

# Usages en recherche de logiciels (dits d'« intelligence artificielle ») de production de texte, code ou images : questions d'intégrité scientifique – partie 1

64<sup>e</sup> congrès de la Société des Anglicistes de l'Enseignement supérieur

Panel 1 – Transitions induites par l'Intelligence Artificielle dans la recherche en anglistique

6 juin 2025

Catherine Tessier  
[referent-integrite-ethique@onera.fr](mailto:referent-integrite-ethique@onera.fr)

# « Intelligence artificielle » ?

## L'intelligence artificielle

- est une **discipline scientifique**
- qui recouvre un ensemble de techniques différentes, voir [ici](#) par exemple



## Terme utilisé par abus de langage pour parler :

- de techniques d'apprentissage machine, quelles qu'elles soient
- de ce qui nous concerne ici :

des **agents conversationnels**, donc des **logiciels**,  
fondés sur **plusieurs techniques d'apprentissage machine**  
qui produisent du texte, du code informatique ou des images,  
en réponse à des **requêtes** d'un utilisateur

# Comment fonctionnent de tels logiciels ?

---

En *très bref* et *très schématique*, pour le texte (ou le code informatique) :

- L'utilisateur, par ses requêtes\* (*prompts*), **programme** l'outil
  - L'outil produit de « nouvelles » données, par exemple la séquence de mots **la plus probable** après le *prompt*, à partir de caractéristiques communes « apprises » sur un **corpus de données de très grande taille** [CNPEN23]
- corrélations statistiques + filtrage de résultats + annotations humaines

Les **grands modèles de langue** (LLM - *Large Language Model*) sont entraînés sur un corpus de textes

- généraliste (internet)
- ou spécifique (exemple : corpus bibliographique)

La *Liste relative au vocabulaire de l'intelligence artificielle (termes, expressions et définitions adoptés)*, JORF n° 0212 du 6 septembre 2024, préfère le terme d' « instruction générative » - [lien](#)

[CNPEN23] Comité national pilote d'éthique du numérique. *Avis 7 – Systèmes d'intelligence artificielle générative : enjeux d'éthique*, juin 2023. [lien](#)

# Par nature (1/2)

---

- Requêtes de l'utilisateur

S'il s'agit d'un outil ouvert (en ligne, ex : ChatGPT)

→ **divulcation des informations** contenues dans les requêtes

- sécurité
- vie privée
- propriété intellectuelle

- Résultats fournis

→ issus de corpus contenant des éléments **originaux**

→ respect de la propriété intellectuelle des tiers ?

# Par nature (2/2)

---

C'est un logiciel

→ ne « comprend » pas

→ illusion de « dialogue »

Fonctionnement statistique

→ possibilité de résultats **erronés** (manque de robustesse)

→ pas de notion de « vérité »

→ **reproductibilité ?**