# AI Scientist

Présenté par AUBOURG Thomas, DEMEUDE Edgar, THIEBAUD Enzo et VU Anh Duy

# Introduction

# Objectifs : Vers une IA scientifique autonome

### Exploration autonome

Capacités d'hypothèse, raisonnement sophistiqué et auto-correction

### Interprétabilité centrale

Comprendre et expliquer les mécanismes, pas seulement produire des résultats

# Enjeux scientifiques

## Accélération de la découverte

- Automatiser la recherche expérimentale
- Explorer des domaines où la combinatoire dépasse les capacités humaines

### Chimie

Millions de combinaisons moléculaires possibles

### Physique des matériaux

Propriétés émergentes complexes

### Biologie moléculaire

Interactions protéiques multidimensionnelles

# Enjeux cognitifs

### Compréhension profonde

Au-delà du pattern matching vers la saisie conceptuelle

### Curiosité artificielle

Générer des questions pertinentes, explorer l'inconnu

### Métacognition

Réfléchir sur son propre raisonnement et ses limites

# Enjeux éthiques et épistémologiques

*"Science sans conscience n'est que ruine de l'âme."*

— François Rabelais (1494-1553)

**Le paradoxe du résultat**

Veut-on des résultats ou le plaisir de réfléchir pour les trouver ?

**Acceptabilité**

GPT en co-auteur d'articles : jusqu'où accepter cette collaboration ?

**Illusion de compréhension**

Quelles exigences d'explicabilité pour les découvertes par IA ?

**Notion de sacrifice**

Veut-on privilégier les scientifiques IA au détriment des scientifiques humains ?

Problématique et approches

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

---

## Deux grandes voies depuis les années 1970 :

### Approche symbolique

Raisonnement explicite et formalisation (top-down)

### Approche connexionniste

Apprentissage à partir des données (bottom-up)

Ces deux traditions, longtemps opposées, convergent aujourd'hui vers des **approches neuro-symboliques** prometteuses.

# Plan



# I - L'approche symbolique

**Raisonner pour comprendre**

# L’idée générale

- **fondée sur la logique et les règles**
- **Manipule des symboles pour raisonner**
- **Objectif :**
  - **comprendre et expliquer ses décisions**
- **Basée sur des connaissances explicites**

# Jacques Pitrat & CAIA

- **CAIA = Chercheur Artificiel en Intelligence Artificielle**

- **Apprend, s’auto-corrige, expérimente**

- **Composé de plusieurs agents :**
  - **MALICE – résout les problèmes**
  - **MONITOR – surveille**
  - **MANAGER – planifie**
  - **ADVISOR – évalue**
  - **ZEUS – contrôle**

# RefPerSy et MARS

- **RefPerSy = Reflective Persistent System**
- **Observe ses actions et s’adapte**
- **Garde en mémoire ses apprentissages**
- **Exemple : MARS (Middleware for Adaptive Reflective Systems)**
  - **Système capable d’ajuster son comportement selon son environnement**
  - **Inspiré par la réflexivité et la persistance**

# Conscience et Métaconnaissance

- **Métaconnaissance = réfléchir sur sa propre connaissance**
- **Conscience = comprendre ses propres actions**
- **Sert à planifier, contrôler, s’améliorer**
- **Vers une conscience artificielle fonctionnelle, utile et autonome**

# Critiques de l’approche symbolique

- **Forces :**
  - **Raisonnement clair et explicable**
  - **Bonne compréhension des processus**

- **Limites :**
  - **Trop complexe à formaliser**
  - **Peu adaptable à des situations nouvelles**

- **Aujourd’hui :**
  - **on tend vers une IA hybride (symbolique + Deep Learning)**

# Conclusion

- **L’approche symbolique cherche à comprendre et expliquer les raisonnements.**
- **Elle repose sur la logique, les règles et la métaconnaissance.**
- **Mais ces systèmes restent rigides et demandent beaucoup de connaissances humaines.**
- **Pour surmonter ces limites, une autre approche est apparue :**
  - **les réseaux de neurones, capables d’apprendre à partir de données sans règles prédéfinies.**

# II - L’approche connexioniste

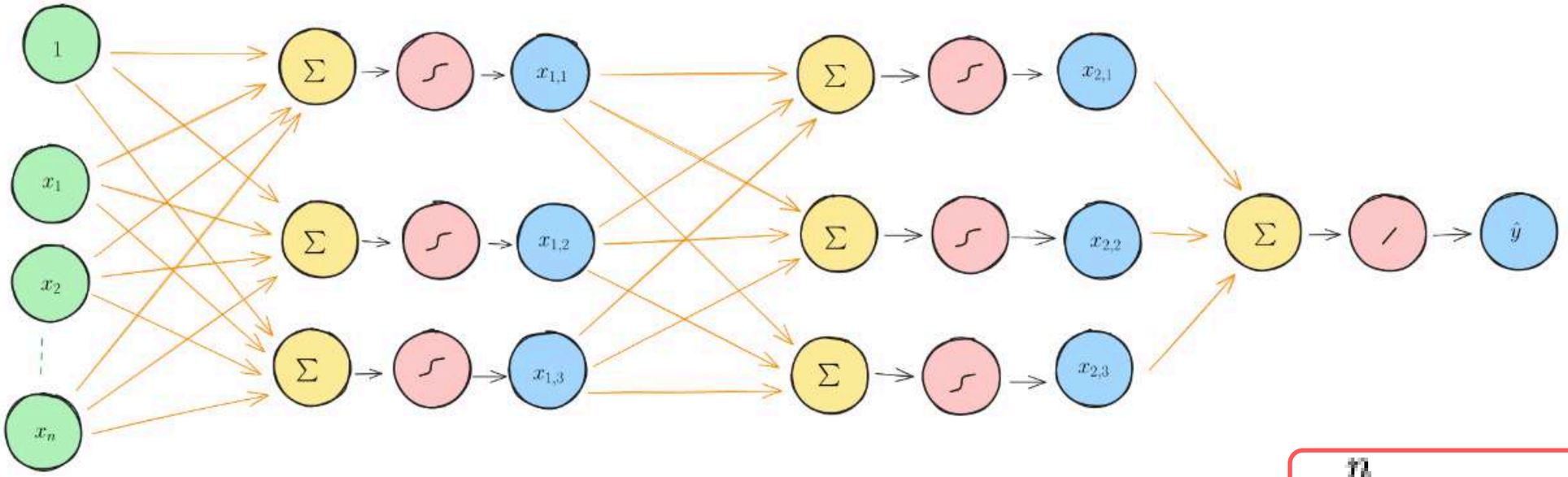
**Apprendre à découvrir**

# A - Les réseaux de neuronnes “purs et durs”

---

# B - Vers plus d’interprétabilité - Kolmogorov-Arnold Networks (KAN)

# A - Les réseaux de neuronnes "purs et durs"




$$y_i = \sigma\left(\sum_{j=1}^{n} w_{ij}x_j + b_i\right)$$

## Universal Approximation Theorem (UAT)

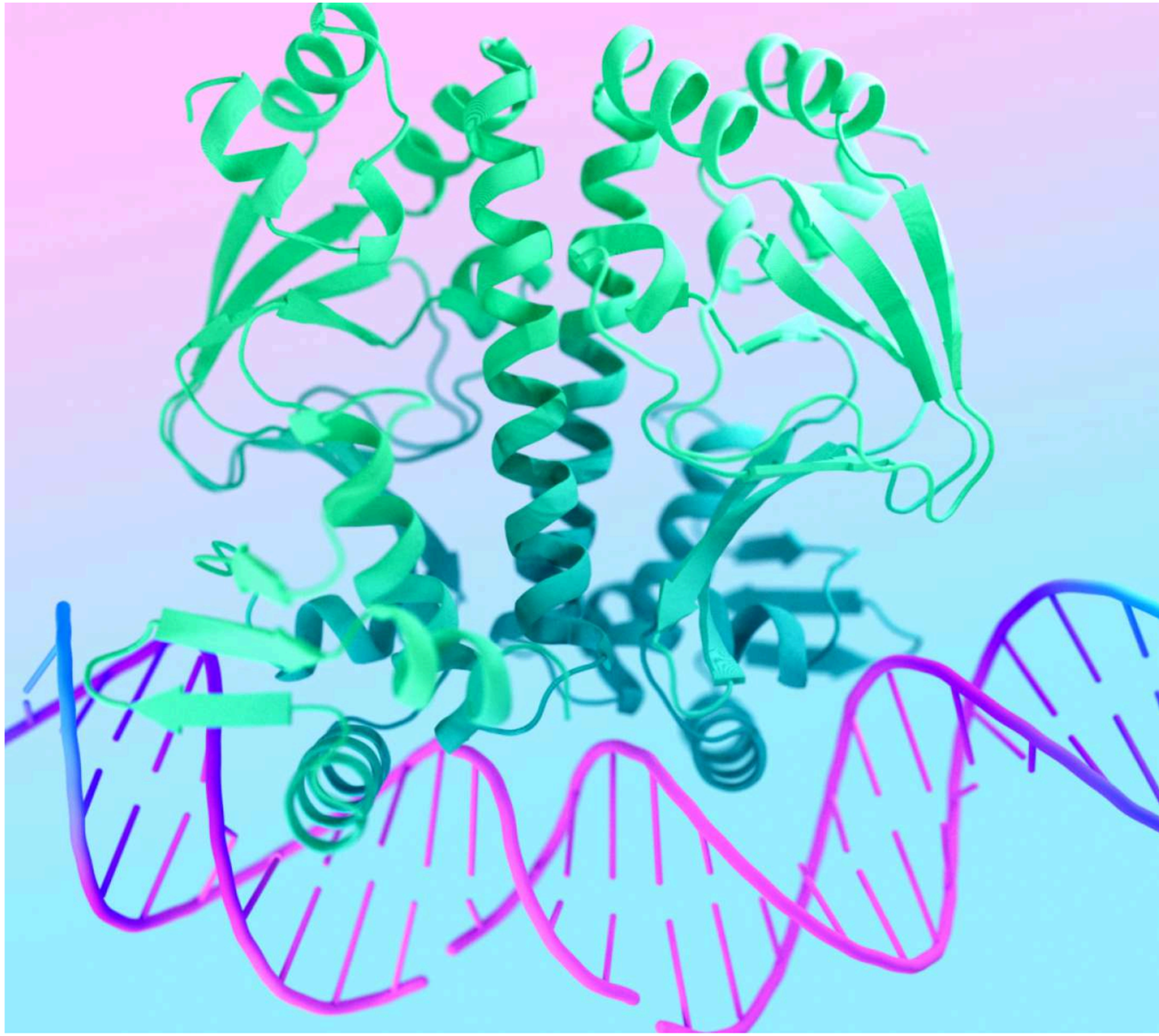
*Le Théorème d'Approximation Universelle (AUT) établit qu'un réseau de neurones à deux couches peut* ***approximer n'importe quelle fonction continue*** *à une erreur ϵ près.*

*Image : T. Deschamps (2025), Cours de Deep Learning, Université Lyon 1.*

# AlphaFold, un tournant pour l'IA scientifique

*(DeepMind, 2018/2020)*

*AlphaFold est un système d'intelligence artificielle basé sur des réseaux de neurones, qui a résolu le problème de la prédiction très précise de la structure 3D des protéines à partir de leur séquence d'acides aminés.*



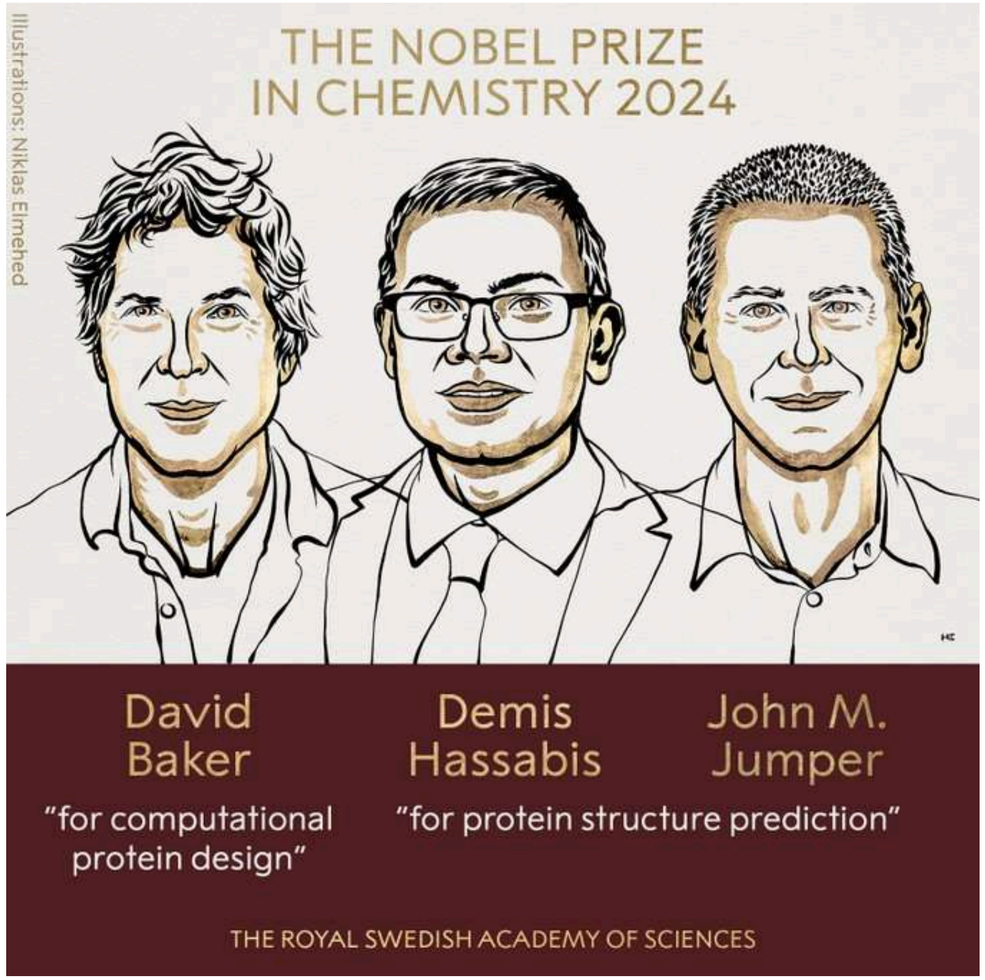
*Jumper, J., Hassabis, D, Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., …*
*Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873), 583-589. https://doi.org/10.1038/s41586-021-03819-2*

# **AlphaFold**, *Google DeepMind*

## Critical Assessment of Structure Prediction (CASP)

*Concours communautaire où des groupes de recherche doivent prédire les **structures tridimensionnelles** à partir de **séquences protéiques.***

| | **Scores moyen GDT_TS** | **Nombre approximatif de structures connues** |
|---|---|---|
| ***Meilleure avant AlphaFold*** | ***~75** (≈ 0.3-0.4 nm)* | **~170 000** |
| **AlphaFold 2** | **92.4** (< 0.1 nm) | **> 230 millions** |




*Rives, A. et al. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. Proc. Natl Acad. Sci. USA 118, e2016239118 (2021)*

*Durairaj, J. et al. Uncovering new families and folds in the natural protein universe. Nature 622, 646–653 (2023). https://www.nature.com/articles/s41586-023-06622-3*

# Critique de AlphaFold et de l'approche 100% MLP

## Réussites

+ **Génération de prédictions fiables et rapides**
  Accélération massive de la recherche en biologie structurale
+ **Résultats empiriques**
+ **Découverte de patterns implicites**
  Identification de motifs et relations invisibles à l'œil nu ou aux méthodes classiques
+ **Approximation de fonctions inconnues**
  Modélisation de systèmes non linéaires ou mal compris par des équations analytiques

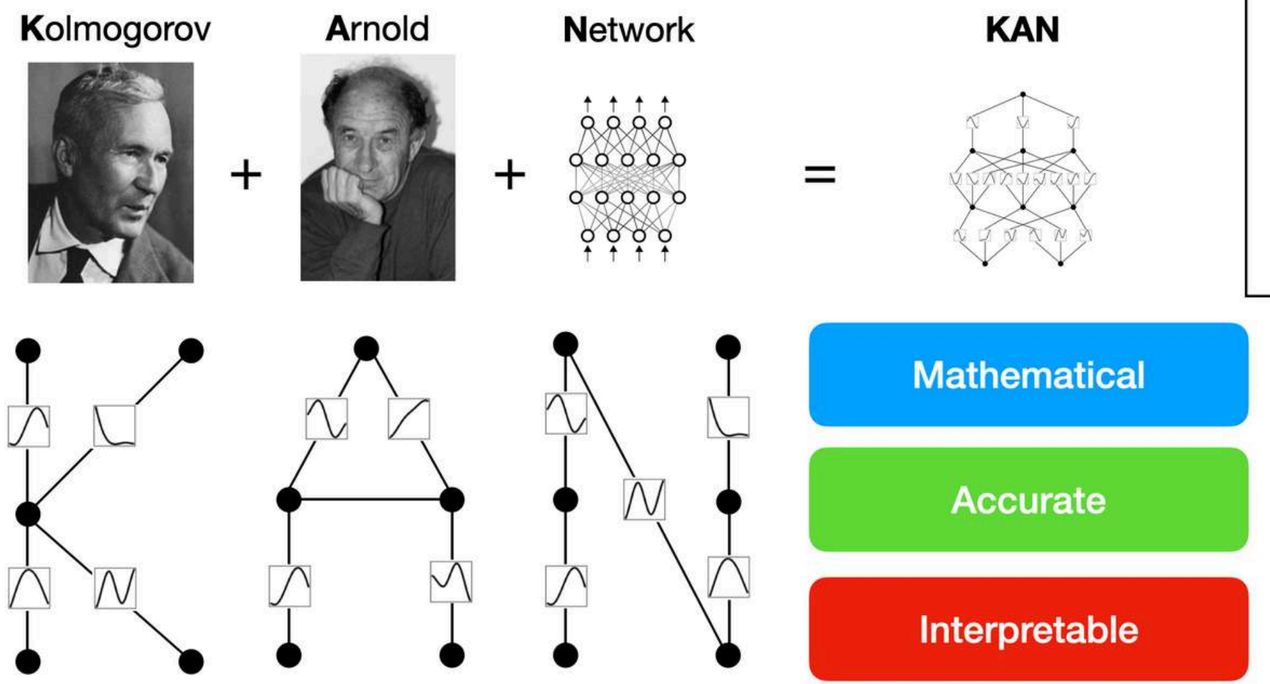
## Limites

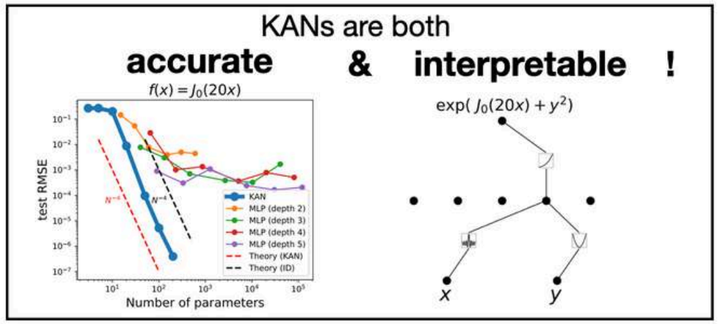
− **Limité aux données**
  N'arrive pas à proposer des protéines "orphelines"
− **Pas de raisonnement causal**
  *Découvertes empiriques mais non explicatives*
− **Pas de métacognition**
  *Pas de réflexivité ni d'auto-correction*
− **Opacité**
  *Boîte noire non interprétable*

# B - Vers plus d'interprétabilité

## Kolmogorov-Arnold Networks (KAN)

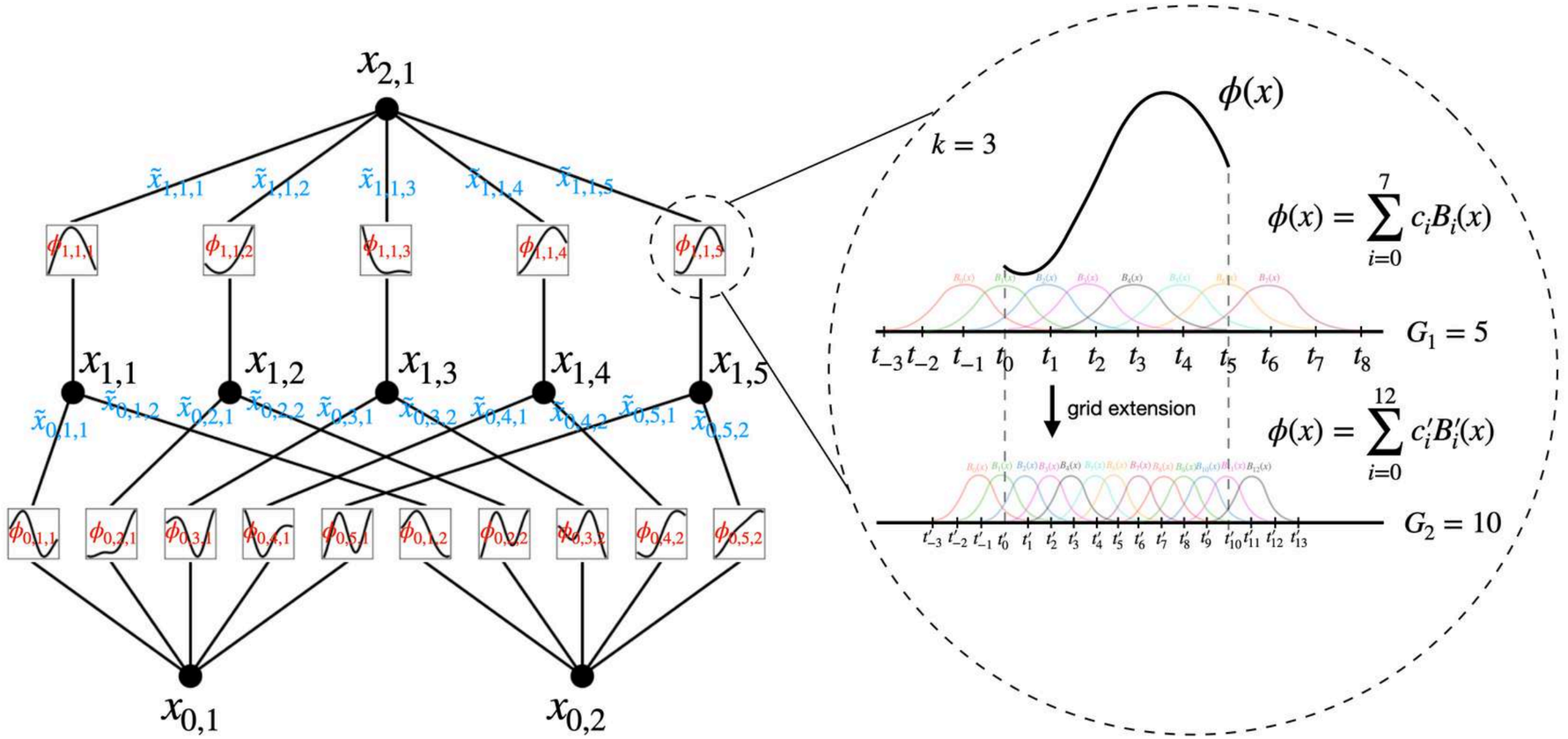
***Motivation :*** *Rendre les réseaux neuronaux plus interprétables et plus efficaces pour modéliser des fonctions scientifiques.*




*Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljačić, M., Hou, T. Y., & Tegmark, M. (2025). KAN : Kolmogorov-Arnold Networks (No. arXiv:2404.19756). arXiv. https://arxiv.org/html/2404.19756v5#S4*

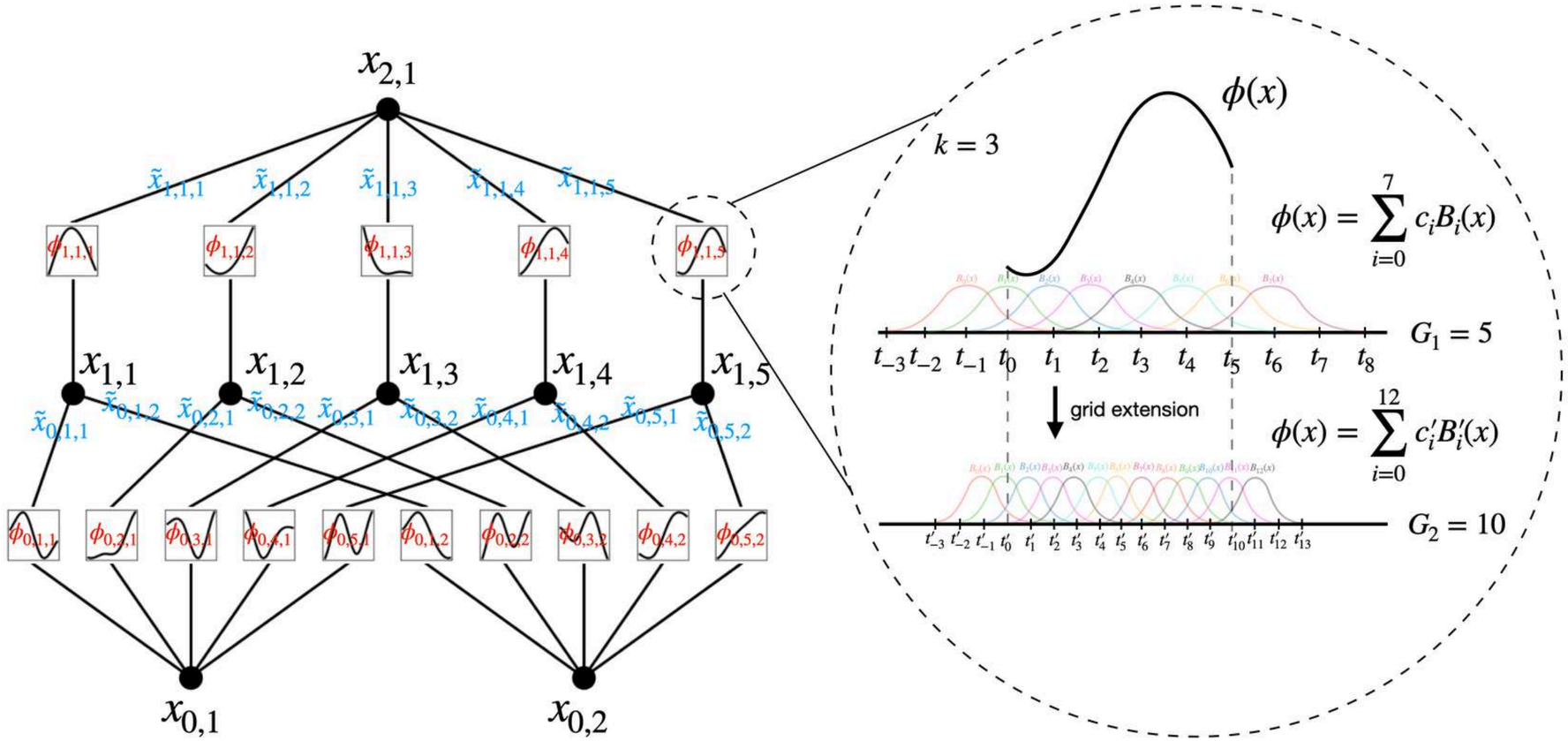
***Résumé :*** *les KANs cherchent à remplacer les* ***opérations de pondération (linéaires) et d'activation*** *dans les MLPs par des* ***fonctions d'activation apprenables (paramétrées par des B-splines) placées sur les arêtes****.*




*Alex Davies, Petar Veli˘ckovi´c, Lars Buesing, Sam Blackwell, Daniel Zheng, Nenad Tomašev,Richard Tanburn, Peter Battaglia, Charles Blundell, András Juhász, et al. Advancing mathematics by guiding human intuition with ai. Nature, 600(7887):70–74, 2021. https://www.nature.com/articles/s41586-021-04086-x*

**Résultat en Théorie des Noeuds:**

Dans des essais sur la classification de la signature des nœuds, les KANs ont surpassé le MLP de DeepMind en atteignant une **précision de test de 81,6 %** (contre 78,0 % du MLP).
Réalisé avec une **architecture extrêmement petite** (environ **$2\times10^2$ paramètres pour KAN** contre **$3\times10^5$ pour le MLP**)




*Alex Davies, Petar Veli˘ckovi´c, Lars Buesing, Sam Blackwell, Daniel Zheng, Nenad Tomašev,Richard Tanburn, Peter Battaglia, Charles Blundell, András Juhász, et al. Advancing mathematics by guiding human intuition with ai. Nature, 600(7887):70–74, 2021. https://www.nature.com/articles/s41586-021-04086-x*

# Critique des KAN

## Réussites

+ **Résultats empiriques proches du MLP**
+ **Révéler les structures compositionnelles**
  $y \approx f(x_1, x_2, \cdots, x_d)$
+ **Découvrir les relations structurelles**
  $f(x_1, x_2, \cdots, x_d) \approx 0.$
+ **Modèles plus petit que MLP**
  *Plus de pouvoir expressif sur chaque paramètre*

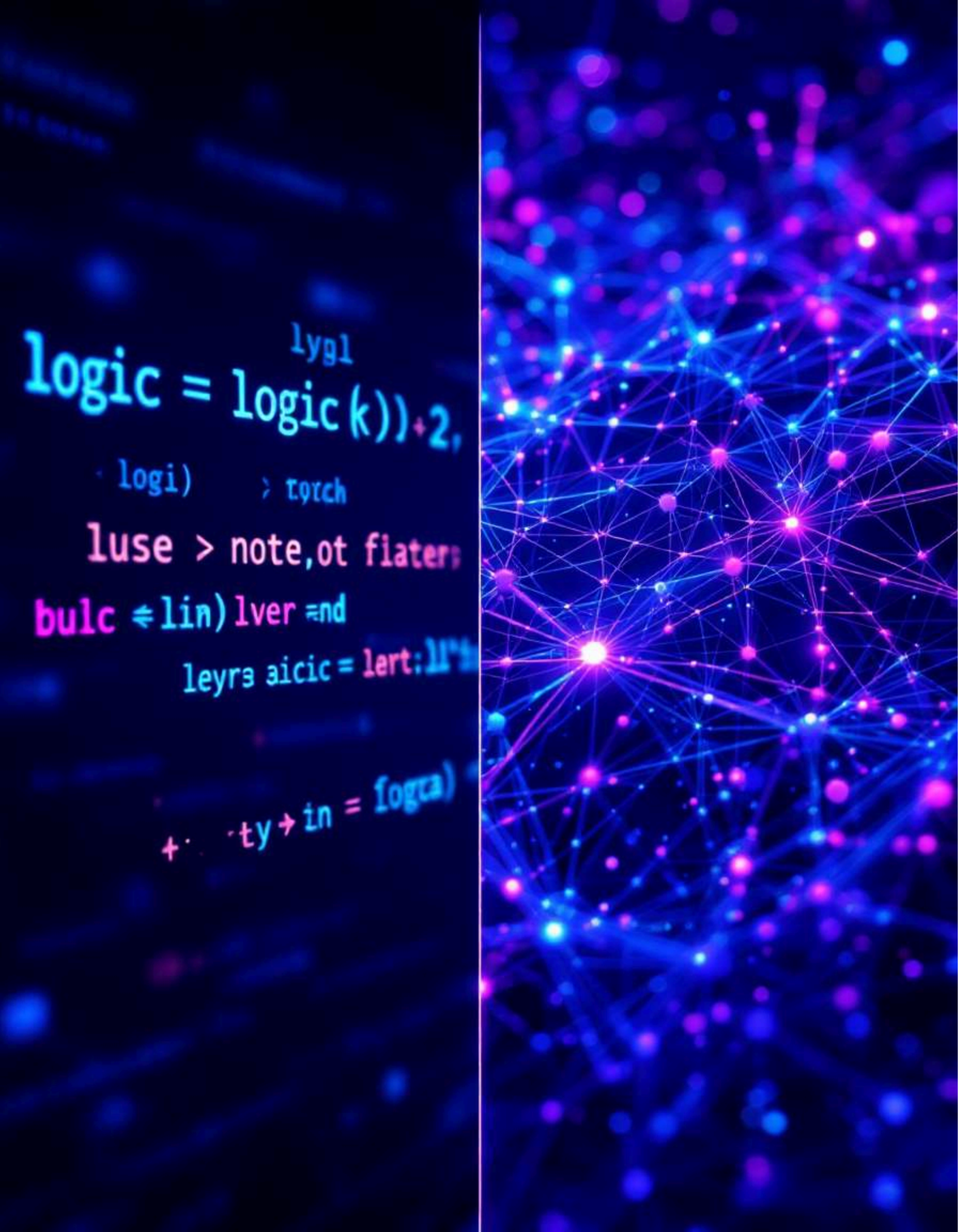
## Limites

− **Plus lent qu'un MLP classique**
  Pas de batch possible pour l'instant
− **Théories mathématiques**
  *Question sur la profondeur des réseaux*
− **Pas d'interprétabilité globale**
  *Uniquement locale*
− **Pas de métacognition**
  *Pas de réflexivité ni d'auto-correction*

*Alex Davies, Petar Veli˘ckovi´c, Lars Buesing, Sam Blackwell, Daniel Zheng, Nenad Tomašev,Richard Tanburn, Peter Battaglia, Charles Blundell, András Juhász, et al. Advancing mathematics by guiding human intuition with ai. Nature, 600(7887):70–74, 2021. https://www.nature.com/articles/s41586-021-04086-x*

# III - L'approche neuro-symbolique

**Raisonner et apprendre**



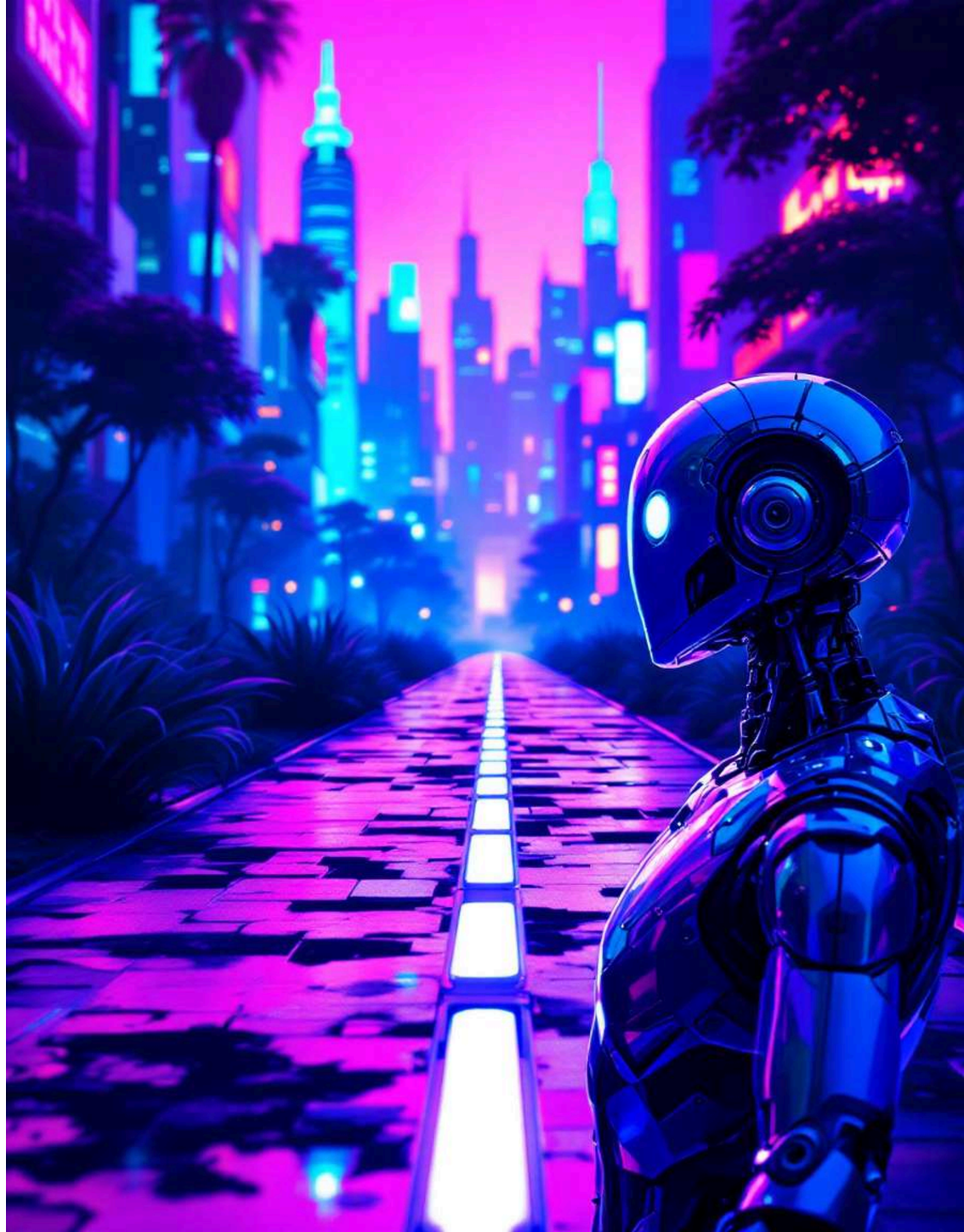

# Deux Paradigmes, Deux Limites

| Approche Symbolique | Approche Neuronale |
|---|---|
| Règles explicites, logique formelle | Apprend de vastes jeux de données |
| Raisonnement fort, explicable | Excellente reconnaissance de motifs |
| Adaptabilité limitée aux nouvelles données | Prédictions difficiles à expliquer |
| Cadres rigides définis par l'homme | Souvent opaque : décisions "boîte noire" |

*Aucune approche seule n'offre apprentissage profond et compréhension réelle, d'où le défi de l'intégration neuro-symbolique.*

# The Core Problem

L'IA doit à la fois prédire avec précision et expliquer ses décisions. Les systèmes symboliques raisonnent mais n'apprennent pas; les systèmes neuronaux apprennent mais n'expliquent pas.



# Neuro-Symbolique : La Solution

La neuro-symbolique unit IA : les réseaux neuronaux apprennent les modèles, le raisonnement symbolique crée un savoir interprétable.

1. Apprentissage Neuronal

   Détection de modèles de données.

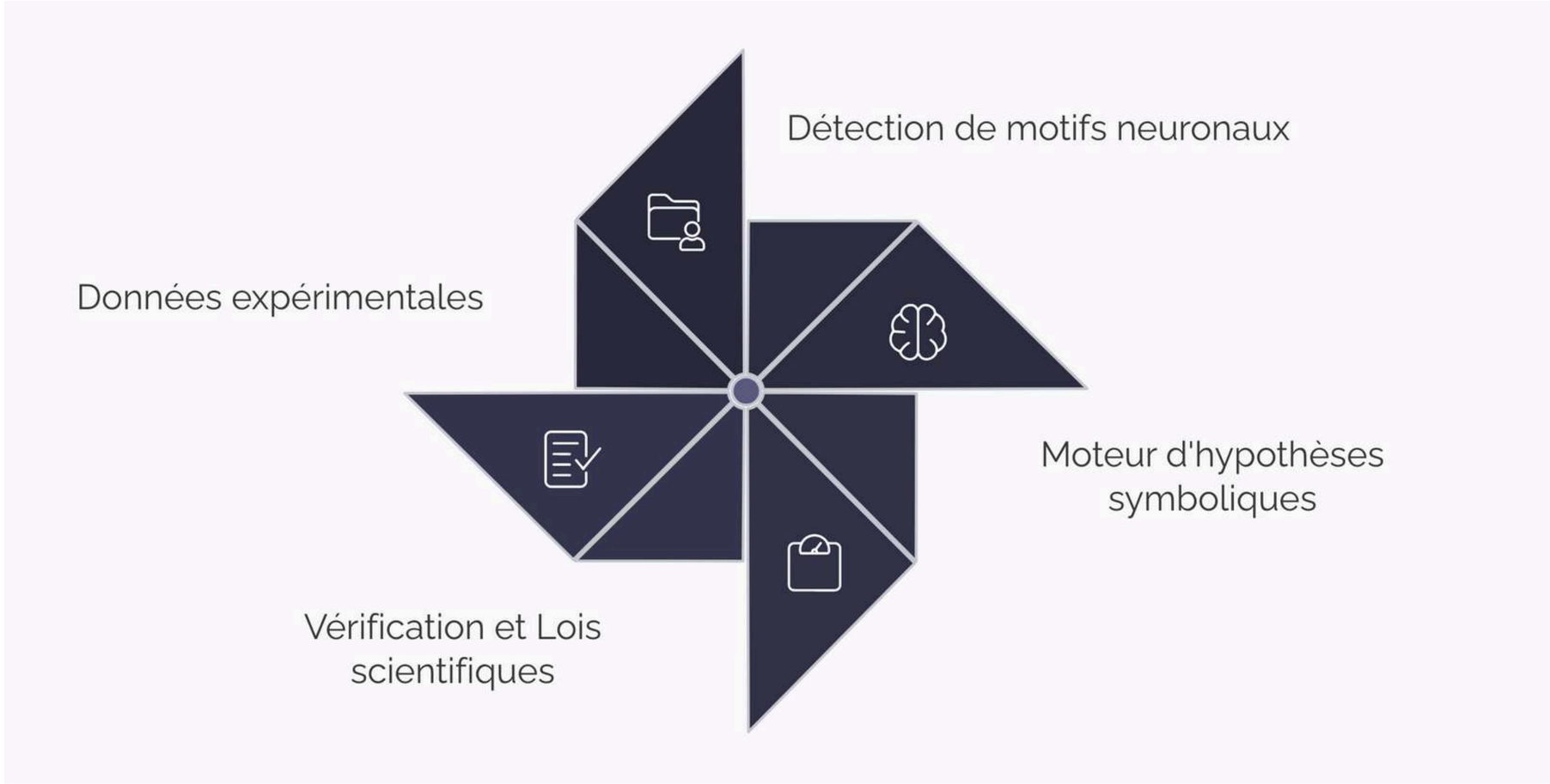
2. Extraction de Connaissances

   Conversion en règles et équations.

3. Connaissances Explicites

   Compréhension scientifique interprétable.

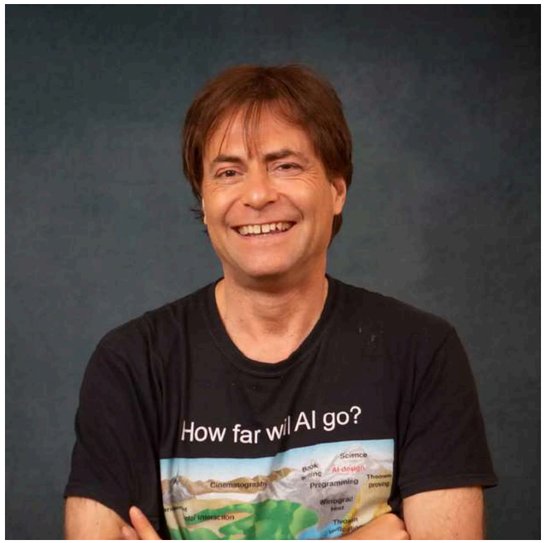
# Le pipeline Données-Découverte



# AI Feynman : Une Étude de Cas Révolutionnaire

AI Feynman (MIT, 2020) : une IA neuro-symbolique qui a redécouvert les lois de la physique à partir de données brutes, sans équations préétablies.

Silviu-Marian Udrescu




Max Tegmark


# Comment AI Feynman fonctionne : Trois phases

**Phase 1: Approximation neuronale**

Le réseau neuronal détecte motifs, symétries et propriétés mathématiques.

**Phase 2: Recherche symbolique**

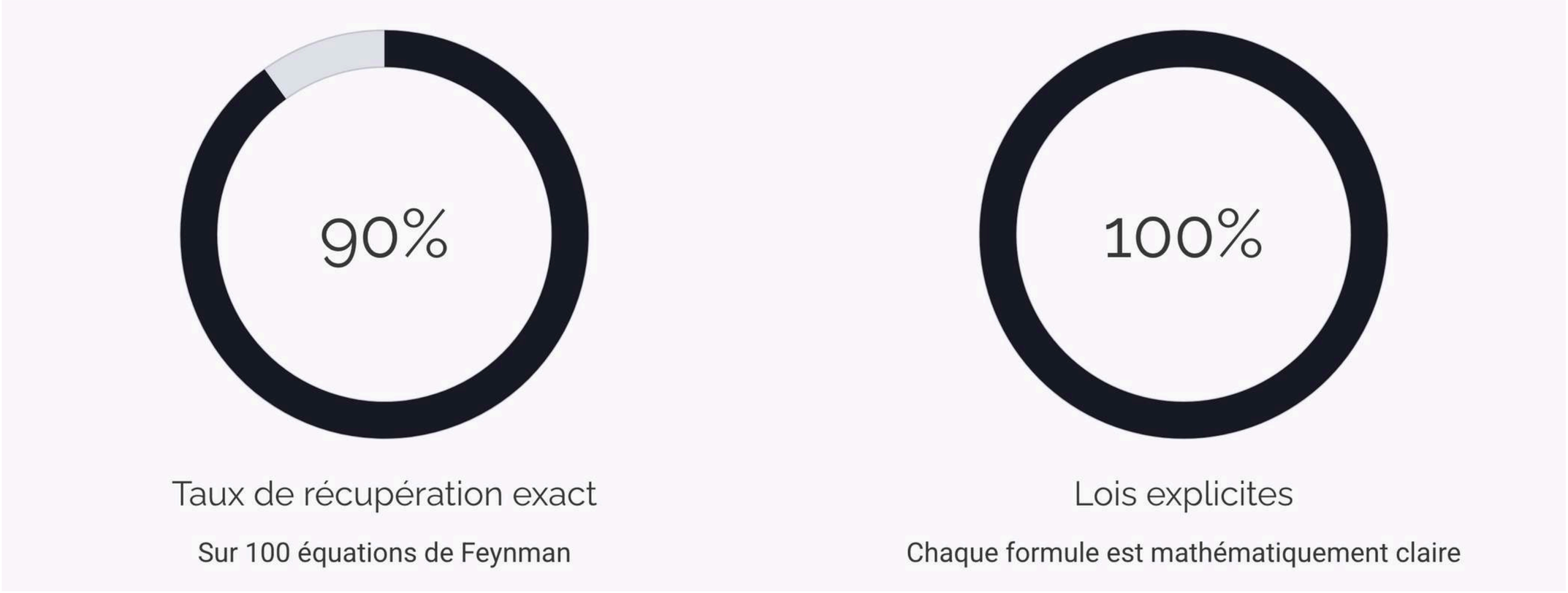
Un moteur algébrique explore les équations possibles. Les données neuronales filtrent les formules impossibles.

**Phase 3: Affinage itératif**

Les modules neuronal et symbolique collaborent pour tester les hypothèses et affiner la loi exacte.

# Succès Empirique : Redécouverte de la Physique

AI Feynman a retrouvé 90 équations sur 100 des célèbres cours de physique de Feynman, souvent plus vite que les physiciens humains.



# Le Chemin à Suivre : L'Intelligence Unifiée

L'IA neuro-symbolique représente une transition fondamentale, combinant la puissance d'apprentissage des réseaux neuronaux avec la clarté du raisonnement symbolique. Cela permet des systèmes capables d'une véritable perspicacité scientifique, découvrant les principes les plus profonds du monde, au-delà de la simple analyse.

L'avenir de l'IA n'est ni purement neuronal ni purement symbolique, mais un dialogue intelligent entre les deux—une synthèse produisant du savoir, pas seulement des schémas

# Critiques du neuro-symbolisme

## Selon Yang et al. (2025), "Neuro-Symbolic Artificial Intelligence: Towards Improving the Reasoning Abilities of Large Language Models"

*https://arxiv.org/pdf/2508.13678*

# Architectures hybrides avancées

## Difficulté d'intégrer apprentissage massif et raisonnement formel

Les approches neuronales restent fondamentalement statistiques et corrélatives, pas logiques.

### Objectif des architectures neuro-symboliques

- Puissance d'apprentissage à grande échelle des réseaux de neurones (big data, flexibilité)
- Rigueur du raisonnement symbolique (logique, preuve, consistance)

### Problèmes d'intégration

- Manque de cohérence structurelle entre les deux paradigmes
- Problèmes de scalabilité et d'efficacité (ralentissement des calculs, optimisations complexes)
- Architectures souvent "emboîtées" plutôt que fusionnées

Le vrai défi n'est pas de coller un moteur logique sur un LLM, mais de créer une architecture cohérente et différentiable où apprentissage et raisonnement se renforcent mutuellement.

# Compréhension Théorique

## Manque de fondements théoriques solides

Il n'existe aujourd'hui aucun cadre théorique clair pour expliquer pourquoi (et quand) l'intégration symbolique améliore réellement les capacités de raisonnement d'un modèle neuronal.

- Théorie de la généralisation des modèles hybrides
- Formulation de la fonction d'optimisation (LLM + feedback symbolique)
- Analyse des "raccourcis de raisonnement" (reasoning shortcuts)
- Le lien entre lois de scaling et capacité logique

Sans théorie solide, le neuro-symbolisme reste une collection d'expériences empiriques.

# Conclusion

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

# Êtes-vous pour ou contre la démocratisation des scientifiques IA ?

**Si la science a pour but d’apporter des réponses objectives, peut-on réellement dissocier un résultat scientifique de son auteur ?**

# L’émergence des intelligences artificielles scientifiques risque-t-elle, à long terme, d’appauvrir la créativité et le renouveau scientifique ?

# L’objectivité scientifique a-t-elle encore un sens lorsque les outils de découverte eux-mêmes (les IA) reposent sur des biais statistiques ou de données ?

# Accepter des résultats issus d’un modèle non explicable, est-ce encore faire de la science ?

**En laissant les IA explorer le savoir à notre place, prenons-nous le risque de devenir ignorants de nos propres découvertes ?**