



Institut européen

des métiers de la **traduction** | IEMT

Université de Strasbourg

# Web, corpus, traduction : exploitations Mdèles de langue

Enzo Doyen

2025 - M1

# Pourquoi parler des modèles de langue ?

- Modèles dont le fonctionnement repose sur des corpus / jeux de données textuelles
  - L'usage principal des corpus tend maintenant vers **l'entraînement de modèles de langue**
- Possibilité d'utilisation émergente des modèles de langue pour la **réalisation d'analyses linguistiques/de corpus**
- Impact des modèles de langue sur le domaine de la traduction :
  - **automatisation des pans du métier** (MTPE, traduction automatique...) ;
  - mais aussi **accompagnement du traducteur ou de la traductrice** (aide à la recherche documentaire/traduction créative, entre autres).



# I. Introduction aux modèles de langue

# Modèles de langue

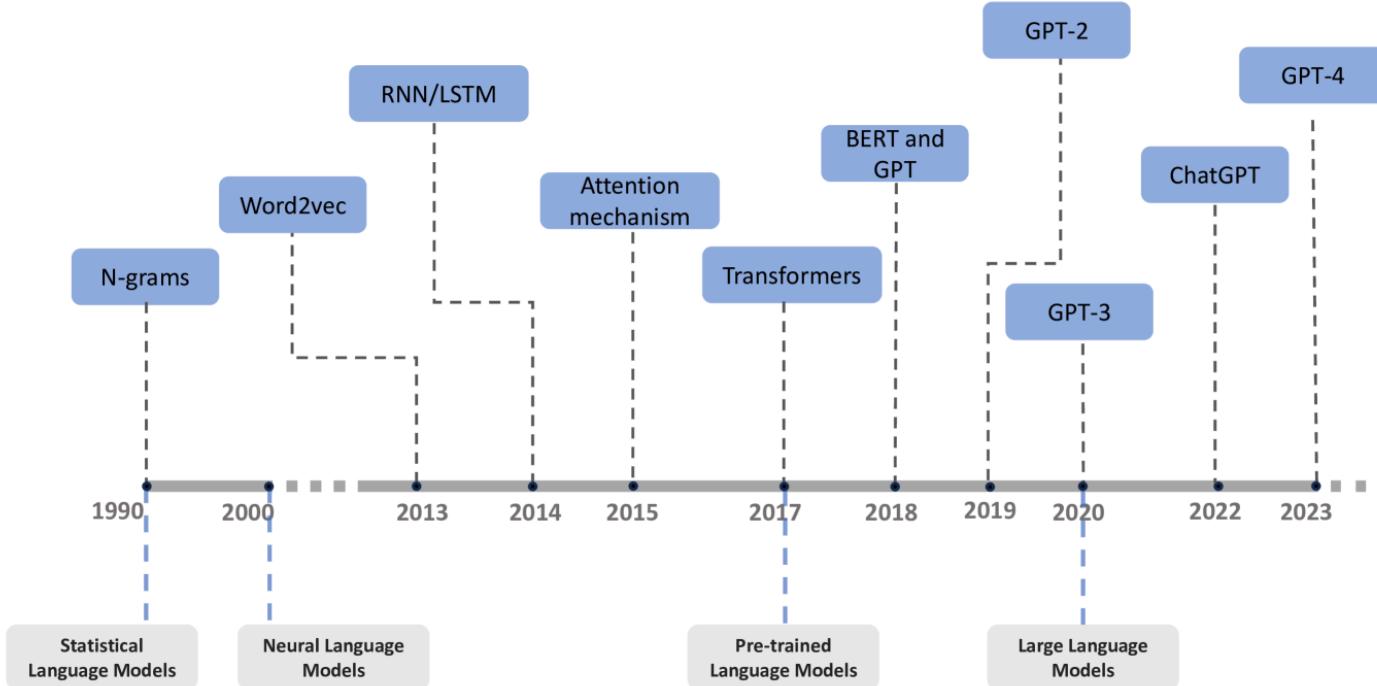
**Modèle de langue** : modèle utilisé pour prédire la probabilité d'une séquence de mots.

On distingue trois types de modèles de langue majeurs :

- les **modèles statistiques** (années 1990) ;
- les **modèles neuronaux** (années 2010) ;
- les **grands modèles de langue** (LLM) (à partir de 2017), qui incluent les **modèles à base d'instructions** (2022 et sortie de ChatGPT).

Ces modèles sont tous « entraînés » sur de grands corpus textuels pour prédire la probabilité d'une séquence de mots.

# Historique des modèles de langue



Wang et al. (2024)

# Modèles de langue statistiques

Soit la séquence de mots : « Il était une \_\_ ».

Quelle est la probabilité que le mot suivant soit « fois » ?

$$P(\text{fois} \mid \text{il était une})$$

La même logique est appliquée pour un mot  $w_m$  après une séquence de mots  $w_1, w_2, \dots, w_{m-1}$  :

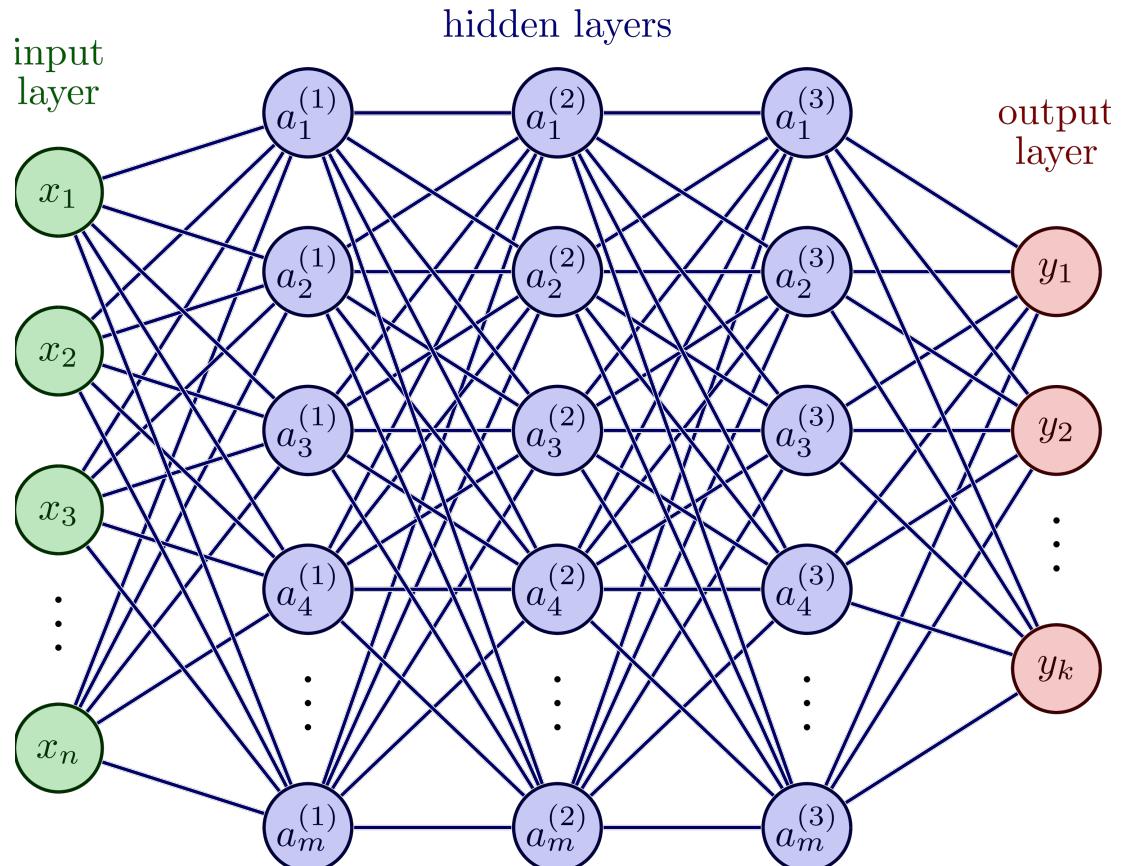
$$P(w_m \mid w_1, w_2, \dots, w_{m-1})$$

# **Modèles de langue neuronaux**

Mots convertis en vecteurs (liste de nombres) qui capturent différentes propriétés des mots en fonction de leur usage dans les corpus et, indirectement, leur signification.

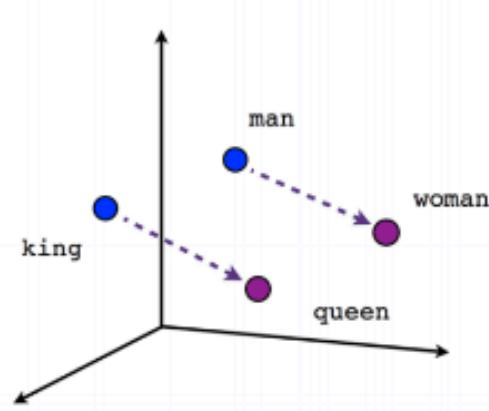
# Exemple d'un réseau de neurones

$$y = \mathbf{x}\mathbf{w} + b$$

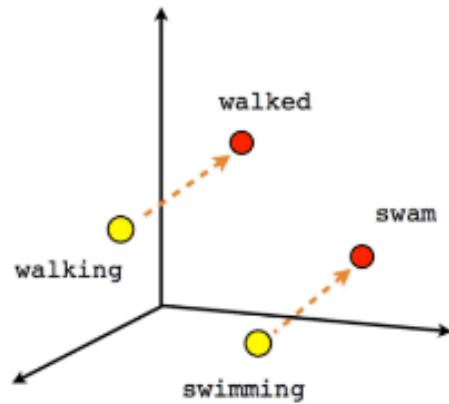


# Modèles de langue neuronaux

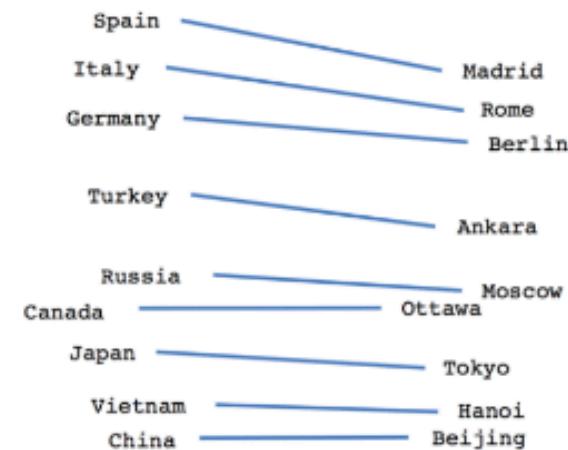
Représentation des mots dans un espace vectoriel.



Male-Female



Verb tense



Country-Capital

# Modèles de langue autorégressifs

Évolution par rapport aux modèles de langue neuronaux par un meilleur accès aux ressources : plus de données, plus de puissance de calcul, meilleures architectures (Transformers notamment).

Appellation « **grands modèles de langue** » (LLM) en référence au grand nombre de paramètres utilisés pour l'entraînement.

# Modèles de langue autorégressifs

Les élections présidentielles américaines sont un processus démocratique fondamental pour élire le président des États-Unis. Voici un aperçu de ce processus :

- Refresh un 79.69%
- les 8.88%
- il 4.91%
- une 2.49%
- le 2.20%

### 1. Calendrier des élections

Les élections présidentielles américaines se déroulent tous les quatre ans. Les deux dernières élections présidentielles ont lieu les 5 novembre de ce mois.

### 2. Éligibilité aux élections présidentielles

Pour être éligible à l'élection présidentielle, un citoyen américain doit :

- \* Être âgé de 35 ans au moins au moment de son élection.

# **Modèles à base d'instructions**

Grands modèles de langue affinés sur des instructions pour répondre à des requêtes, converser, générer du contenu textuel, etc.

Model



Instruction fine-tune on many tasks

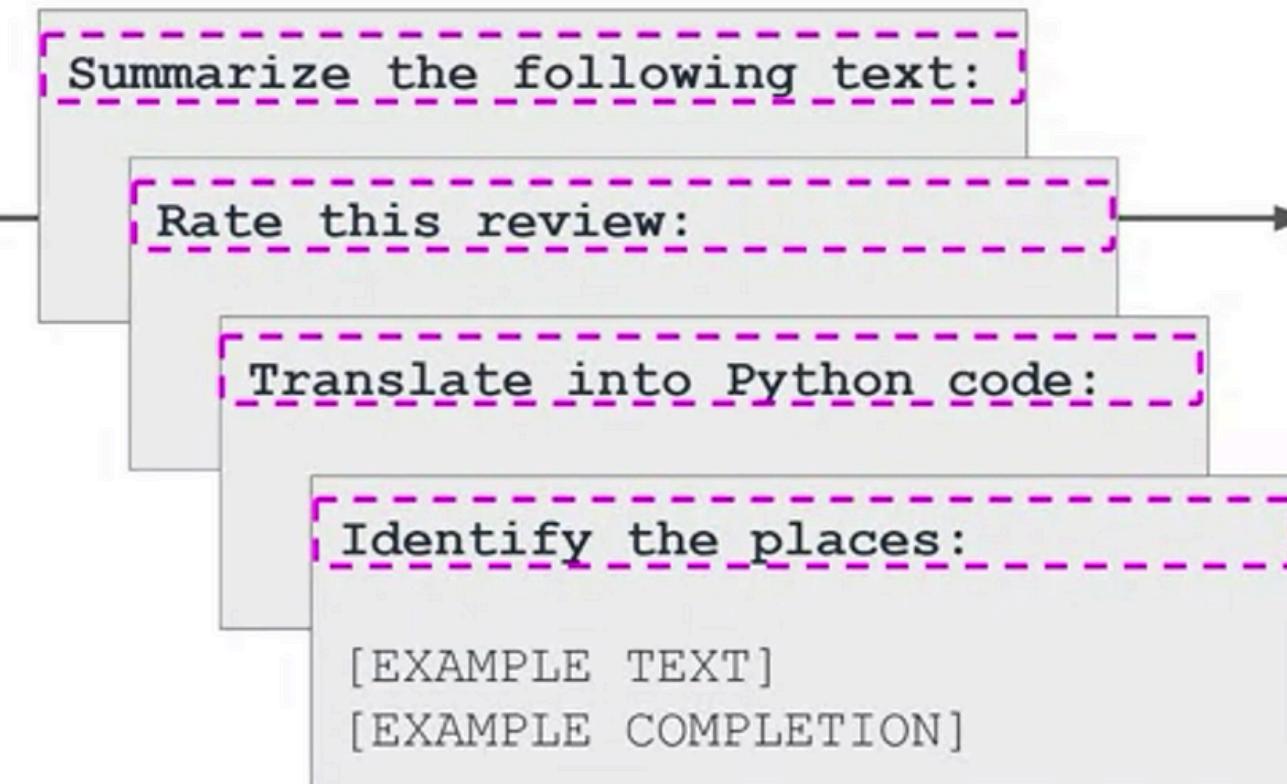


Fig. 3. – Source : <https://ogre51.medium.com/instruction-fine-tuning-of-langs-a-comprehensive-guide-e2f197e19c36>

AntConc

Target Corpus

Name: AmE06  
Files: 500  
Tokens: 1017879

KWIC Plot File View Cluster N-Gram Collocate Word Keyword Wordcloud ChatAI

Tokens prompt: 655 completion: 38 total: 693 session total: 2158

How can I help you?

You How is the word "gauge" used in the KWIC results?

gpt-3.5-turbo-0613 In the KWIC results, the word "gauge" is used in various contexts such as "gauge the fiscal impact," "gauge the public's will," "gauge the results," "gauge the vitality," "gauge of the country's economic health," "gauge his health," "gauge quantitatively," "gauge whether or not," "gauge transformations," "gauge whether they will be successful," and "gauge with respect to."

You What types of nouns does the word "gauge" collocate with?

gpt-3.5-turbo-0613 The word "gauge" collocates with nouns such as "impact," "public's will," "results," "vitality," "character," "economic health," "action," "group," "health," "glass transition temperature," "shotgun," "transformation," "knowledge," and "issue."

Inc. Legacy  Max Tokens 50 Temp 1.0

Model gpt-3.5-turbo Source KWIC

Based on the results, is the word "gauge" generally used with positive or negative sentiment?

Progress [progress bar]

Clear Submit

Time taken (creating chatai results): 1.6353 sec

Fig. 4. – Source : Laurence Anthony, Corpus AI: Integrating Large Language Models (LLMs) into a Corpus Analysis Toolkit

# **Modèles à base d'instructions dans AntConc**

- possibilité d'interroger un modèle sur l'ensemble des documents d'un corpus ;
- accès facile aux résultats proposés par AntConc ;
- analyse supplémentaire en dehors des fonctionnalités de base d'AntConc (p. ex., analyse de sentiments).



## II. Modèles de langue pour l'aide à la traduction

# Modèles de langue pour l'aide à la traduction

Est-il envisageable d'exploiter les capacités des modèles de langue (LLM) pour améliorer la traduction humaine ?

---

Sujet de la coopération entre « intelligence artificielle » et traducteur·ices abordé depuis longtemps :

- Kay (1980), *The Proper Place of Men and Machines in Language Translation*
- Melby (1992), *The translator workstation*
- Alonso et Nunes Veira (2017), *The Translator's Amanuensis 2020*

Plus récemment :

- Jiao et al. (2025), *To eat or to feed: can large language models provide useful feedback in translation education?*

# Modèles de langue pour l'aide à la traduction

Est-il envisageable d'exploiter les capacités des modèles de langue (LLM) pour améliorer la traduction humaine ?

---

**Hao et Zhuang (2025), *Investigating the Integration of LLMs into Trainee Translators' Practice and Learning***

Étude menée par **Hao et Zhuang en 2025** sur **50** apprentis traducteurs et traductrices chinois ↔ anglais invité·es à utiliser des LLM sur trois phases du processus de traduction (préproduction/production/postproduction).

**Données** : questionnaire concernant les tâches effectuées avec LLM, leur fréquence et le taux de satisfaction (1–5) des traducteur·ices.

# Coopération avec LLM pour la traduction

Scénarios	Score fréq.	Score satisf.
Summarizing the content of the source text	1.10	3.25
Highlighting key terms or phrases that require special attention	0.76	3.41
Providing background knowledge or external resources for understanding the source text	1.51	3.64
Answering specific questions about terminology, grammar, or cultural references	1.71	3.77
Suggesting appropriate style, tone, and register for the translation	0.80	3.35
Providing translation references for sentences or paragraphs	1.22	3.13
Identifying (and correcting) grammatical, semantic, or stylistic issues in the target text	1.27	3.47
Examining whether the translation meets the standard of classic translation norms like « faithfulness, expressiveness and elegance »	0.56	3.14
Providing feedback from the target audience's perspective	0.39	3.58
Providing suggestions for future translations based on past feedback	0.51	3.33

Hao et Zhuang (2025), *Investigating the Integration of LLMs into Trainee Translators' Practice and Learning*

# Coopération avec LLM pour la traduction

Scénarios	Score fréq.	Score satisf.
Summarizing the content of the source text	1.10	3.25
Highlighting key terms or phrases that require special attention	0.76	3.41
Providing background knowledge or external resources for understanding the source text	1.51	3.64
Answering specific questions about terminology, grammar, or cultural references	1.71	3.77
Suggesting appropriate style, tone, and register for the translation	0.80	3.35
Providing translation references for sentences or paragraphs	1.22	3.13
Identifying (and correcting) grammatical, semantic, or stylistic issues in the target text	1.27	3.47
Examining whether the translation meets the standard of classic translation norms like « faithfulness, expressiveness and elegance »	0.56	3.14
Providing feedback from the target audience's perspective	0.39	3.58
Providing suggestions for future translations based on past feedback	0.51	3.33

Hao et Zhuang (2025), *Investigating the Integration of LLMs into Trainee Translators' Practice and Learning*

# Coopération avec LLM pour l'interprétation

**Table 4.** Percentage data of students' responses to AI-generated feedback and peer-generated feedback.

Dimensions	Sum	Percentage			
		Agree	Argue	Disagree	Disregard
<b>Coherence</b>	<b>266</b>				
AI-generated	174	0.66	0.16	0.01	0.17
Peer-generated	92	0.78	0.01	0.00	0.19
<b>Completeness</b>	<b>294</b>				
AI-generated	193	0.57	0.31	0.03	0.09
Peer-generated	101	0.70	0.04	0.01	0.17
<b>Consistency</b>	<b>376</b>				
AI-generated	228	0.61	0.30	0.02	0.06
Peer-generated	148	0.75	0.12	0.01	0.11
<b>Terminology and grammar</b>	<b>327</b>				
AI-generated	223	0.69	0.22	0.00	0.09
Peer-generated	104	0.82	0.07	0.00	0.09
<b>Average</b>	<b>4.87</b>				
<b>Overall: AI-generated</b>	<b>818</b>	0.64	0.26	0.02	0.08
<b>Average</b>	<b>2.42</b>				
<b>Overall: Peer-generated</b>	<b>445</b>	0.78	0.08	0.00	0.13

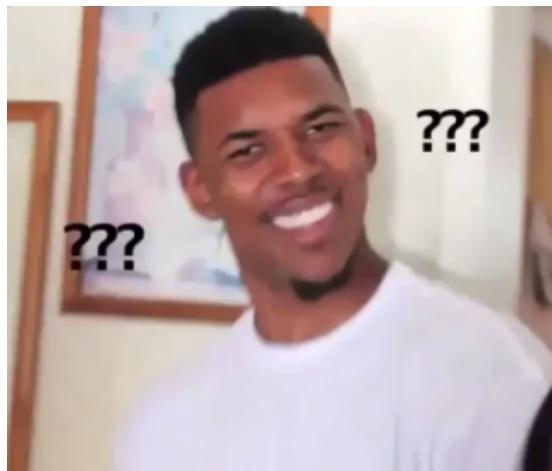
# Aide à la recherche documentaire

Vous traduisez dans un domaine technique, et tombez sur le texte suivant :

*This app is configured as a DLC. Please go to the parent app for this DLC to configure its Linux Runtime Mappings.*

# Aide à la recherche documentaire

*This app is configured as a DLC. Please go to the parent app for this DLC to configure its [Linux Runtime Mappings](#).*



do you know what is a "Linux runtime mapping"?



The term "Linux runtime mapping" isn't a standard, widely recognized term on its own, but it can be interpreted based on the context of Linux systems, runtime environments, and memory mapping.

## Possible Interpretations of "Linux Runtime Mapping"

### 1. Memory Mapping at Runtime:

- This could refer to the process where the Linux operating system maps files or devices into the address space of a process at runtime. For example, using `mmap()` to map a file to memory, allowing a program to access the file as if it were a part of its memory.
- Memory mappings are crucial for managing how processes interact with physical memory and virtual memory. These mappings occur at runtime and are managed by the kernel.

### 2. Dynamic Linker and Runtime Symbol Mapping:

- In the context of executing programs, "runtime mapping" might refer to how the dynamic linker (`ld-linux.so`) maps shared libraries and resolves symbols (functions, variables) at runtime.
- When a program is executed, the linker maps the necessary shared libraries into the process's address space and ensures that all function calls and references are correctly linked.

# Aide à la recherche documentaire : ajout de contexte



Given the context you've provided, "Linux runtime mapping" within the Steam Runtime SDK environment refers to the method by which Steam ensures that games and applications run in a consistent and predictable runtime environment across various Linux distributions. Here's how this works:

## 1. **LD\_LIBRARY\_PATH Runtime (Steam Runtime 1 'scout')**:

- The Steam client and many native Linux games rely on a predefined set of libraries provided by the Steam Runtime. These libraries are stored in a specific directory (e.g., `~/.steam/root/ubuntu12_32/steam-runtime`).
- **Runtime Mapping** in this context refers to the process of adjusting the `LD_LIBRARY_PATH` environment variable to include paths to these libraries. This ensures that the games use the exact versions of libraries provided by Steam, rather than those installed by the system.

# Aide à la recherche documentaire (mais douteux pour la trad.)

given the context, what could be a good translation of "Linux runtime mapping" into French?

 Given the context of Steam's Linux runtime, a good translation of "Linux runtime mapping" into French could be:

**"Cartographie de l'environnement d'exécution Linux"**

Here's a breakdown:

- **"Cartographie"** refers to mapping, which in this case involves how libraries or environments are organized or configured.
- **"Environnement d'exécution"** is the common translation for "runtime," representing the set of libraries and configurations needed for software to run.
- **"Linux"** remains unchanged, as it refers to the operating system.

# Aide à la recherche documentaire (mais douteux pour la trad.)

**Texte source** : *This app is configured as a DLC. Please go to the parent app for this DLC to configure its Linux Runtime Mappings.*

**Traduction finale** : ~~Cartographie~~ **Mappage** de l'environnement d'exécution  
Linux

# Aide à la recherche documentaire

Autre exemple :

*Feet IK now repositions to idle pose if the feet have pinned at a largely different pose to that authored*

# Aide à la recherche documentaire

Essayons une instruction plus complète :

You are an advanced AI system with **expertise across all domains**. Your task is to **analyze and explain sentences or sets of sentences** provided by users, **along with any accompanying context**. Focus on **clarifying technical terms, concepts, and nuances** within the given text. **Utilize the context provided to enhance your understanding and provide more accurate explanations**. Your goal is to ensure the user comprehends the sentence(s) perfectly. **Offer detailed, yet accessible explanations of complex ideas, breaking them down into simpler components when necessary**. If multiple interpretations are possible, present them clearly. **Draw connections between different elements of the text** when relevant. **If any ambiguities remain, highlight them and explain why they exist**. Be prepared to **answer follow-up questions about any aspect of your explanation**. Maintain a professional, informative tone throughout your responses. Begin your analysis now, based on the user's input.

Sentence: Feet IK now repositions to idle pose if the feet have pinned at a largely different pose to that authored

Context: first-person shooter video game changelog



In this sentence, the changelog entry describes a change to how the "Feet IK" system behaves in a first-person shooter video game. Here's a breakdown of the terms and concepts involved:

1. **Feet IK:** "IK" stands for **Inverse Kinematics**, a technique used in animation and game development. It allows the game engine to automatically adjust a character's limb positions, like their feet, to align with the environment or animations realistically. For example, if a character walks on uneven ground, the IK system can adjust the character's feet to match the surface.
2. **Repositions to idle pose:** This means that the Feet IK system will adjust the character's feet to a default "idle" position. The "idle pose" is a standard, neutral stance that a character returns to when not actively performing another action.
3. **If the feet have pinned at a largely different pose to that authored:** This indicates that sometimes the character's feet may end up "pinned" or locked into a position that is significantly different from what was originally intended or "authored" by the animators. This could occur due to various in-game actions or environmental interactions.

So, the change described here is that if the character's feet end up locked in an unusual or unintended position, the game will now automatically reset them back to the idle position. This likely improves visual consistency and helps avoid awkward animations when the feet are stuck in an

# Aide à la recherche documentaire

**Texte source** : *Feet IK now repositions to idle pose if the feet have pinned at a largely different pose to that authored*

**Traduction finale** : *Les pieds en cinématique inverse se reposent maintenant à la pose de repos s'ils ont été fixés dans une pose très différente de celle d'origine.*

# Aide à la recherche documentaire

- importance de rédiger des instructions (*prompts*) précises ;
- privilégier la rédaction des instructions en anglais (langue la plus présente dans les données d'entraînement des modèles) ;
- accompagner sa requête de contexte (type de projet, autres textes similaires, informations trouvées en ligne, etc.) ;
- résultats dépendants des corpus sur lesquels les modèles ont été entraînés.

# Aide à la traduction créative

Soit le texte suivant à traduire, utilisé pour désigner un élément cosmétique avec un corbeau :

*Ravenous Raven*

# Aide à la traduction créative

Soit le texte suivant à traduire, utilisé pour désigner un élément cosmétique avec un corbeau :

*Ravenous Raven*

# Aide à la traduction créative

Soit le texte suivant à traduire, utilisé pour désigner un élément cosmétique avec un corbeau :

*Ravenous Raven*

*Corbeau ... ?*

# Aide à la traduction créative

Soit le texte suivant à traduire, utilisé pour désigner un élément cosmétique avec un corbeau :

*Ravenous Raven*

*Corbeau vorace ?* 

give synonyms of "vorace" in French starting with "c"



Here are some synonyms of "vorace" in French that start with "c":

- **Carnassier** – often used to describe animals with a predatory or meat-eating nature.
- **Cupide** – implies greed or an intense desire, though it usually refers to money or material wealth rather than appetite.
- **Convulsif** – though slightly less direct, this can describe an intense, almost uncontrollable craving or action.

Let me know if you're looking for other synonyms or particular nuances!

give synonyms of "vorace" in French starting with "c"



Here are some synonyms of "vorace" in French that start with "c":

- **Carnassier** – often used to describe animals with a predatory or meat-eating nature.
- **Cupide** – implies greed or an intense desire, though it usually refers to money or material wealth rather than appetite.
- **Convulsif** – though slightly less direct, this can describe an intense, almost uncontrollable craving or action.

Let me know if you're looking for other synonyms or particular nuances!

# Aide à la traduction créative

Soit le texte suivant à traduire, utilisé pour désigner un élément cosmétique avec un corbeau :

*Ravenous Raven*

*Corbeau carnassier*

Préservation de l'allitération et du sens !

# Vérification dans Sketch Engine

## CONCORDANCE

Get more space



French Web 2023 (frTenTen23)



simple **corbeau carnassier** • 2



less than 0.01 • 7.2e-9%



Details

Left context

KWIC

Right context

1	<input type="checkbox"/> <a href="#">(i) rosamystica.fr</a>	domine, c'est le fief de l'âpre dominer , ce être le fief de le âpre	<b>corbeau carnassier</b> corbeau carnassier	.</s><s>Disons, pour préciser . dire , pour préciser	
2	<input type="checkbox"/> <a href="#">(i) fandom.com</a>	ecs apparaissent comme des cs apparaître comme de+les	<b>corbeaux carnassiers</b> corbeau carnassier	, alors qu'ils semblaient juste ê , alors que ils sembler juste ê	

# Aide à la traduction créative

Possibilité d'obtenir des résultats encore meilleurs en :

- fournissant du contexte ;
- fournissant des exemples de traductions précédentes ;
- utilisant des instructions plus complexes.

# LMArena

<https://lmarena.ai/>

- plateforme en ligne pour tester différentes versions de modèles de langue (propriétaires et open source) ;
- possibilité d'utiliser les versions payantes de modèles gratuitement (sous réserve de ne pas dépasser le cap de requêtes).

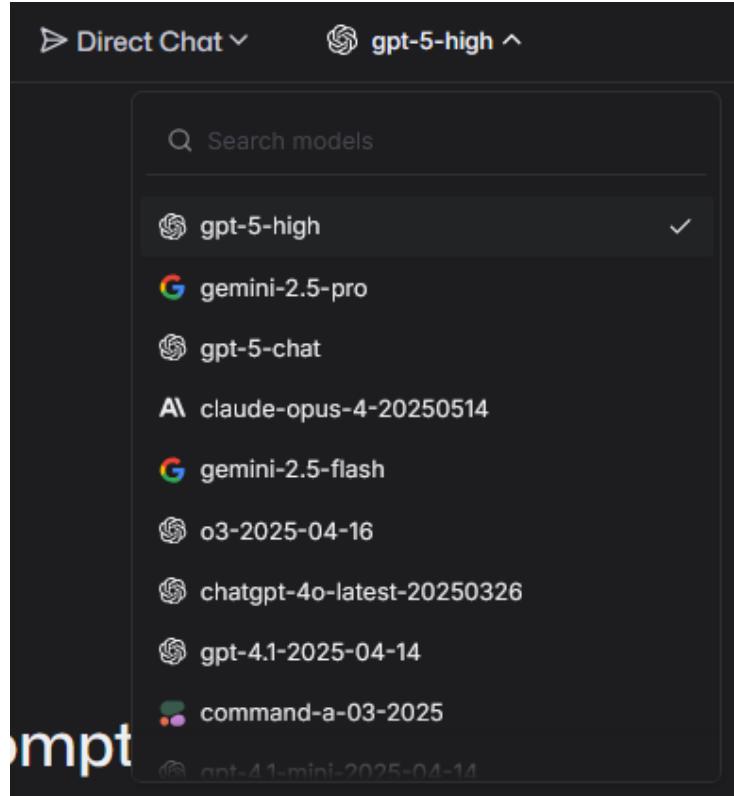
# LMArena

<https://lmarena.ai/>

- plateforme en ligne pour tester différentes versions de modèles de langue (propriétaires et open source) ;
- possibilité d'utiliser les versions payantes de modèles gratuitement (sous réserve de ne pas dépasser le cap de requêtes).

**Attention : site à visée académique ; les messages sont conservés et peuvent être utilisés à des fins de recherche.**

# LMArena



# Pour aller plus loin

- 3Blue1Brown, 2024. *Large Language Models explained briefly* (lien sur Moodle)
- Daniel Jurafsky and James H. Martin. 2024. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models, 3rd edition*. Online manuscript released August 20, 2024. [https://web.  
stanford.edu/~jurafsky/slp3](https://web.stanford.edu/~jurafsky/slp3).
- Cours de l'Université de Stanford CS224N: *Natural Language Processing with Deep Learning* : <https://web.stanford.edu/class/cs224n/>.