

On fait le point sur...

L'IA, un outil en question

IA générative, quelles limites ?



Problématique

Avec la révolution numérique liée à l'émergence des Intelligences artificielles, de nombreuses questions et inquiétudes émergent. Quelles sont les limites de l'IA à prendre en considération ?

Mots-clés

Mot 1 : Biais, hallucinations, Intelligence biologique

Enjeux majeurs de la question

La question des biais algorithmiques et des hallucinations

Les modèles génératifs fonctionnent sur un modèle probabiliste : ils génèrent une réponse statistique complexe (souvent la plus probable) à partir des données sur lesquelles ils ont été entraînés. Cela provoque deux dangers :

- **Les hallucinations** : C'est un phénomène où un système d'intelligence artificielle, en particulier un grand modèle de langage (LLM), génère des **informations incorrectes, fictives ou trompeuses** tout en les présentant comme des faits avérés. Les hallucinations d'IA se manifestent de plusieurs manières : **Génération de fausses informations** (événements ou des faits qui n'ont aucun fondement dans la réalité) ; **Réponses hors sujet** ; **Mélange d'informations** (combiner des éléments véridiques avec des éléments fictifs, rendant difficile la distinction entre le vrai et le faux)

Les causes principales des hallucinations d'IA incluent : des données d'entraînement insuffisantes ou biaisées, la complexité élevée des modèles, des ambiguïtés sémantiques dans les requêtes des utilisateurs, le surajustement (overfitting) des modèles.

- **Les biais algorithmiques** : Ce phénomène est l'un des plus graves problèmes à ce jour. Il s'agit d'une « déviation par rapport à un résultat censé être neutre, loyal ou encore équitable » et on en distingue deux grandes familles : Les **biais cognitifs**, qui irriguent tous les autres (biais de confirmation, tendance à détecter de fausses corrélations entre événements) mais aussi **les biais de données ou biais de représentativité** (inadéquation entre données utilisées pour l'entraînement de l'algorithme et données cibles sur lequel il sera déployé). Les **préjugés et stéréotypes** sont donc souvent reproduits et amplifiés. L'IA reflète donc les données sur lesquelles elles sont entraînées : elles sont conçues dans un milieu très masculin, blanc, par des personnes de classes sociales élevées et provenant des pays occidentaux.

La question des limites techniques de l'IA

Pour le chercheur Frédéric Alexandre (*dir. de recherche chez Inria et responsable de l'équipe-projet Mnemosyne*), la force et la spécificité de **l'intelligence biologique** que n'a pas l'intelligence artificielle, c'est de nous permettre de **répondre à des situations complexes en usant de stratégies** mêlant créativité, adaptabilité et émotions. Pour ce chercheur, l'IA essaie de reproduire au mieux certaines de nos fonctions cognitives. Elle le fait avec beaucoup de performances pour certaines (reconnaître) et beaucoup moins pour d'autres (aspects émotionnels). Elle reste donc

Ressources du pôle de compétences académique Histoire-Géographie et numérique

On fait le point sur...

spécialisée à certaines fonctions sans parvenir pour l'instant à faire émerger une faculté d'intelligence plus globale comme l'**Intelligence biologique**.

L'IA générative produit des modèles de langages qui s'appuient sur des **connaissances statiques**. En avoir conscience est indispensable pour prendre en compte le fait que les réponses générées ne seront que le résultat lié aux informations de base dont dispose l'IA. Elles **ne peuvent donc intégrer des connaissances réactualisées** au regard d'événements récents, ni même des corrections à des données intégrées au préalable avec des erreurs et qui n'auraient pas été corrigées. Cela pose d'autant plus de problèmes lorsque ces IA sont consultées dans des domaines comme la santé, les questions politiques ou de sujets sociétaux qui évoluent parfois rapidement.

Pour aller plus loin

Sitographie

Avis des autorités :

- [CNIL](#)
- [EU AI act](#)
- EMI : Class code IA [jeu pour comprendre le fonctionnement de l'IA](#)

Articles et rapports:

- [les biais sexistes, en voie d'amélioration \(TV5 monde\)](#)
- [Le livre Blanc de l'IA générative \(Université de Strasbourg\)](#)

Articles et vidéos en ligne :

- <https://www.youtube.com/watch?si=2GfztVnKRwkDApdX&v=xXHWTC4mJBM&feature=youtu.be>
- <https://www.lemonde.fr/blog/binaire/2018/05/14/intelligence-artificielle-et-pensee-humaine/>

Quand l'aborder avec nos élèves

Au collège

En EMC : Il pourrait être intéressant d'aborder ces limites dans le cadre du programme de 5^e « **agir pour l'égalité homme femme** » via différentes activités montrant la génération d'images « biaisées » lors d'un travail sur les stéréotypes.

Au lycée

En HGGSP, en classe de Première, lors du Thème 4 : « **S'informer : un regard critique sur les sources et les modes de communication** » ou lors objet de travail conclusif : « **l'information à l'heure d'internet** »,

En EMC, 2nd le sujet peut être abordé lors de la **thématique** « **Liberté et responsabilité : l'exemple de l'information (vecteurs, nécessité et enjeux)** »