

# **DERRIÈRE LA RÉCENTE VAGUE D'ASSISTANTS IA POUR LA RECHERCHE**

## **UNE COURSE AUX DONNÉES ET LLM SCIENTIFIQUES**

Patrice Lopez, science-miner

# LLM ET ASSISTANTS IA



# LARGE LANGUAGE MODEL (LLM)

Grand modèle de langage (10-500 milliards de paramètres) entraîné sur un grand volume de données (>500 milliards de tokens), exemple: GPT-3 (2020)

- modèle génératif : réalise une « complétion automatique »... un token après l'autre
- nécessite une expertise pour l'utilisation

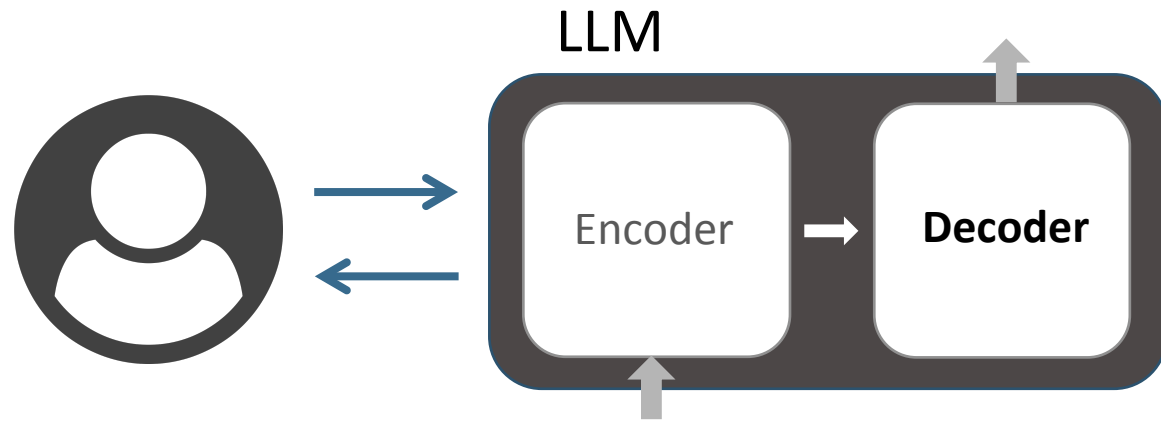
# LLM ENTRAÎNÉ EN CHATBOT

LLM entraîné pour le dialogue (dizaines de milliers de prompts/réponses évalués manuellement), exemple GPT-3 -> ChatGPT (2022)

Capacité de générer du contenu simplement avec des instructions ou un peu de données :

- usage direct, plus besoin de compétence technique !
- IA devient directement accessible au grand public

# LLM ENTRAÎNÉ EN CHATBOT : APPLICATION DIFFICILE POUR LA SCIENCE ET LA RECHERCHE



## Utilisation directe LLM/Chatbot:

- LLMs ne peuvent générer que des connaissances générales/consensuelles
- Non compatible aux tâches scientifiques: « hallucinations »
- **Galactica** (LLM scientifique de Meta) débranché après 3 jours

couverture



correction

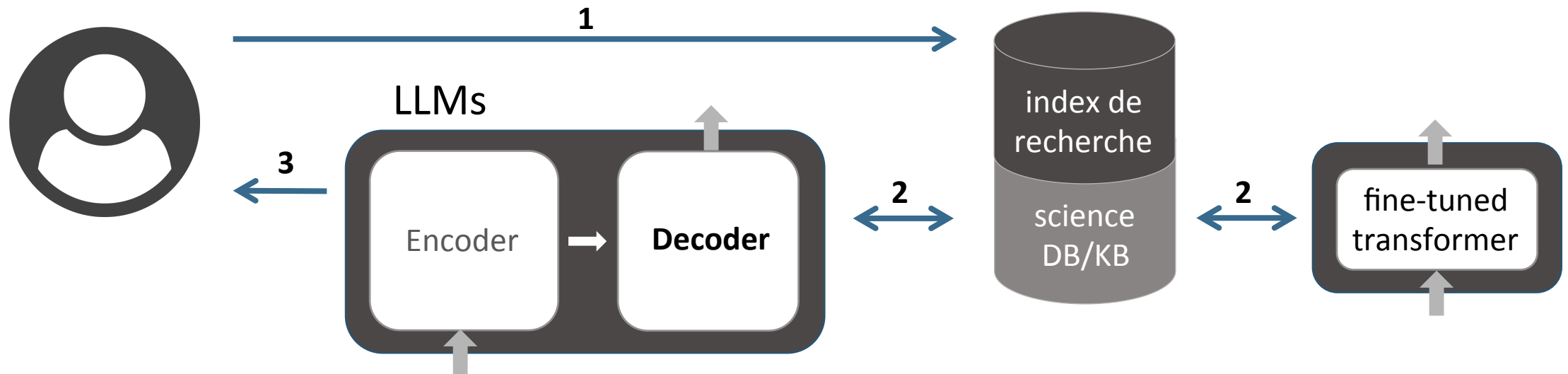


fraîcheur



# LLM ET ASSISTANTS IA : RETRIEVAL AUGMENTED GENERATION (RAG)

1. Appel à un moteur de recherche/base de données pour obtenir des résultats fiables
2. Ré-arrangement/enrichissement des meilleurs résultats
3. Le LLM est utilisé pour générer une réponse en langage naturelle à partir des meilleurs résultats



## Approche standard de tous les assistants IA actuellement

- Permet d'injecter des données correctes et à jour à l'application
- Très simple à prototyper à partir d'un moteur de recherche existant
- Mais **différenciateur et valeur** se situent alors dans la BD/moteur de recherche !

# ASSISTANTS IA POUR LA RECHERCHE



# 2023: UNE VAGUE D'ASSISTANTS IA POUR LA RECHERCHE

Répond à des questions sur la base d'un index de publications scientifiques :

- Consensus
- **scite.ai assistant** ←
- Zeta alpha
- Dimensions AI assistant (beta)
- Scopus.ai (beta)
- Ask R Discover (beta)
- Core-GPT (announced, CORE)

Répond à des questions concernant une ou plusieurs publications fournies par l'utilisateur :

- Elicit
- ResearchGPT/sSciSpace ([typeset.io](https://typeset.io))
- Curie (Springer Nature) academic writing
- [getconverse.com](https://getconverse.com)
- ChatPDF
- HeyScience (beta) review prior to submission
- JSTOR AI pilot

**scite\_ demo !**

# LIMITES ACTUELLES DES APPROCHES "RAG"

- La qualité informationnelle d'une réponse dépend du **moteur de recherche**
- Le LLM est limité à une génération de surface qui "recolle" des morceaux de textes (snippets) : le "contenu" provient du moteur de recherche et l'utilisation est plus de l'ordre de la **paraphrase**
- **Beaucoup du potentiel des LLM est neutralisé** : lier des informations de façons nouvelles, injecter des information supplémentaires, suggestions, multimodalité, désambiguisation par le dialogue, etc.
- On est loin du potentiel des LLM à cause essentiellement des **risques de génération non factuelle pour un usage de recherche** (« hallucinations »)



# LLM SCIENTIFIQUES



# MANQUE DE DONNÉES SCIENTIFIQUES DANS LES LLM MAINSTREAM ET SPÉCIALISÉS

- Exemple Galactica (Meta) : 8-9 M d'articles de plein textes pour le pré-entraînement (40 M de résumés)
  - 6-7% sur un corpus d'au moins 120 M de publications scientifiques
  - dont potentiellement 40 M en accès ouvert
- Pas de contenu scientifique particulier pour GPT-3, ajouts a priori limités pour GPT-3.5 & ChatGPT
- Images, figures, grandes bases de données scientifiques (GenBank, UniPROT, PubChem, RefSeq, etc.), alignements texte/nomenclature
- Le défi n'est pas seulement l'entraînement des modèles durant des semaines/mois, mais aussi **collecter et structurer des données d'entraînement de qualité**

# INITIATIVES LLM SCIENTIFIQUES

Réalisées:

- **Galactica** (Meta, 2022) : généraliste science (106 Mrds tokens)
- **Minerva** (Google, 2022) : mathématique (dérivé de PaLM avec arXiv et pages web du domaine)
- **Llemma** (EleutherAI, 2023): mathématique (dérivé de Llama, 55 Mrd tokens, similaire à Minerva)

En cours :

- **OLMo**: Allen Institute on AI (US), LUMI (Finlande)
  - **SciLOnS** : collaboration avec des chercheurs du DFKI (OpenGPT-X), de CommonCrawl, ... 35 millions de publications en access ouvert préparées via Grobid, Pub2TEI, LaTeXML (ressources de calcul NSF ACCESS avec UT Austin, 50 TB de données, 300 Mrd tokens)
- L'accès et l'ingestion de ces données est en soi un défi !

# MOTS DE LA FIN : IA GÉNÉRATIVE ET INFORMATION SCIENTIFIQUE



# IA GÉNÉRATIVE : POSSIBLE CHANGEMENT DE PARADIGME EN INFORMATION SCIENTIFIQUE

## IA générative et copyrights?

- Jurisprudence USPTO : pas de copyrights pour les parties d'une oeuvre générées par ML
- GitHub Copilot: code logiciel mixte avec/sans copyrights ?
- Publications scientifiques assistées par ML et accès fermé ?



## Évolution de l'aspect narratif des publications scientifiques :

- La dimension textuelle et narrative d'une publication scientifique reposera de plus en plus sur de l'IA générative et suivrait donc une dévalorisation
- La présentation des résultats de recherche devient une "commodité" automatisée et contextualisée
- Opportunités vers de nouvelles formes de publications moins textuelles : data-ification, formalisation argumentaire (hypothèse, réfutation, protocoles, etc.), logiciels



# MERCI!

---



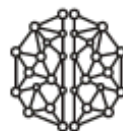
PATRICE.LOPEZ@SCIENCE-MINER.COM




[HTTPS://SCIENCE-MINER.COM](https://science-miner.com)





Démo scite.ai




### Ask simple questions and get reliable answers from the full-texts of millions of research articles


"How does the structure of a protein affect its function?" 


"What are the cognitive and neural mechanisms underlying decision-making and risk-taking behavior?" 

"How many rats live in NYC?" 


### Get unblocked on whatever you're writing, whether it's a simple blog post, essay, or a grant proposal


"What are some major unanswered questions about empathy?" 


"What are some controversies in the field of nephrology?" 

"What are some common arguments against raising the minimum wage?" 

### Effectively use information from research articles to support your research tasks

"Find competing evidence for the paragraph below" 

"Summarize the following in two sentences with no references:" 

"Find a source for the following sentence:" 

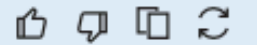


What do you need help with?





are dogs more intelligent than cats? 



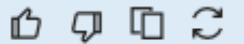
To determine whether dogs are more intelligent than cats, it is essential to consider various aspects of intelligence and behavior in both species. While some studies suggest that dogs may exhibit more human-oriented behaviors and emotional closeness compared to cats ([Fraga et al., 2020](#); [González-Ramírez & Hernández, 2021](#)), other research indicates that cats and dogs are capable of forming amicable relationships and living comfortably in each other's presence ([Prior & Mills, 2020](#)). Additionally, a study found that children have a higher level of empathy for dogs and cats compared to other animals, suggesting that both species have the capacity to engage with humans ([Shimatani & Koda, 2021](#)).

Furthermore, it is important to note that the intelligence and behavior of cats and dogs can be influenced by their evolutionary history and domestication. For example, dogs tend to be open and expressive, while cats are elusive and cautious ([Jia et al., 2022](#)). This may reflect differences in their social and cognitive development, which could impact their perceived intelligence.

Moreover, it is crucial to consider the cognitive abilities of both cats and dogs. While some studies have suggested differences in their responses to human ostensive cues and problem-solving tasks ([Pongrácz & Onofer, 2020](#)), other research has highlighted similarities in their cognitive processing, such as their ability to learn and categorize stimuli ([French et al., 2004](#); [Mareschal et al., 2002](#); [Mareschal et al., 2002](#)). These findings indicate that both species possess cognitive capabilities, albeit with variations in their behavioral responses.

In conclusion, the question of whether dogs are more intelligent than cats is complex and multifaceted. While some studies suggest differences in their behavior and social interactions, it is important to consider the diverse cognitive and emotional capacities of both species. Ultimately, intelligence and

are dogs more intelligent than cats? 



To determine whether dogs are more intelligent than cats, it is essential to consider various aspects of intelligence and behavior in both species. While some studies suggest that dogs may exhibit more human-oriented behaviors and emotional closeness compared to cats (Fraga et al., 2020; González-Ramírez & Hernández, 2021), other research indicates that cats and dogs are capable of forming amicable relationships and living comfortably in each other's presence (Prior & Mills, 2020). Additionally, studies have shown that both species have unique cognitive abilities for dogs and cats compared to other animals, suggesting that both species have their own forms of intelligence.

Furthermore, it is important to note that the process of domestication has influenced the social and cognitive development of both species. For example, dogs tend to exhibit more social behaviors and are often more dependent on human interaction, while cats are generally more independent and self-sufficient.

Moreover, it is crucial to consider the cognitive abilities of both species in relation to human ostensive cues and problem-solving skills. Dogs are known for their ability to learn and categorize human gestures and sounds, while cats are often more adept at recognizing and responding to subtle cues. Both species possess cognitive capabilities, albeit with variations in their behavioral responses.

In conclusion, the question of whether dogs are more intelligent than cats is complex and multifaceted. While some studies suggest differences in their behavior and social interactions, it is important to consider the diverse cognitive and emotional capacities of both species. Ultimately, intelligence and

"...However, emotional closeness is greater with dogs than with cats...."

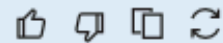
### Pet-Human Relationships: Dogs versus Cats

[González-Ramírez<sup>1</sup>](#), [Hernández<sup>2</sup>](#) 2021 *Animals*

 11 |  1 |  7 |  0

[View full text](#) [Add to dashboard](#) [Cite](#)

are dogs more intelligent than cats? 



To determine whether dogs are more intelligent than cats, it is essential to consider various aspects of intelligence and behavior in both species. While some studies suggest that dogs may exhibit more human-oriented behaviors and emotional closeness compared to cats (Fraga et al., 2020; González-Ramírez & Hernández, 2021), other research indicates that cats and dogs are capable of forming amicable relationships and living comfortably in each other's presence (Prior & Mills, 2020). Additionally, a study found that children have a higher level of empathy for dogs and cats compared to other animals, suggesting that both species have the capacity to engage with humans (Shimatani & Koda, 2021).

Furthermore, it is important to note that the intelligence and behavior of cats and dogs can be influenced by their evolutionary history and domestication. For example, dogs tend to be open and expressive, while cats are elusive and cautious (Jia et al., 2022). This may reflect differences in their social and cognitive development, which could impact their perceived intelligence.

Moreover, it is crucial to consider the cognitive abilities of both cats and dogs. While some studies have suggested differences in their responses to human ostensive cues and problem-

## References

### Reference #1

"...However, both wolves (with similar anatomy) and cats (with a flexible neck Graf et al 1995 ; Zhang et al 2014 ) were shown to perform less human-oriented behaviours in a problem-solving context than dogs (Brubaker et al 2017 ; Miklósi et al 2005)...."

▼ See full context

Section: Discussion

### Who turns to the human? Companion pigs' and dogs' behaviour in the unsolvable task paradigm

Paula Pérez Fraga<sup>1</sup>, László Gerencsér<sup>2</sup>, Melinda Lovas<sup>3</sup> et al.  
2020 *Anim Cogn*

 11 |  2 |  5 |  0

[View full text](#) [Add to dashboard](#) [Cite](#)

In conclusion, the question of whether dogs are more intelligent than cats is complex and multifaceted. While some studies suggest differences in their behavior and social interactions, it is important to consider the diverse cognitive and emotional capacities of both species. Ultimately, intelligence and behavior in cats and dogs are influenced by a range of factors, including their evolutionary history, domestication, and individual differences.

^ View searches run to ground this response

[problem solving skills dogs cats](#)

[learning behavior dogs cats](#)

[intelligence comparison dogs cats](#)

#### ^ View publications consulted

[1] Zentall "Comparative Cognition Research Demonstrates the Similarity between Humans and Other Animals" *Animals* (2023) doi:10.3390/ani13071165

[2] Pearson "DOGS, HISTORY, AND AGENCY" *History and theory* (2013) doi:10.1111/hith.10683

[3] Wijker et al. "Effects of Dog Assisted Therapy for Adults with Autism Spectrum Disorder: An Exploratory Randomized Controlled Trial" *Journal of autism and developmental disorders* (2019) doi:10.1007/s10803-019-03971-9

[4] Spelke and Lee "Core systems of geometry in animal minds" *Philosophical transactions of the royal society b biological sciences* (2012)

#### ^ View publications used

[1] Fox "Beyond AI: Multi-Intelligence (MI) Combining Natural and Artificial Intelligences in Hybrid Beings and Systems" *Technologies* (2017) doi:10.3390/technologies5030038

[2] Uccheddu et al. "Comparing Pears to Apples: Unlike Dogs, Cats Need Habituation before Lab Tests" *Animals* (2022) doi:10.3390/ani12213046

[3] Duncan et al. "Sub-clinical infection and asymptomatic carriage of *Cryptococcus gattii* in dogs and cats during an outbreak of cryptococcosis" *Medical mycology* (2005) doi:10.1080/13693780500036019

[4] Stern et al. "Evaluation of a human urine barbiturate test to screen for