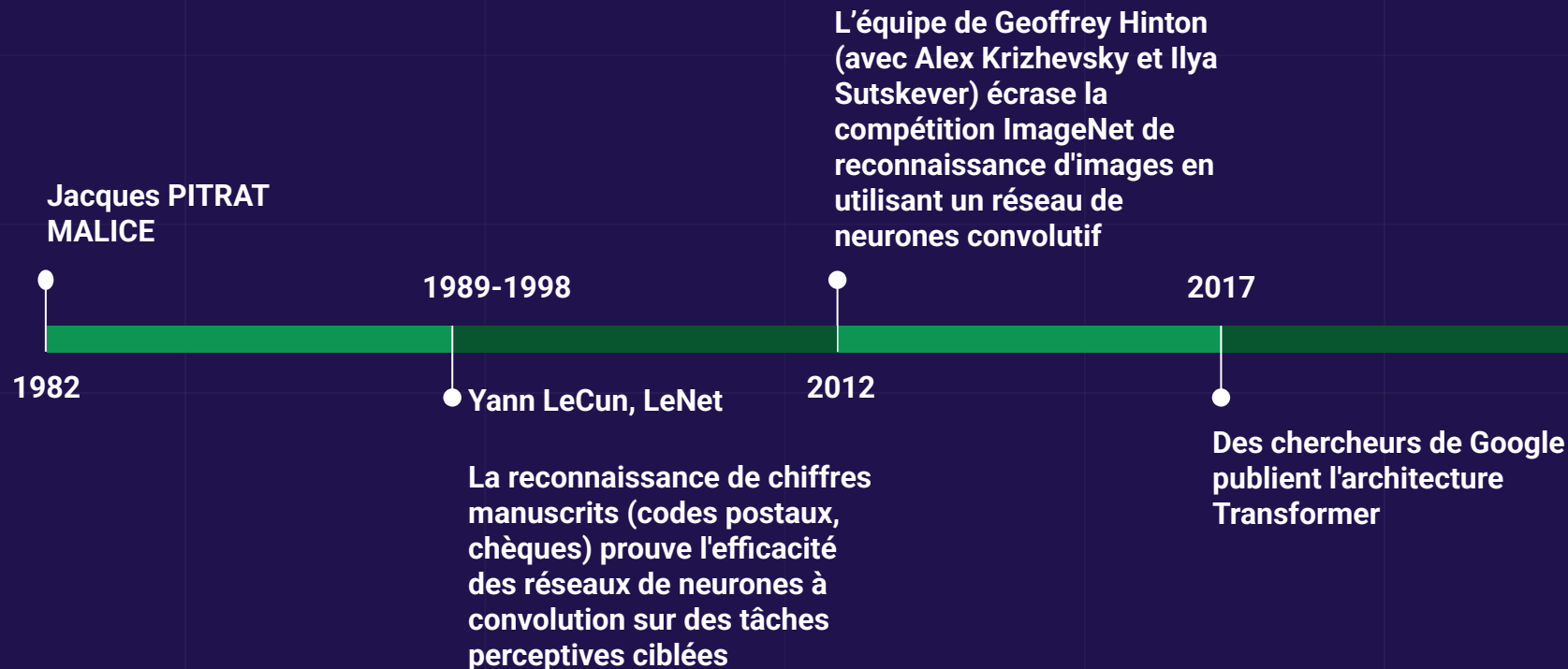
- Domaine restreint : logique et problèmes combinatoires.
- Formalisation des connaissances longue et coûteuse pour l'humain.
- Code complexe, difficile à maintenir et à transmettre.
- Puissance de calcul limitée à l'époque.

Bilan

CAIA a posé les bases et les méthodes à suivre pour rendre les machines autonomes, sans atteindre l'objectif d'un chercheur scientifique généraliste.

Et après PITRAT ?

Le désintérêt autour de l'IA symbolique



Et après PITRAT ?

Le problème de la boîte noire

IA Symbolique

Boîte transparente (Règles & Logique)

Entrée → [Règles / Logique] → Décision explicable

- **Fonctionnement** : Repose sur des symboles, la logique formelle et des règles explicites.
- **Avantage** : Explicabilité totale, chaque étape du raisonnement est traçable.
- **Limite** : Rigidité et difficulté à passer à l'échelle sur des données non structurées.

IA Connexionniste

Boîte noire (Deep Learning & Neurones)

Données → [? Boîte Noire ?] → Prédiction / Résultat

- **Fonctionnement** : Apprentissage automatique par réseaux de neurones distribués.
- **Avantage** : Haute performance sur la vision, le langage et les motifs complexes.
- **Limite** : Opacité, impossibilité d'expliquer précisément le cheminement interne.

Une alternative interprétable aux réseaux classiques

MLP Classique

- **Arêtes** : Poids scalaires fixes
 - **Nœuds** : Fonctions d'activation fixes
 - **Principe** : Combinaisons transmises à une activation fixe.
- boîte noire

Réseau KAN (Liu et al. (2024))

- **Arêtes** : Fonctions apprenables
 - **Nœuds** : Simples sommations
 - **Principe** : Décomposition de relations complexes en équations.
- Boîte lisible : extrait des lois mathématiques explicites.

D'un point de vu cognitif :

Le système reste dépourvu de métacognition, de réflexivité et de compréhension du sens des équations qu'il extrait.

L'IA dans la découverte scientifique

Encore Jacques PITRAT

À Montpellier, en 1986, Jacques Pitrat fait une conférence sur la métaconnaissance :

Chercheur du MIT

Comment gérez-vous le problème de la terminaison et de la preuve d'un programme ?

Jacques PITRAT

"Des gens qui boguent, il y en a plein les asiles !"



Faire de la recherche = des conjectures incertaines, des impasses, tourner en rond,...

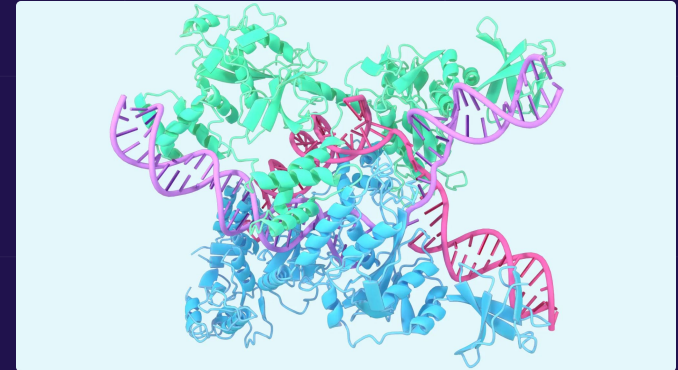
La pensée humaine ne dispose d'aucune garantie formelle de terminaison

AlphaFold (Google DeepMind 2018/2020)

Prédiction très précise de la structure 3D des protéines à partir de leur séquence d'acides aminés grâce aux réseaux de neurones.

Résultats :

- **En 2021** : Publication détaillée de l'architecture dans Nature et mise à disposition publique de la base de données.
- **En 2024** : Attribution du prix Nobel de chimie à Demis Hassabis et John Jumper pour leurs travaux sur la prédiction de la structure des protéines via AlphaFold



Google DeepMind

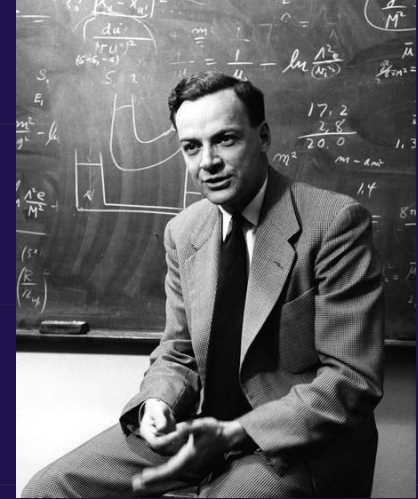
Une précision atomique mais toujours pas de métacognition...

AI Feynman (MIT / Udrescu & Tegmark 2020)

IA autonome conçue pour découvrir de manière autonome des lois/équations physiques à partir de données expérimentales et d'un réseau de neurones.

Résultats :

- Le système a été évalué sur 100 équations fondamentales tirées des célèbres cours de physique de Feynman
- Il a redécouvert de manière exacte 90 équations sur 100 sans connaissance préalable des formules.



Richard Feynman, Caltech

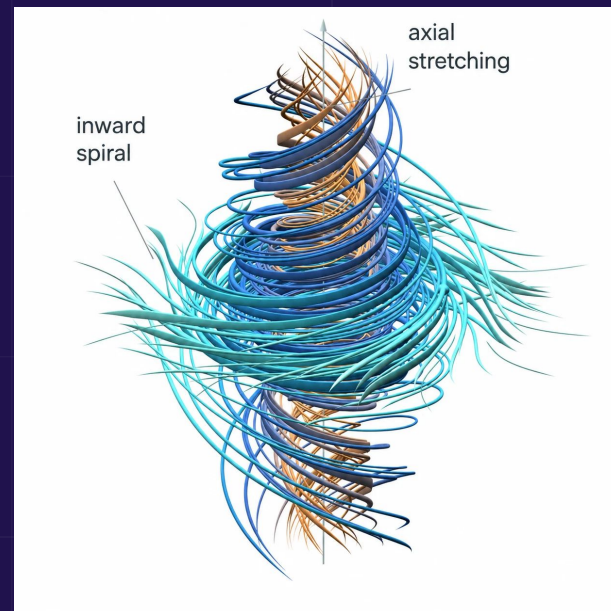
Résultats explicables (formules) mais toujours pas de métacognition ou de compréhension sémantique

L'IA dans la découverte scientifique

OpenAI 2026

Depuis le 8 Septembre 2026, OpenAI affirme avoir un modèle pouvant résoudre des problèmes mathématiques encore ouverts.

- **Vérification formelle sous Lean**
- **Une métacognition illusoire**



Problème de Navier-Stokes, OpenAI

Questions du débat

DISCUSSION

SCIENCE

Un article produit par une IA et validé par les pairs constitue-t-il une contribution scientifique légitime ?

COGNITION

L'autonomie scientifique exige-t-elle une conscience, ou une capacité fiable d'auto-évaluation suffit-elle ?

ÉTHIQUE

Comment éviter la saturation de la littérature par des articles générés en masse ?

CONTRÔLE

Qui choisit les objectifs de recherche et qui est responsable des erreurs ?

TECHNIQUE

Une approche hybride, mêlant règles explicites et modèles de langage, pourrait-elle réduire les erreurs ?

Conclusion

L'AI Scientist prolonge une ambition ancienne : faire participer la machine à la production de connaissances, et pas seulement à l'exécution de tâches.

L'héritage de Pitrat

Rendre explicites les méthodes et permettre au système de les examiner.

La promesse actuelle

Accélérer la lecture, la programmation et l'exploration d'hypothèses.

La condition

Vérifier les résultats, documenter les limites et garder une supervision humaine.

Références de départ

- Pitrat, J. (2009). Artificial Beings: The conscience of a conscious machine. Wiley-ISTE.
- Sabah, G. (2022). Conscience et Intelligence artificielle(s) vues par Jacques Pitrat. ROIA.
- Lu, C. et al. (2024). The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery. arXiv.
- Mitchener, L. et al. (2025). Kosmos: An AI Scientist for Autonomous Discovery. arXiv.