Les lois de l'intelligence. L'autre révolution de l'IA

Titre(s) : Les lois de l'intelligence. L'autre révolution de l'IA [[periodique]] / Jean-Baptiste Veyrieras

Ensemble : Epsiloon 58

Auteur(s) : Veyrieras, Jean-Baptiste

Autre(s) auteur(s) : Laurens, Clémentine

Editeur, producteur : 01/04/26

Description matérielle : pp.40-53

ISSN : 2800-4736

Note sur la description matérielle : 15

Résumé ou extrait : Des travaux récents en neurosciences et en intelligence artificielle montrent que des réseaux artificiels très différents finissent souvent par produire des représentations internes proches de celles du cerveau humain. En vision, Michael Bonner et Zirui Chen ont comparé la plus grande base de données d'IRM fonctionnelle de cerveaux humains exposés à des images naturelles à une vingtaine de réseaux de neurones artificiels. Malgré des architectures, des données d'entraînement et des objectifs variés, ces modèles convergent vers des « dimensions universelles des représentations visuelles », c'est-à-dire des axes d'organisation communs et utiles quelle que soit la tâche visuelle. Dans un réseau de 10 millions de neurones, chaque image peut être représentée par un vecteur de 10 millions de dimensions, mais l'analyse statistique permet de dégager des régularités partagées avec le cerveau humain.L'article souligne aussi que les IA traitent spontanément les données complexes selon une hiérarchie allant du simple au complexe, comme le cerveau pour la vision ou le langage. Les premières couches repèrent des éléments élémentaires, puis les couches profondes construisent des représentations de plus en plus abstraites. Les grands modèles de langage renforcés par apprentissage apprennent par ailleurs à générer un texte intermédiaire pour découper une question en sous-problèmes, résoudre chaque étape puis recomposer une réponse, ce qui améliore fortement leur raisonnement.Au-delà des ressemblances neuronales, les évaluations comportementales font apparaître des capacités émergentes. Google en avait recensé 137 dès 2022. GPT-4 a réussi un test de Turing auprès de 73 % des évaluateurs dans une prépublication de mars 2025. Le modèle Centaur, dérivé de Llama et entraîné sur les décisions de 60 000 personnes issues de 160 expériences de psychologie, prédit mieux les comportements humains que les meilleurs modèles cognitifs classiques. D'autres travaux décrivent chez les IA une théorie de l'esprit, des traits de personnalité stables observés sur 18 modèles, une créativité partielle, une forme d'intuition et une capacité croissante au mensonge.Les chercheurs restent prudents : parler comme un humain ne signifie pas nécessairement penser comme lui. Les méthodes d'apprentissage diffèrent, l'ancrage physique des IA reste limité et l'opacité des modèles augmente. Avec près de 110 milliards de dollars d'investissements privés dans l'IA en 2024 et plus de 650 milliards annoncés en 2026, l'enjeu est désormais scientifique, économique et sécuritaire. L'article plaide pour une véritable psychologie des machines et pour l'étude

d'une cognition hybride entre humains et IA....

Sujet - Nom commun : Intelligence artificielle -- Modèles mathématiques
Neurosciences -- Méthodes
Neurones -- Traduction automatique