Compte rendu - Présentation et débat sur l'IA Scientifique (AI Scientist)

Problématique : Jusqu'où une intelligence artificielle peut-elle réellement participer à la production de connaissances scientifiques ?

La présentation a exploré cette question à travers trois grandes approches de l'intelligence artificielle : l'approche symbolique, l'approche connexionniste et les tentatives actuelles d'intégration dites neuro-symboliques.
L'objectif central du domaine est de concevoir une IA capable non seulement d'apprendre et de produire des résultats, mais aussi de formuler des hypothèses, de raisonner, de s'auto-corriger et d'expliquer ses propres conclusions. Cette exigence d'interprétabilité est au cœur de l'ambition d'un « scientifique artificiel ».

I. L'approche symbolique : raisonner pour comprendre

L'approche symbolique cherche à reproduire le raisonnement scientifique explicite : formuler des hypothèses, les tester, les corriger. Elle repose sur la métaconnaissance, c'est-à-dire la capacité du système à comprendre sa propre activité de recherche.

Jacques Pitrat en a été l'un des pionniers avec le projet CAIA (Chercheur Artificiel en Intelligence Artificielle). Ce système se compose d'une société d'agents :
- Alice, l'exécutant, génère des hypothèses ;
- Monitor évalue leurs performances ;
- Manager planifie les tâches ;
- Advisor conseille à partir des erreurs passées ;
- Zeus coordonne l'ensemble et veille au bon fonctionnement global.

CAIA incarne l'idée d'une IA capable de raisonner sur sa propre recherche. Ses réussites ont montré qu'un tel système pouvait découvrir des contraintes ou des stratégies inédites. Néanmoins, son déploiement reste limité par la formalisation initiale des problèmes, souvent dépendante de l'intervention humaine, et par une difficulté à reformuler les tâches de manière autonome.

Pitrat lui-même soulignait que la création d'un véritable « chercheur artificiel » nécessiterait encore plusieurs décennies, voire un siècle.

II. L'approche connexionniste : apprendre à découvrir

L'approche connexionniste s'appuie sur l'apprentissage automatique et sur la puissance des données. Elle a connu un tournant majeur avec AlphaFold, développé par Google DeepMind, qui a permis de prédire avec une précision remarquable la structure 3D des protéines à partir de leur séquence d'acides aminés. Cette réussite, saluée par le Prix Nobel de chimie 2024, illustre la capacité des réseaux de neurones à produire des découvertes empiriques à grande échelle.

Ces modèles offrent des prédictions rapides et fiables, accélérant la recherche en biologie structurale et révélant des régularités invisibles aux méthodes classiques. Cependant, ils fonctionnent comme des boîtes noires : leurs mécanismes internes restent opaques, leur raisonnement difficile à interpréter. Ils reposent sur la corrélation plus que sur la causalité et manquent de réflexivité.

Pour pallier cette opacité, de nouveaux modèles tels que les Kolmogorov-Arnold Networks (KAN) ont été développés. En remplaçant les activations fixes des réseaux classiques par des fonctions apprenables (B-splines), ces modèles visent une meilleure interprétabilité locale. Les KAN permettent d'identifier les variables importantes et de visualiser certaines transformations internes. Néanmoins, ils demeurent privés d'une compréhension globale ou d'une réelle métacognition.

III. L'approche neuro-symbolique : raisonner et apprendre

Afin de combiner les atouts des deux paradigmes précédents, les approches neuro-symboliques tentent d'articuler l'apprentissage empirique des réseaux neuronaux avec la rigueur du raisonnement symbolique.
L'objectif est de concevoir une IA capable d'apprendre à partir des données tout en manipulant des concepts et des règles explicites. Autrement dit, une IA qui comprend ses découvertes.

Le système AI Feynman illustre bien cette ambition. Il associe apprentissage neuronal et raisonnement symbolique pour redécouvrir, à partir de simples jeux de données, des équations issues des Feynman Lectures on Physics. Le réseau repère d'abord des invariances, puis un moteur symbolique explore l'espace des formules possibles jusqu'à retrouver une équation explicite.
AI Feynman a ainsi pu redécouvrir 90 formules exactes sur 100, démontrant la complémentarité des deux approches. Ses limites résident toutefois dans la sensibilité au bruit des données et dans une intégration encore superficielle entre les deux paradigmes, le symbolique étant souvent ajouté au-dessus du réseau plutôt que fusionné en profondeur. Il est pour l'instant trop tôt pour dire que le neuro-symbolisme sera capable de résoudre à la fois les problèmes du connexionnisme et du symbolisme, sans ajouter de nouveaux problèmes limitants.

Conclusion et synthèse du débat

La présentation a mis en lumière la trajectoire de l'IA scientifique, oscillant entre le raisonnement déductif et explicatif de l'approche symbolique (CAIA) et la découverte empirique ultra-efficace de l'approche connexionniste (AlphaFold)

Les échanges qui ont suivi ont ouvert une réflexion sur les implications éthiques et épistémologiques de ces nouvelles formes de recherche automatisée.
Cette discussion a soulevé la question cruciale de la confiance dans les résultats produits par une IA, même lorsque cette dernière est dotée d'un système de justification pour expliquer son raisonnement et l'obtention de ses résultats.

Les positions exprimées au cours du débat ont révélé une diversité d'attitudes face à l'IA Scientifique : certains participants se montrent prêts à l'utiliser comme un simple outil d'assistance, d'autres envisagent de lui confier davantage de responsabilités à mesure que la technologie progressera, tandis que plusieurs restent profondément inquiets quant à son intégration dans le processus scientifique. Cette pluralité d'opinions illustre la manière dont, même au sein de la communauté de recherche en IA, les visions divergent quant à la place et au rôle futur d'un « AI Scientist ».

Des pistes ont tout de même été évoquées pour encadrer ces usages : valoriser la relecture humaine, renforcer la traçabilité des découvertes et encourager le développement d'outils explicatifs.