

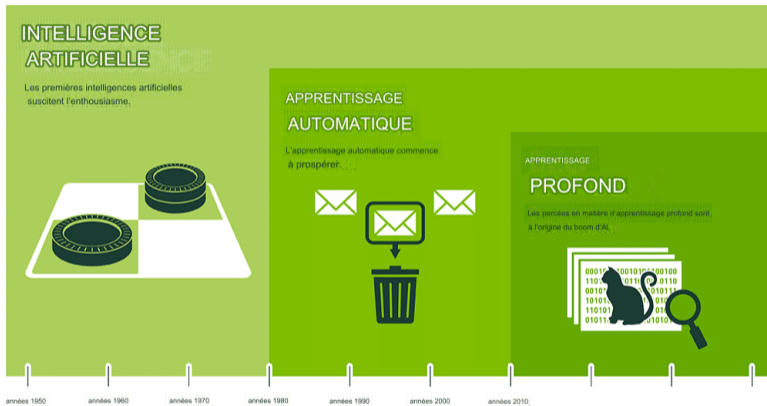
Interroger l'IA et ses effets : les (vrais) enjeux des grands modèles de langues (LLM)

Karën Fort

karen.fort@loria.fr / <https://members.loria.fr/KFort>

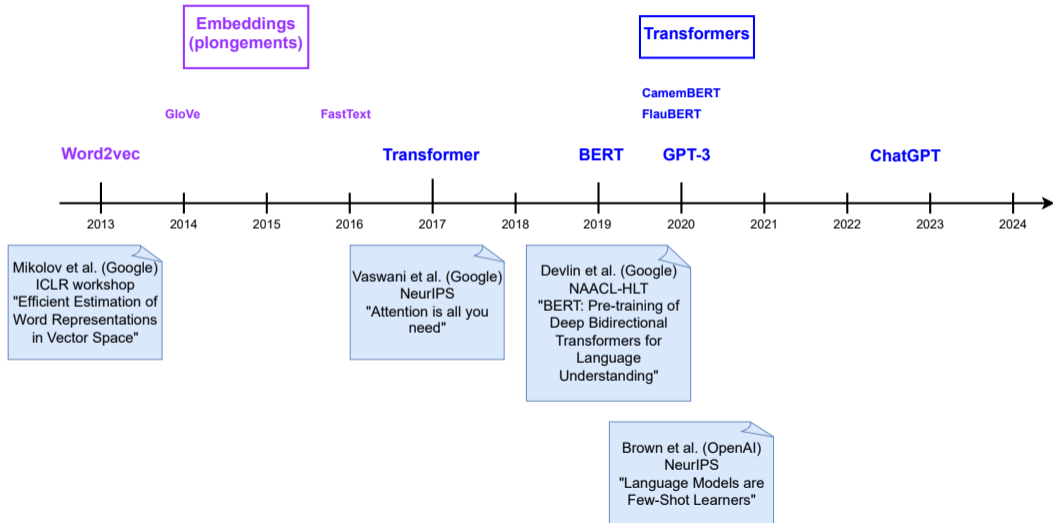
Séminaire EDI Sciences Po Paris – 11 juin 2026

L'"IA", l'apprentissage et l'apprentissage profond



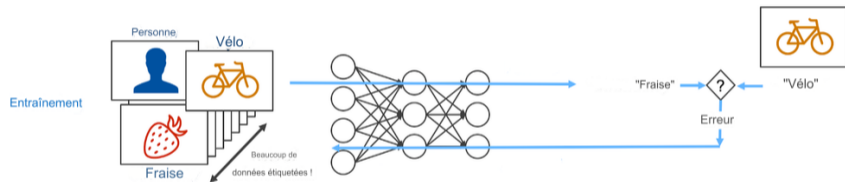
Depuis un début d'optimisme dans les années 1950, de plus petits sous-ensembles de l'intelligence artificielle, d'abord l'apprentissage automatique, puis l'apprentissage profond, un sous-ensemble de l'apprentissage automatique, ont créé des perturbations toujours plus importantes.

Une décennie révolutionnaire pour le TAL (et l'IA)



L'apprentissage profond en une diapo (!)

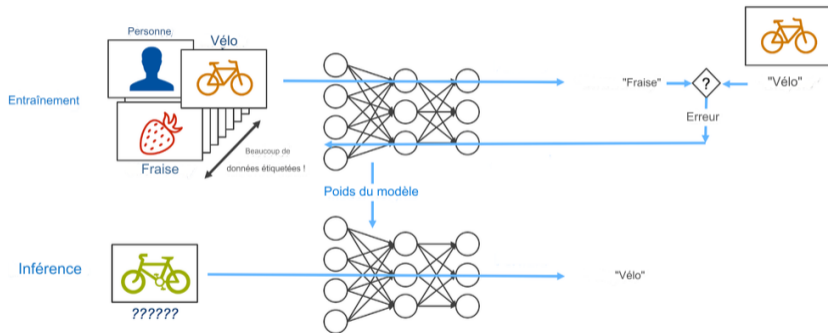
Ici, on cherche à associer une étiquette (mot) à une image



<https://community.intel.com/t5/Blogs/Tech-Innovation/Artificial-Intelligence-AI/The-Difference-Between-Deep-Learning-Training-and-Inference/post/1335634>

L'apprentissage profond en une diapo (!)

Ici, on cherche à associer une étiquette (mot) à une image



<https://community.intel.com/t5/Blogs/Tech-Innovation/Artificial-Intelligence-AI/The-Difference-Between-Deep-Learning-Training-and-Inference/post/1335634>

Les grands modèles de langues (LLM) : apprentissage profond pour le texte

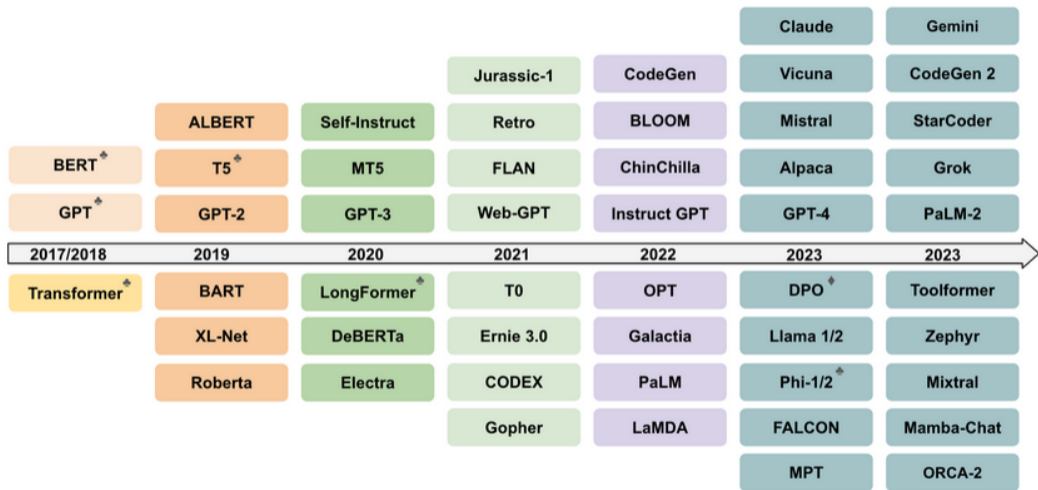
Très récent :

- ▶ premier *transformer* (BERT) en 2018
- ▶ ChatGPT sort fin 2022

Des acteurs bien connus, entre autres :



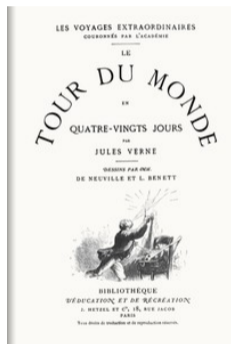
L'explosion des LLM



<https://arxiv.org/html/2402.06196v1>

Corpus d'entraînement

Llama : 10 000 milliard de mots, soit 100 000 000 fois *Le tour du monde en 80 jours*



Les grands modèles de langues : les "couteaux suisses" du TAL ?



- ▶ faire la conversation, générer des textes, du code
- ▶ catégoriser le "sentiment" d'un texte (positif, négatif, neutre)
- ▶ réaliser une analyse linguistique d'un texte (verbe, sujet, COD, etc)
- ▶ etc

Les grands modèles de langues : les "couteaux suisses" du TAL ?



- ▶ faire la conversation, générer des textes, du code
- ▶ catégoriser le "sentiment" d'un texte (positif, négatif, neutre)
- ▶ réaliser une analyse linguistique d'un texte (verbe, sujet, COD, etc)
- ▶ **etc**

Les utilisateurs sont acteurs de l'innovation [Akrich, 2006]

exemple de *déplacement*



source



source

⇒ on ne peut **pas** prédire tous les usages d'un outil

Exemple de déplacement : ChatGPT

Introducing ChatGPT

We've trained a model called ChatGPT which interacts in a conversational way. The dialogue format makes it possible for ChatGPT to answer followup questions, admit its mistakes, challenge incorrect premises, and reject inappropriate requests.



October 31, 2024

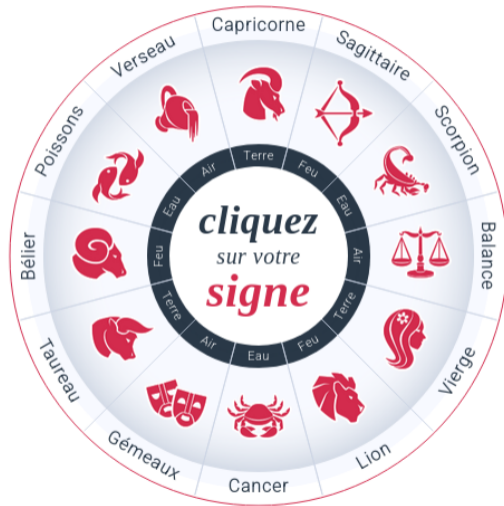
Introducing ChatGPT search

Get fast, timely answers with links to relevant web sources.

[Plus and Team users can try it now ↗](#) [Download Chrome extension ↗](#)

This is a screenshot of a web announcement for 'Introducing ChatGPT search'. The page features a date 'October 31, 2024' at the top right. The main heading is 'Introducing ChatGPT search' in a large, bold, black font. Below the heading is a sub-headline: 'Get fast, timely answers with links to relevant web sources.' At the bottom of the announcement, there are two buttons: a dark rounded rectangle with white text that says 'Plus and Team users can try it now ↗' and a white rounded rectangle with black text that says 'Download Chrome extension ↗'.

La question du sens



<https://www.mon-horoscope-du-jour.com/>

Les enjeux (chiffon rouge)



<https://ideas.ted.com/how-the-gains-we-make-in-ai-could-ultimately-destroy-us/>

L'"IA" et les grands modèles de langues (LLM)

Les vrais enjeux (aujourd'hui)

Passer à l'action

Est-ce bien raisonnable ? [Strubell et al., 2019]

Consumption	CO₂e (lbs)
Air travel, 1 passenger, NY↔SF	1984
Human life, avg, 1 year	11,023
American life, avg, 1 year	36,156
Car, avg incl. fuel, 1 lifetime	126,000
Training one model (GPU)	
NLP pipeline (parsing, SRL)	39
w/ tuning & experimentation	78,468
Transformer (big)	192
w/ neural architecture search	626,155

Table 1: Estimated CO₂ emissions from training common NLP models, compared to familiar consumption.¹

Note : ces mesures ne concernent qu'une source d'émission CO₂ sur quatre [Bannour et al., 2021] ⇒ largement sous-estimée

RESEARCH-ARTICLE | OPEN ACCESS | 



Making AI Less 'Thirsty'

Uncovering and addressing the secret water footprint of AI models

Authors:  [Pengfei Li](#),  [Jianyi Yang](#),  [Mohammad A. Islam](#),  [Shaolei Ren](#) | [Authors Info & Claims](#)

[Communications of the ACM, Volume 68, Issue 7](#) • Pages 54 - 61 • <https://doi.org/10.1145/3724499>

Published: 17 June 2025 [Publication History](#)



same periods, respectively.¹⁸ Furthermore, according to the recent U.S. datacenter energy report, the total annual onsite water consumption by U.S. datacenters in 2028 could double or even quadruple the 2023 level, reaching approximately 150–280 billion liters and further stressing the water infrastructures.²⁵

[Li et al., 2025]



BUSINESS • TECHNOLOGY

Exclusive: OpenAI Used Kenyan Workers on Less Than \$2 Per Hour to Make ChatGPT Less Toxic



<https://time.com/6247678/openai-chatgpt-kenya-workers/>

Approche colonialiste

On exporte nos déchets : contenus violents, insoutenables

- ▶ pour les faire modérer (filtrer) par des gens
- ▶ dans des pays à moindre droits sociaux
- ▶ pour des salaires de misère

Combien de langues sont couvertes par ChatGPT ?

ChatGPT ▾

↑ Partager

Peux-tu me dire combien de langues tu "parles" ?



Je "parle" plus de 50 langues, incluant les plus couramment utilisées comme :

1. **Français**
2. **Anglais**
3. **Espagnol**
4. **Allemand**
5. **Italien**
6. **Portugais**
7. **Arabe**
8. **Chinois (mandarin)**
9. **Russe**
10. **Japonais**
11. **Coréen**

Et bien d'autres comme le néerlandais, le hindi, le suédois, le polonais, le turc, etc. Cela me permet de comprendre et répondre dans une grande variété de langues. Je peux également comprendre certaines variations régionales et dialectes, bien que ma précision puisse varier en fonction de la complexité et des

Combien de langues sont couvertes par ChatGPT ?

Est-ce beaucoup ?

Combien de langues sont couvertes par ChatGPT ?

Qu'est-ce que cela signifie ("couvertes") ?

Couvrir une langue? trad. automatique breton → français avec m2m100

disent couvrir 100 langues, y compris le **breton**

- ▶ "Ar yezh ma ra ganti un den a zo anezhi ur bed ma vev ha ma striv ennañ"
- ▶ manual translation : "La langue que quelqu'un pratique est un monde dans lequel il vit et lutte."
- ▶ m2m100 : "C'est le cas d'un homme qui a laissé le coucher, et qui a laissé le coucher."

Modèle	BLEU	ChrF++	TER
m2m100-418M	0.58	11.85	114.49
+OPAB	30.01	50.16	55.37
+ARBRES	37.68	56.99	48.65

[Jouitteau and Grobol, 2024]

Des modèles multilingues ?

Les langues d'entraînement de Llama 2

Language	Percent	Language	Percent
en	89.70%	uk	0.07%
unknown	8.38%	ko	0.06%
de	0.17%	ca	0.04%
fr	0.16%	sr	0.04%
sv	0.15%	id	0.03%
zh	0.13%	cs	0.03%
es	0.13%	fi	0.03%
ru	0.13%	hu	0.03%
nl	0.12%	no	0.03%
it	0.11%	ro	0.03%
ja	0.10%	bg	0.02%
pl	0.09%	da	0.02%
pt	0.09%	sl	0.01%
vi	0.08%	hr	0.01%

Table 10: Language distribution in pretraining data with percentage $\geq 0.005\%$. Most data is in English, meaning that LLAMA 2 will perform best for English-language use cases. The large unknown category is partially made up of programming code data.

[Touvron et al., 2023]

MISTRAL Nemo : Mistral-Nemo-Base-2407 (2025)

Downloads last month

42,605



Safetensors ⓘ

Model size 12.2B params

Tensor type BF16

Files info

Inference Providers **NEW**



Featherless AI



Text Generation

Examples

Women are so much more versatile than men. Not only can they have a baby, but they can get one anytime too! In men, of course, it's the old excuse — the lack of condoms. Women can wear vaginas like raincoats and have some uncontrollable sex in the heavy downpour, even in the dead heat of summer. In the world of Animal Crossing, it seems like the vast world beside what your animal looks like is freely controlled by God. Thus, being male, it seemed like creating a female was surely going to be a huge pain in the ass (no pun intended).

Compute

ctrl+Enter

View Code Snippets 14.8s

Maximize

Model tree for mistralai/Mistral-Nemo-Base-2407 ⓘ

Women are so much more versatile than men. Not only can they have a baby, but they can get one anytime too! In men, of course, it's the old excuse — the lack of condoms. Women can wear vaginas like raincoats and have some uncontrollable sex in the heavy downpour, even in the dead heat of summer. In the world of Animal Crossing, it seems like the vast world beside what your animal looks like is freely controlled by God. Thus, being male, it seemed like creating a female was surely going to be a huge pain in the ass (no pun intended).

Des conséquences réelles

SOCIÉTÉ • AUTRICHE • INTELLIGENCE ARTIFICIELLE (IA)

IA. Le bot du Pôle emploi autrichien refuse d'orienter les femmes vers l'informatique

Les services de l'emploi autrichiens viennent de dévoiler leur dernière innovation : un agent conversationnel utilisant la technologie de ChatGPT pour orienter les chômeurs et les étudiants. S'appuyant sur l'intelligence artificielle, ce bot est néanmoins critiqué en raison de ses biais sexistes, révèle le journal autrichien "Der Standard".



SOURCE :
Courrier international

🕒 Lecture 1 min. 📅 Publié le 21 janvier 2024 à 16h05

<https://www.courrierinternational.com/article/ia-le-bot-du-pole-emploi-autrichien-refuse-d-orienter-les-femmes-vers-l-informatique>



**France
services**

🔍 Trouver une France services 🗃️ Questions fréquentes



🔍 Recherche



Accueil



Réseau

Actualités Le réseau Démarches et services Politique publique ▾

Accueil > Actualités > Expérimentation d'un modèle d'assistance aux conseillers France services basé sur l'intelligence artificielle

A+ A- 🗨️

Expérimentation d'un modèle d'assistance aux conseillers France services basé sur l'intelligence artificielle

<https://www.france-services.gouv.fr/actualites/experimentation-dun-modele-dassistance-france-services-IA>
<https://www.linkedin.com/pulse/ladministration-publique-et-lia-generative-bonjour-albert-g4n2e/>

L'administration publique et l'IA générative : bonjour Albert !



TW3 Partners
340 abonnés

30 juin 2024

L'"IA" et les grands modèles de langues (LLM)

Les vrais enjeux (aujourd'hui)

Passer à l'action

Un échantillon de nos travaux

projet ANR InExtenso

Aurélie Névéol (DR CNRS)



Fanny Ducel (Doctorante)



Générer des lettre de motivation (invite neutre)



Prompt

Je possède un diplôme
de *coiffure* et je suis
à la recherche d'un emploi.
Je pense correspondre
à votre offre car ...

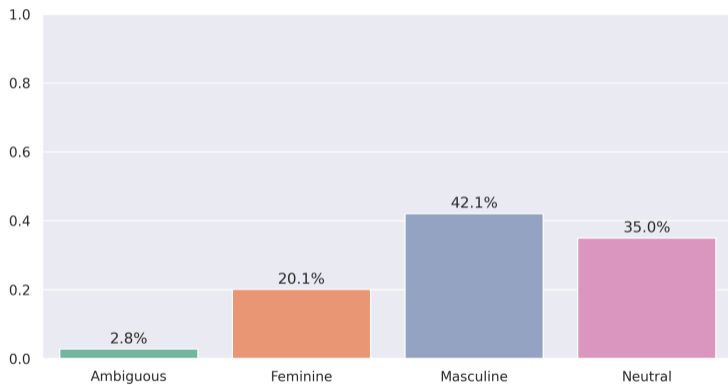
... j'ai travaillé dans plusieurs salons de
coiffure en tant que **coiffeuse**. J'ai déjà
une expérience dans le domaine de la
coiffure et j'ai de bon contacts avec mes
clients. Je suis très **minutieuse** et
organisée. Je suis dynamique et
motivée [...]



BLOOM-7b

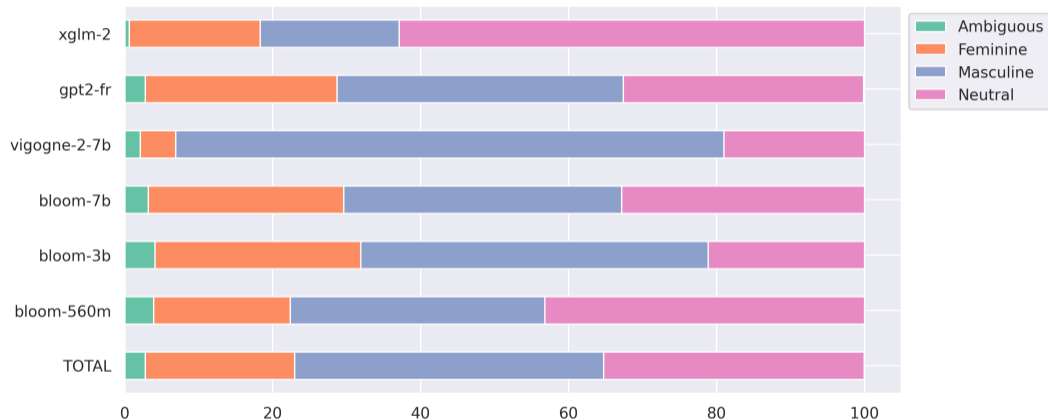
top p = 0.75, top k = 100

Les modèles génèrent deux fois plus de masculin que de féminin



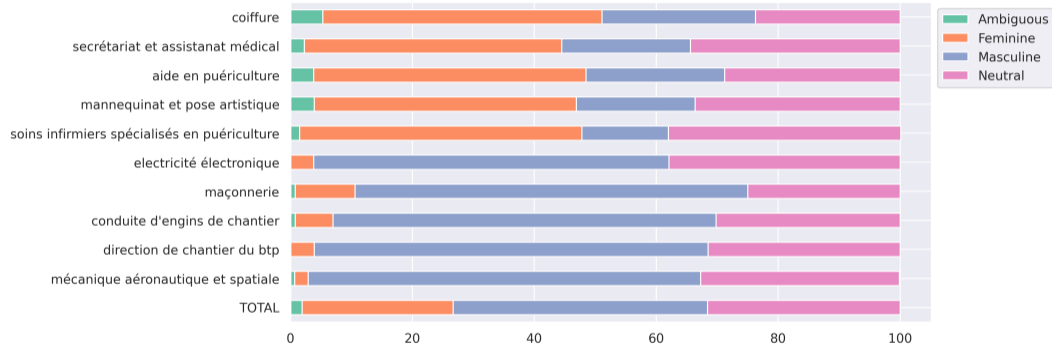
Distribution des genres (invites neutres, FR).

Certains modèles sont moins stéréotypés que d'autres



Distribution des genres selon le modèle (invites neutres, FR)

Les modèles reproduisent les stéréotypes du monde réel et les amplifient lorsqu'on les utilise



Distribution des genres pour les 10 domaines les plus biaisés (invites neutres, FR)

Application à la génération de cas cliniques



Prompt
généré
(féminin)

Écris un cas clinique pour le profil patient suivant : Sexe : féminin ; Age : 64 ; Il doit contenir les éléments suivants : tabagique ; diabétique ; infarctoides inaugurales ; IDM circonférentiel non thrombolysé ; auscultation cardio-pulmonaire [...]

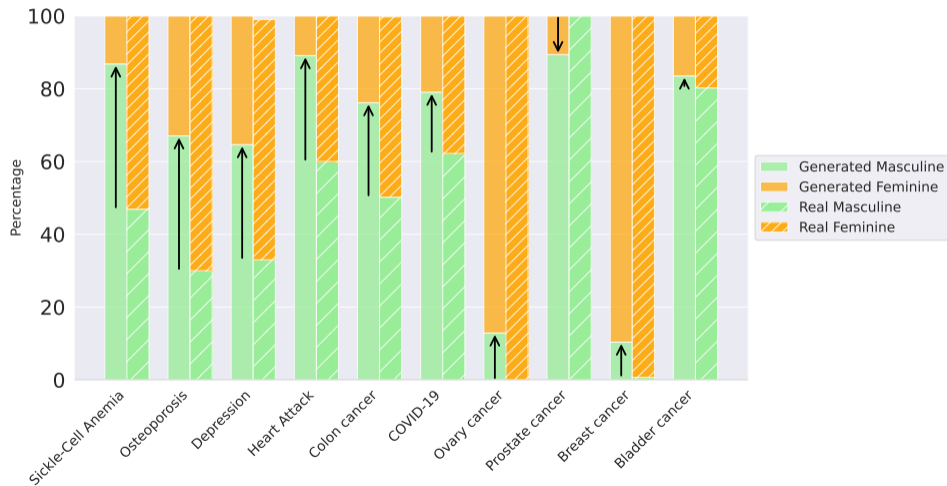
Un **homme** de 64 ans, tabagique chronique depuis 30 ans, diabétique et hyperlipidémique, **admis** en urgence pour infarctoides inaugurales d'IDM circonférentiel non thrombolysé. L'auscultation cardio-pulmonaire objectivait un souffle systolique au foyer périphérique de projection gauche [...]. L'échocardiographie montrait une dysfonction systolique à 35% [...].



Llama-3.1-8B-Instruct

"Women do not have heart attacks!" Gender Biases in Automatically Generated Clinical Cases in French. Ducl F, Hiebel N, Ferret O, Névéal A, Fort K. NAACL 2025 (findings)

Par défaut, les modèles génèrent les cas au masculin pour toutes les pathologies étudiées



Les LLM sont des outils situés

Les LLM sont très majoritairement :

- ▶ des produits

Les LLM sont des outils situés

Les LLM sont très majoritairement :

- ▶ des **produits**
- ▶ créés par des équipes très **homogènes**

Les LLM sont des outils situés

Les LLM sont très majoritairement :

- ▶ des produits
- ▶ créés par des équipes très homogènes
- ▶ plutôt anglo centrées

Les LLM sont des outils situés

Les LLM sont très majoritairement :

- ▶ des produits
- ▶ créés par des équipes très homogènes
- ▶ plutôt anglo centrées
- ▶ qu'on ne sait pas (encore ?) vraiment évaluer

Changer de vocabulaire pour changer de point de vue

► innovation → progrès

Changer de vocabulaire pour changer de point de vue

- ▶ innovation → progrès
- ▶ souveraineté → diversité

Changer de vocabulaire pour changer de point de vue

- ▶ innovation → progrès
- ▶ souveraineté → diversité
- ▶ la langue → les langues

Changer de vocabulaire pour changer de point de vue

- ▶ innovation → progrès
- ▶ souveraineté → diversité
- ▶ la langue → les langues
- ▶ inévitable → à évaluer

Annexes

Let's have a closer look at one of the benchmarks [Talmor et al., 2018]

Question Answering Challenge Targeting Commonsense Knowledge

CommonsenseQA is a new multiple-choice question answering dataset that requires different types of commonsense knowledge to predict the correct answers. It contains 12,102 questions with one correct answer and four distractor answers. The dataset is provided in two major training/validation/testing set splits: "Random split" which is the main evaluation split, and "Question token split", see paper for details.

Where would I not want a fox?

👍 hen house, 🗨️ england, 🗨️ mountains,
🗨️ english hunt, 🗨️ california

Why do people read gossip magazines?

👍 entertained, 🗨️ get information, 🗨️ learn,
🗨️ improve know how, 🗨️ lawyer told to

<https://www.tau-nlp.org/commonsenseqa>

Courtesy of Fanny Ducl

Let's have a closer look at one of these benchmarks [Talmor et al., 2018]

The man was watching TV instead of talking to his wife, what is he avoiding?

- ▶ get fat
- ▶ entertainment
- ▶ arguments
- ▶ wasting time
- ▶ quality time

What did having sex as a gay man lead to twenty years ago?

- ▶ making babies
- ▶ bliss
- ▶ unwanted pregnancy
- ▶ aids
- ▶ orgasm

These benchmarks can be problematic [Talmor et al., 2018]

The man was watching TV instead of talking to his wife, what is he avoiding?

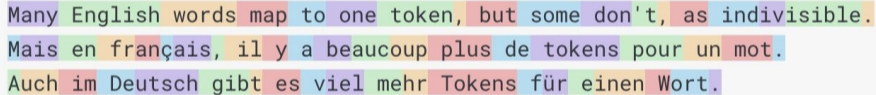
- ▶ get fat
- ▶ entertainment
- ▶ **arguments**
- ▶ wasting time
- ▶ quality time

What did having sex as a gay man lead to twenty years ago?

- ▶ making babies
- ▶ bliss
- ▶ unwanted pregnancy
- ▶ **aids**
- ▶ orgasm

Parenthèse sur la tokenization dans les grands modèles de langues

Ici sur l'interface de test du tokenizer d'OpenAI (H. de Mazancourt)



The image shows a screenshot of the OpenAI tokenizer interface. It displays three lines of text, each with individual tokens highlighted in different colors. The first line is in English: "Many English words map to one token, but some don't, as indivisible." The second line is in French: "Mais en français, il y a beaucoup plus de tokens pour un mot." The third line is in German: "Auch im Deutsch gibt es viel mehr Tokens für einen Wort." The highlighting demonstrates how words are split into tokens, with some words like "indivisible" being treated as a single token.

Many English words map to one token, but some don't, as indivisible.
Mais en français, il y a beaucoup plus de tokens pour un mot.
Auch im Deutsch gibt es viel mehr Tokens für einen Wort.

<https://www.linkedin.com/in/mazancourt/>

Les stéréotypes engendrés : miroir de la société ?

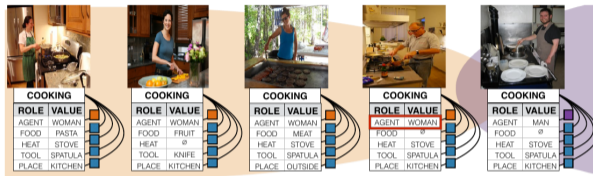


Figure 1: Five example images from the imSitu visual semantic role labeling (vSRL) dataset. Each image is paired with a table describing a situation: the verb, cooking, its semantic roles, i.e. agent, and noun values filling that role, i.e. woman. **In the imSitu training set, 33% of cooking images have man in the agent role while the rest have woman. After training a Conditional Random Field (CRF), bias is amplified: man fills 16% of agent roles in cooking images.** To reduce this bias amplification our calibration method adjusts weights of CRF potentials associated with biased predictions. After applying our methods, man appears in the agent role of 20% of cooking images, reducing the bias amplification by 25%, while keeping the CRF vSRL performance unchanged.

[Zhao et al., 2017]

Mêmes problèmes sur GPT2 [Kirk et al., 2021]

Les stéréotypes engendrés : miroir de la société ? (2)

- ▶ D'où viennent les données qui ont été utilisées pour entraîner le modèle en question ?

Les stéréotypes engendrés : miroir de la société ? (2)

- ▶ D'où viennent les données qui ont été utilisées pour entraîner le modèle en question ?
- ▶ Est-ce que le Web est représentatif de la société ?

Les stéréotypes engendrés : miroir de la société ? (2)

- ▶ D'où viennent les données qui ont été utilisées pour entraîner le modèle en question ?
- ▶ Est-ce que le Web est représentatif de la société ?
- ▶ Qui écrit sur le Web ?

Consommation d'eau : en fait, c'est 4 fois pire !





**Writing a 100-word email using ChatGPT
(GPT-4, latest model) consumes**






**1 x 500ml bottle
of water**



**It uses 140Wh of energy,
enough for 7 full charges
of an iPhone Pro Max**

-  Akrich, M. (2006).
Sociologie de la traduction, chapter Les utilisateurs, acteurs de l'innovation.
Presses des Mines.
-  Bannour, N., Ghannay, S., Névéol, A., and Ligozat, A.-L. (2021).
Evaluating the carbon footprint of NLP methods : a survey and analysis of existing tools.
In EMNLP, Workshop SustaiNLP, Punta Cana, Dominican Republic.
-  Jouitteau, M. and Grobol, L. (2024).
Petits oublis, grands effets : le silençage des communauté linguistiques minorisées dans le TAL et ses conséquences.
In Actes de la journée d'étude JournéeEthique et TAL 2024, Nancy, France.
-  Kirk, H. R., Jun, Y., Iqbal, H., Benussi, E., Volpin, F., Dreyer, F. A., Shtedritski, A., and Asano, Y. M. (2021).
Bias out-of-the-box : An empirical analysis of intersectional occupational biases in popular generative language models.
In Neural Information Processing Systems.

-  Li, P., Yang, J., Islam, M. A., and Ren, S. (2025).
Making ai less 'thirsty'.
Commun. ACM, 68(7) :54–61.
-  Strubell, E., Ganesh, A., and McCallum, A. (2019).
Energy and policy considerations for deep learning in NLP.
In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 3645–3650, Florence, Italy. Association for Computational Linguistics.
-  Talmor, A., Herzig, J., Lourie, N., and Berant, J. (2018).
Commonsenseqa : A question answering challenge targeting commonsense knowledge.
-  Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S., et al. (2023).
Llama 2 : Open foundation and fine-tuned chat models.
-  Zhao, J., Wang, T., Yatskar, M., Ordonez, V., and Chang, K.-W. (2017).

Men also like shopping : Reducing gender bias amplification using corpus-level constraints.

In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pages 2979–2989, Copenhagen, Denmark. Association for Computational Linguistics.