



Centre Jean Piaget | Séminaire interdisciplinaire | 2025

Photo: Ionna Berthoud-Papandropoulou

Intelligences animale, humaine et artificielle : similarités et spécificités

Mercredi 26 mars

Le principe d'inférence comportementale et la question de la conscience artificielle

Par Stefano Palminteri, Directeur de recherche INSERM, Ecole Normale Supérieure, Paris

Les modèles de langage de grande taille (LLM) sont rapidement devenus un sujet central en IA et en sciences cognitives, en raison de leurs performances sans précédent dans une vaste gamme de tâches. En effet, certains y voient même des "étincelles d'intelligence artificielle générale" dans leur faculté apparemment illimitée de conversation et de raisonnement. Leurs facultés émergentes sophistiquées, qui n'avaient pas été initialement anticipées par leurs concepteurs, ont déclenché un débat urgent sur la question de savoir si, et dans quelles circonstances, nous devrions attribuer la conscience à des entités artificielles en général et aux LLM en particulier. Le consensus actuel, fondé sur le fonctionnalisme computationnel, propose que la conscience soit attribuée selon un principe d'équivalence computationnelle. L'objectif de cet article d'opinion est de critiquer cette approche actuelle et de plaider en faveur d'un principe alternatif : le "principe d'inférence comportementale". Nous pensons que ce principe d'inférence comportementale fournira un critère épistémologiquement impartial et opérationnalisable pour évaluer la conscience des machines.



Stefano Palminteri est directeur de recherche à l'INSERM et dirige une équipe de recherche à l'École Normale Supérieure de Paris, où sont étudiées les bases computationnelles de l'apprentissage et de la prise de décision chez l'homme. Lauréat de nombreux prix et financements prestigieux, notamment de la Fondation Schlumberger, de la Fondation Alexander von Humboldt, ainsi que du prestigieux ERC Consolidator Grant, il est reconnu pour son apport significatif dans le domaine des sciences cognitives. Plus récemment, son équipe s'est aventurée dans l'étude de la cognition artificielle, en explorant le raisonnement des grands modèles de langage (LLMs) et les interactions homme-machine.

