

# Résumé automatique pour le juridique

Atelier Décoder l'IA, Institut Robert Badinter,  
17 Mars 2026

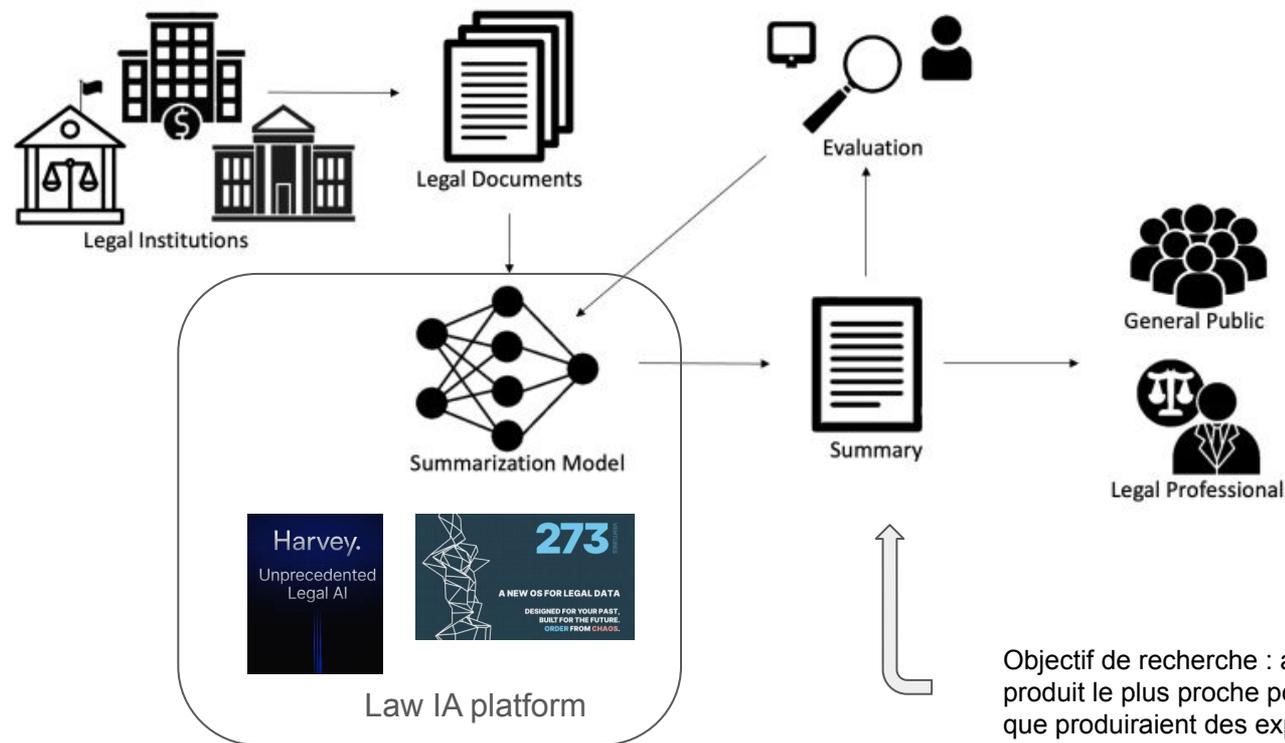
Vincent Guigue  
[vincent.guigue@agroparistech.fr](mailto:vincent.guigue@agroparistech.fr)



Karen Pinel-Sauvagnat  
[karen.sauvagnat@irit.fr](mailto:karen.sauvagnat@irit.fr)



# Pipeline du résumé automatique de documents juridiques



Objectif de recherche : avoir un résumé produit le plus proche possible de celui que produiraient des experts



Mousumi Akter, Erion Çano, Erik Weber, Dennis Dobler and Ivan Habernal : *A Comprehensive Survey on Legal Summarization: Challenges and Future Directions*, 2025, arXiv - <https://arxiv.org/abs/2501.17830>



Congqing He, Yanli Lid, Hao Zhan, Qunjun Zhang: *A survey of large language models for legal tasks: Progress, prospects and challenges*. Computer Science review 60 (2026),



## COMMISSION DELEGATED REGULATION (EU) 2019/980

of 14 March 2019

**supplementing Regulation (EU) 2017/1129 of the European Parliament and of the Council as regards the format, content, scrutiny and approval of the prospectus to be published when securities are offered to the public or admitted to trading on a regulated market, and repealing Commission Regulation (EC) No 809/2004**

(Text with EEA relevance)

THE EUROPEAN COMMISSION,

Having regard to the Treaty on the Functioning of the European Union,

Having regard to Regulation (EU) 2017/1129 of the European Parliament and of the Council of 14 June 2017 on the prospectus to be published when securities are offered to the public or admitted to trading on a regulated market, and repealing Directive 2003/71/EC <sup>(1)</sup>, and in particular Article 9(14), paragraphs (1) and (2) of Article 13, Articles 14(3), 15(2) and 20(11) thereof,

Whereas:

- (1) Regulation (EU) 2017/1129 lays down requirements to be complied with when drawing up prospectuses. The requirements concerning the scrutiny, review, approval and filing of the universal registration document and any amendments thereto, the requirements concerning the format of the prospectus, the base prospectus and the final terms, the specific information to be included in a prospectus, the minimum information to be included in the universal registration document, the reduced information to be included under the simplified disclosure regime for secondary issuances, the reduced content, the standardised format and the sequence of the EU Growth prospectus, the reduced content and standardised format of the specific summary and the scrutiny and approval of prospectuses all need to be specified.
- (2) The content and the format of a prospectus depend on a variety of factors, such as the type of issuer, type of security, type of issuance as well as the possible involvement of a third party as a guarantor and the question of whether or not there is an admission to trading. It is therefore not appropriate to lay down the same requirements for all types of prospectuses. Specific information requirements should be laid down instead and should be combined depending on those factors and the type of prospectus. This should however not prevent an issuer, offeror or person asking for admission to trading on a regulated market to provide in the prospectus the most comprehensive information available.
- (3) To ensure legal certainty and increase transparency for investors, issuers should state in their universal registration document whether the universal registration document has been approved by the competent authority or merely has been filed and published without prior approval.
- (4) The alleviated information requirements for secondary issuances should reflect the specificities of equity and non-equity securities.



EUR-Lex

Access to European Union law



Document juridique

...

## WHAT IS THE AIM OF THE REGULATIONS?

- **Regulation (EU) 2017/1129** (the prospectus regulation), as supplemented by **Delegated Regulations (EU) 2019/979**, **(EU) 2019/980** and **(EU) 2021/528**, aims to help companies, including small and medium-sized enterprises (SMEs), access different forms of finance in the **European Union (EU)**. It does so by simplifying and streamlining the rules and procedures for drawing up, approving and distributing the **prospectus** a company publishes when offering **securities** to the public or admitting securities to trading on a regulated market.
- The legislation reduces costly and burdensome red tape for companies and enables investors to make the right investment decision by providing comprehensible, easy-to-analyse and concise information.



... Et résumé produit par des experts

### Delegated acts

The **European Commission** has adopted the following acts:

- **Delegated Regulation (EU) 2019/979** laying down details of the key financial information in the summary of the prospectus, the publication and classification of the prospectus, the advertisement for the securities and the supplement to a prospectus;
- **Delegated Regulation (EU) 2019/980** laying down details of the precise content and format of the prospectus, and concerning the scrutiny and approval of the prospectus;
- **Delegated Regulation (EU) 2021/528** setting out the minimum information content of the document to be published for a prospectus exemption in connection with a takeover by means of an exchange offer, a merger or a division;
- **Delegated Regulation (EU) 2020/1272** amending and correcting Delegated Regulation (EU) 2019/979;
- **Delegated Regulation (EU) 2020/1273** amending and correcting Delegated Regulation (EU) 2019/980.

D'un point de vue du chercheur en TAL, en plus des difficultés usuelles liées au résumé automatique, des caractéristiques fortes :

- vocabulaire très spécifique
- **contextualisation avec de nouvelles informations**

Comment automatiser  
cette tâche de synthèse ?

Promesses, risques et usages  
maîtrisés

Comment fonctionnent les  
modèles de résumé  
automatique ?

Comment évalue-t-on la qualité  
des résumés produits ?

Quelles adaptations sont  
nécessaires aux domaines  
considérés ?

# Modèles de langue : du pré-entraînement au résumé automatique

# Rappels sur le fonctionnement des modèles de langue

## Statistical Modeling of Texts

Texts splitting = tokens

Large Language Models (LLMs), such as GPT-3 and GPT-4, utilize a process called tokenization. Tokenization involves breaking down text into smaller units, known as tokens, which the model can process and understand. These tokens can range from individual characters to entire words or even larger chunks, depending on the model. For GPT-3 and GPT-4, a Byte Pair Encoding (BPE) tokenizer is used. BPE is a subword tok

Token forecasting = Probabilistic model

Input = Previous tokens

Language Model

Output = Predicted next tokens

Dictionary	Large	0.02
	entire	0.01
	For	0.00
	units	0.00
	...	...
	can	0.00
	may	0.00
...	0.09	
...	...	
...	<b>0.30</b>	

Iterative process = writing text

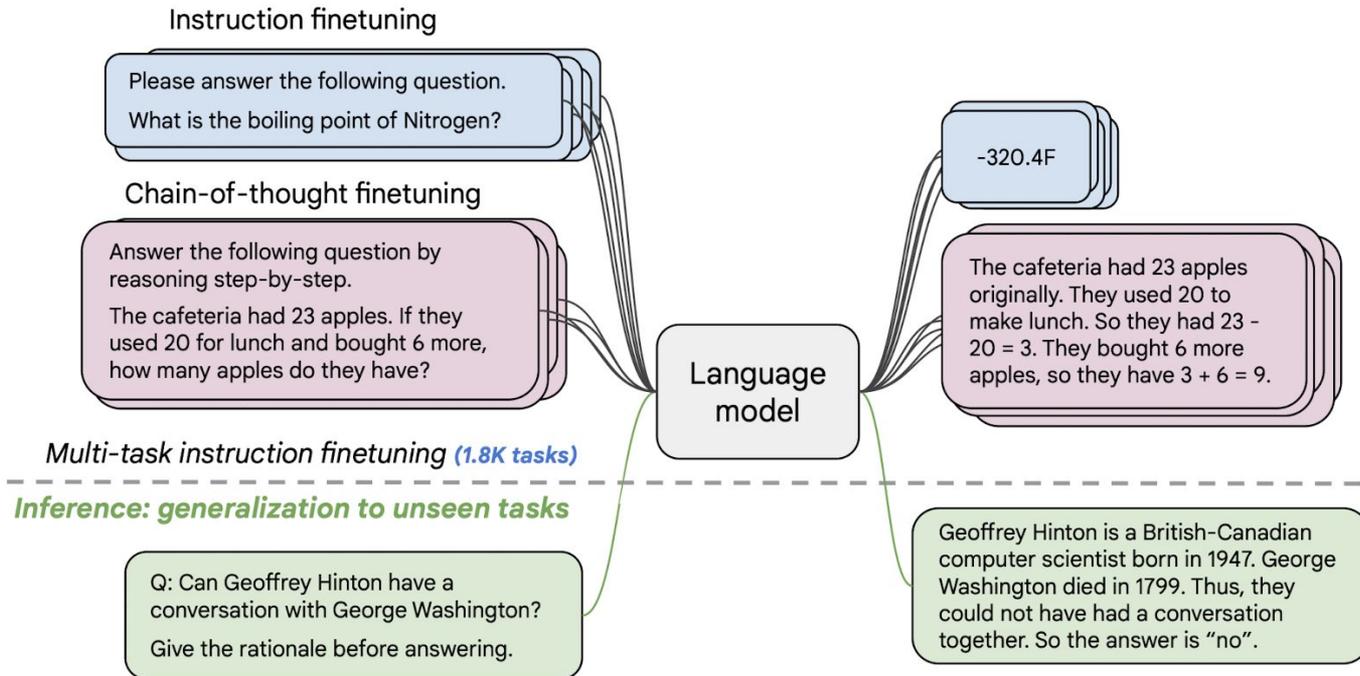
# Vers un apprentissage des tâches

Pré-entraînement = 80% du budget de calcul

Objectif : développer la mémoire paramétrique

Affinage sur les tâches = 20% du budget de calcul...

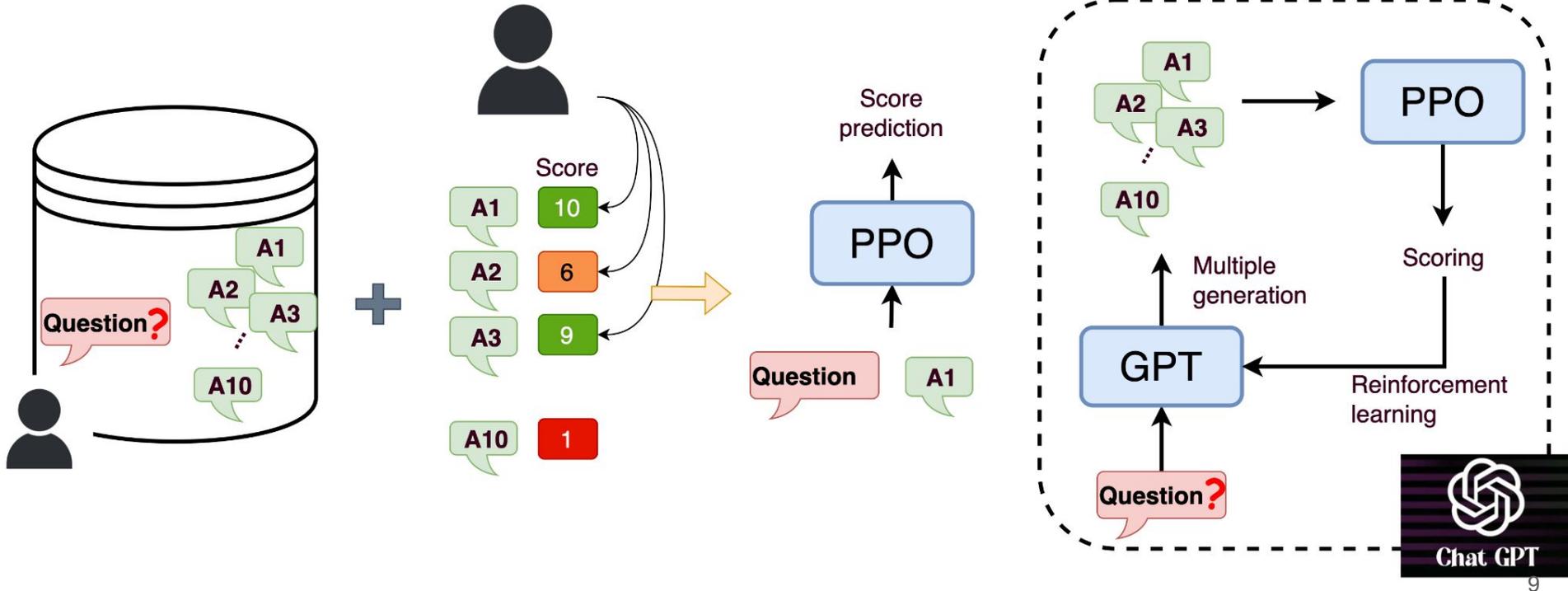
>50% du budget de construction du modèle



# Alignement

Rendre un modèle de langue acceptable du point de vue sociétal

= Développement d'une ligne éditoriale

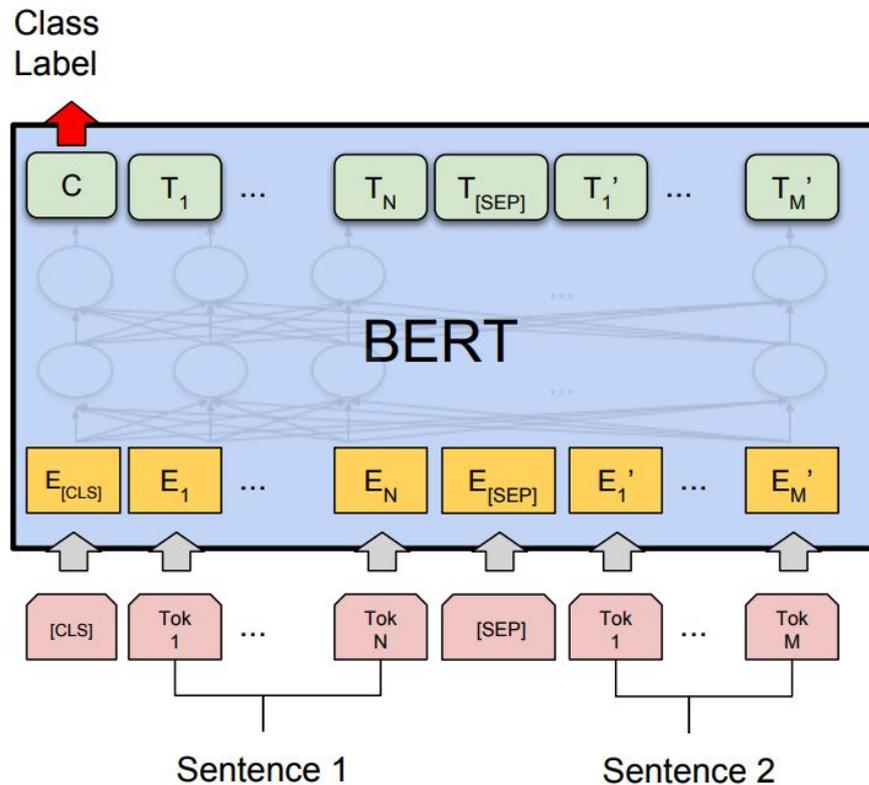


# Développement des capacités en résumé automatique

Différents critères auto-supervisés

Les deux phrases se suivent-elles dans le document?

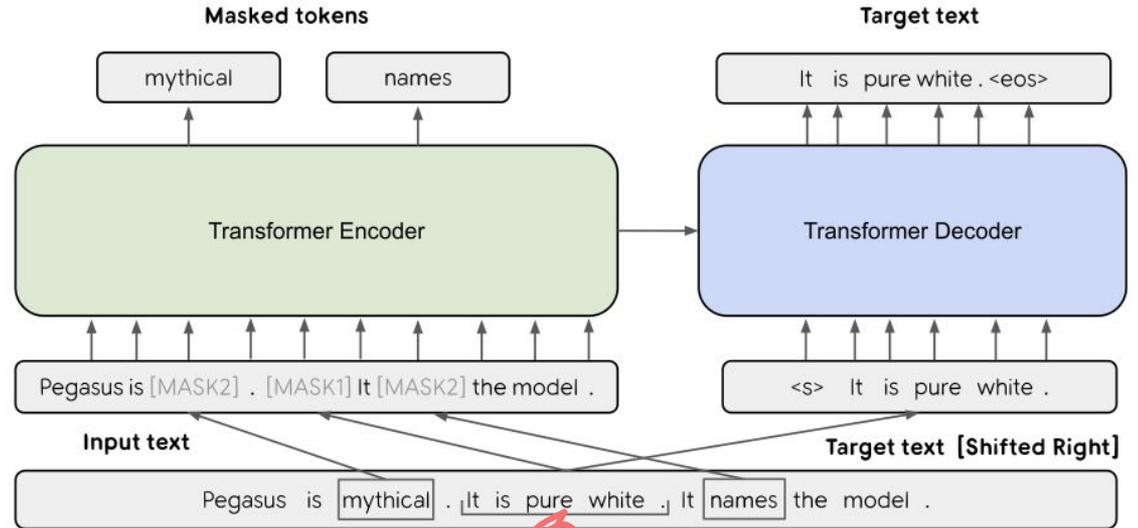
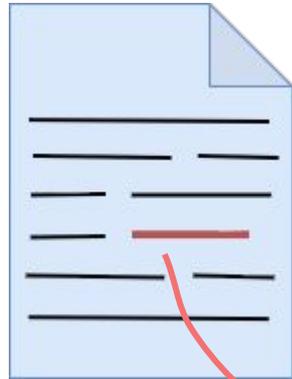
Apprendre à agréger la signification des tokens dans une phrase



# Vers le résumé automatique

Un entraînement spécifique : reconstruire des **phrases clés**

Idée: la phrase qui *partage* le plus de mots avec les autres



# Résumé automatique

Comment imposer la **longueur**, le **style**, la **technicité**, la nature des informations ciblées?

=> Entraînements spécifiques / Instruction tuning

# ELI5

Explain Like I'm 5



Instruction	<MAX_LEN>	Generation	Length
Answer the following instruction using <MAX_LEN> words or less. I'm trying to teach myself to have nicer handwriting. Can you help?	20	Yes, I'd be happy to help you improve your handwriting skills!	12
	40	I'd be happy to help. To improve your handwriting, practice regularly, focus on posture and grip, and try using guidelines or a light box to help you stay within lines.	31
	80	I'd be happy to help you improve your handwriting. Here are a few tips to get you started: * Choose the right tools: Use a pen with good grip and smooth ink flow. Experiment with different types of paper to find one that works well for you. * Practice regularly: Set aside time each day to practice writing. * Focus on posture and grip: Keep your hand steady and your fingers relaxed.	70
I'm trying to teach myself to have nicer handwriting. Can you help?	N/A	Improving your handwriting takes practice, patience, and some guidance. I'd be happy to help you with that. Here are some tips to get you started: 1. <b>**Choose the right tools**</b> : Invest in a few good quality pens or pencils that feel comfortable in your hand. Experiment with different types of pens (e.g., fine tip, medium tip, calligraphy) to find what works best for you. 2. <b>**Practice basic strokes**</b> : Focus on mastering basic strokes such as: .....	456

Following Length

Constraints in Instructions

12

# Du prompt tuning à la personnalisation

Malgré les nombreuses aptitudes des LLM

=> **Raffiner** un modèle de langue sur une **tâche spécifique**

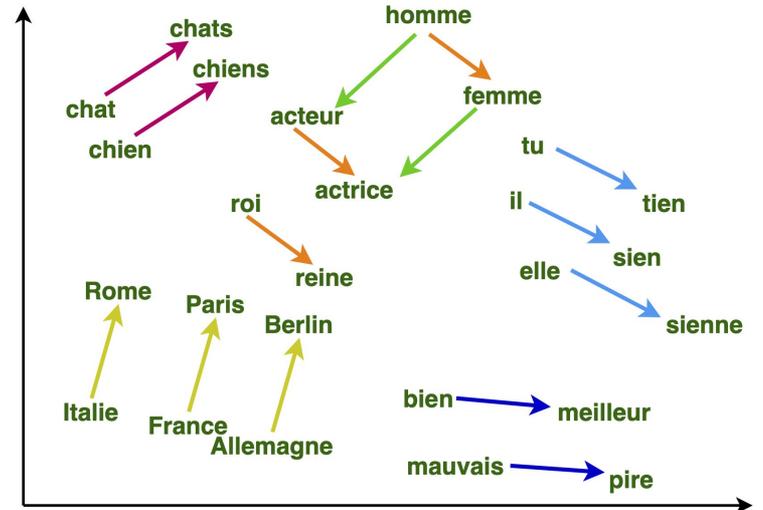
- Utiliser des modèles plus petits
- S'adapter à un vocabulaire spécifique
- Maximiser les performances

(1) Apprendre les positions des mots

(2) Agréger les représentations

(3) Raffiner l'architecture

he curtains open and the moon shining in on the barely  
ars and the cold , close moon " . And neither of the w  
rough the night with the moon shining so brightly , it  
made in the light of the moon . It all boils down , wr  
surely under a crescent moon , thrilled by ice-white  
sun , the seasons of the moon ? Home , alone , Jay pla  
m is dazzling snow , the moon has risen full and cold  
un and the temple of the moon , driving out of the hug  
in the dark and now the moon rises , full and amber a  
bird on the shape of the moon over the trees in front  
But I could n't see the moon or the stars , only the  
rning , with a sliver of moon hanging among the stars  
they love the sun , the moon and the stars . None of  
the light of an enormous moon . The splash of flowing w  
man 's first step on the moon ; various exhibits , aer  
the inevitable piece of moon rock . Housing The Airsh  
oud obscured part of the moon . The Allied guns behind



# Du prompt tuning à la personnalisation

Malgré les nombreuses aptitudes des LLM

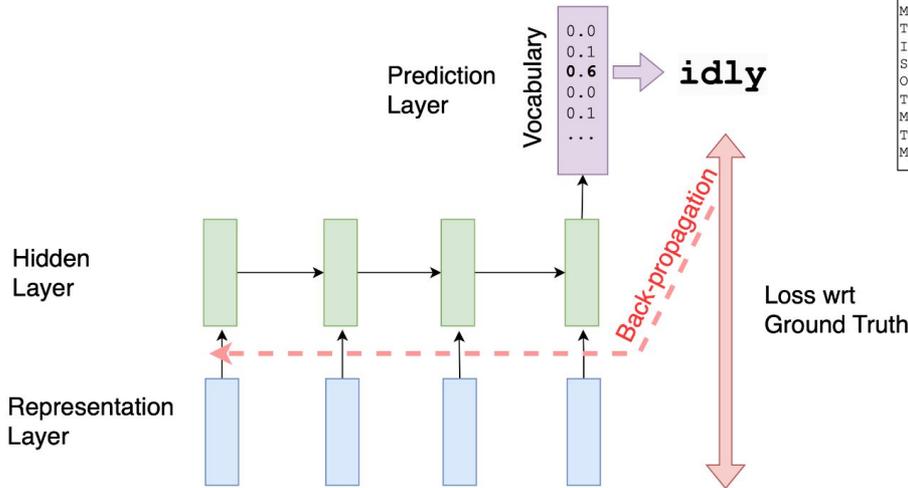
=> **Raffiner** un modèle de langue sur une **tâche spécifique**

- Utiliser des modèles plus petits
- S'adapter à un vocabulaire spécifique
- Maximiser les performances

(1) Apprendre les positions des mots

(2) Agréger les représentation

(3) Raffiner l'architecture



```
The fluffy cat napped lazily in the sunbeam.  
I adopted a stray cat from the shelter last week.  
My cat loves to chase after toy mice.  
The black cat stealthily crept through the dark alley.  
I often find my cat perched on the windowsill, watching birds.  
She gently stroked her cat's fur as it purred contentedly.  
Our neighbor's cat frequently visits our backyard.  
The playful cat swatted at the dangling string with its paw.  
My cat has a preference for fish flavored cat food.  
The cat stealthily stalked a mouse in the garden.  
My grandmother has a collection of porcelain cat figurines.
```

Corpus

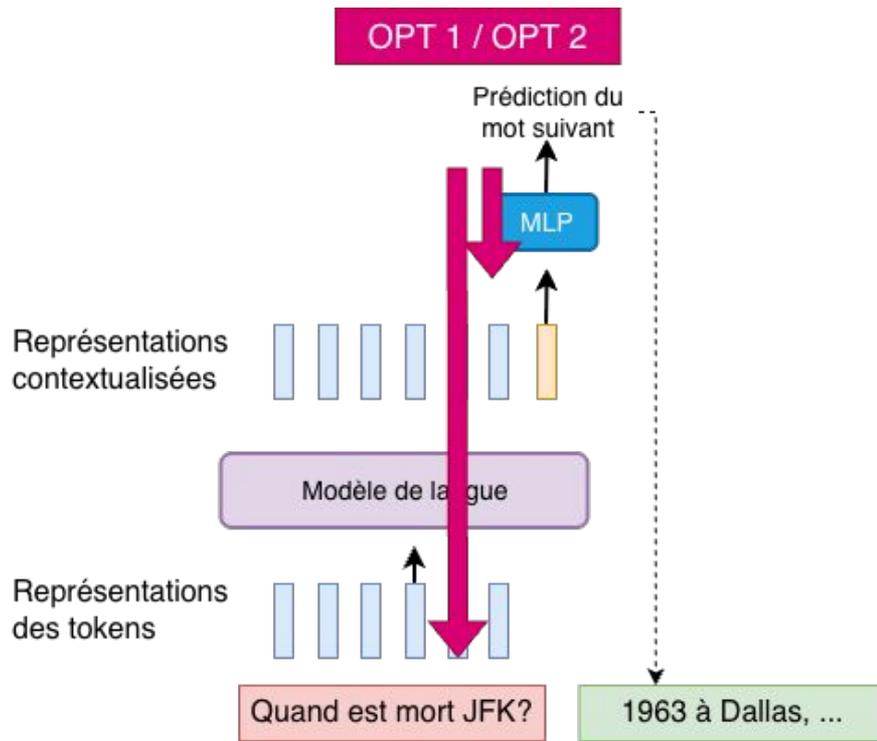
The fluffy cat napped lazily in the sunbeam.

# Du prompt tuning à la personnalisation

Malgré les nombreuses aptitudes des LLM

=> **Raffiner** un modèle de langue sur une **tâche spécifique**

- Utiliser des modèles plus petits
- S'adapter à un vocabulaire spécifique
- Maximiser les performances



(3) Raffiner l'architecture

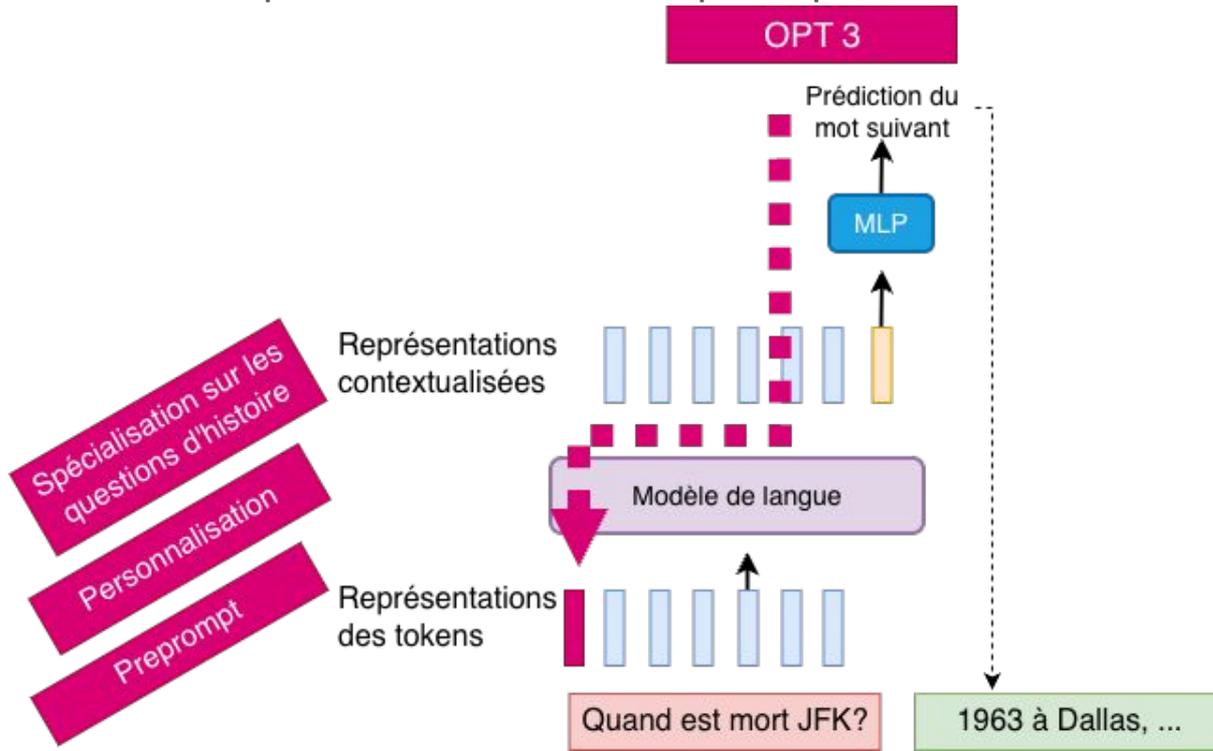
# Du prompt tuning à la personnalisation

Malgré les nombreuses aptitudes des LLM

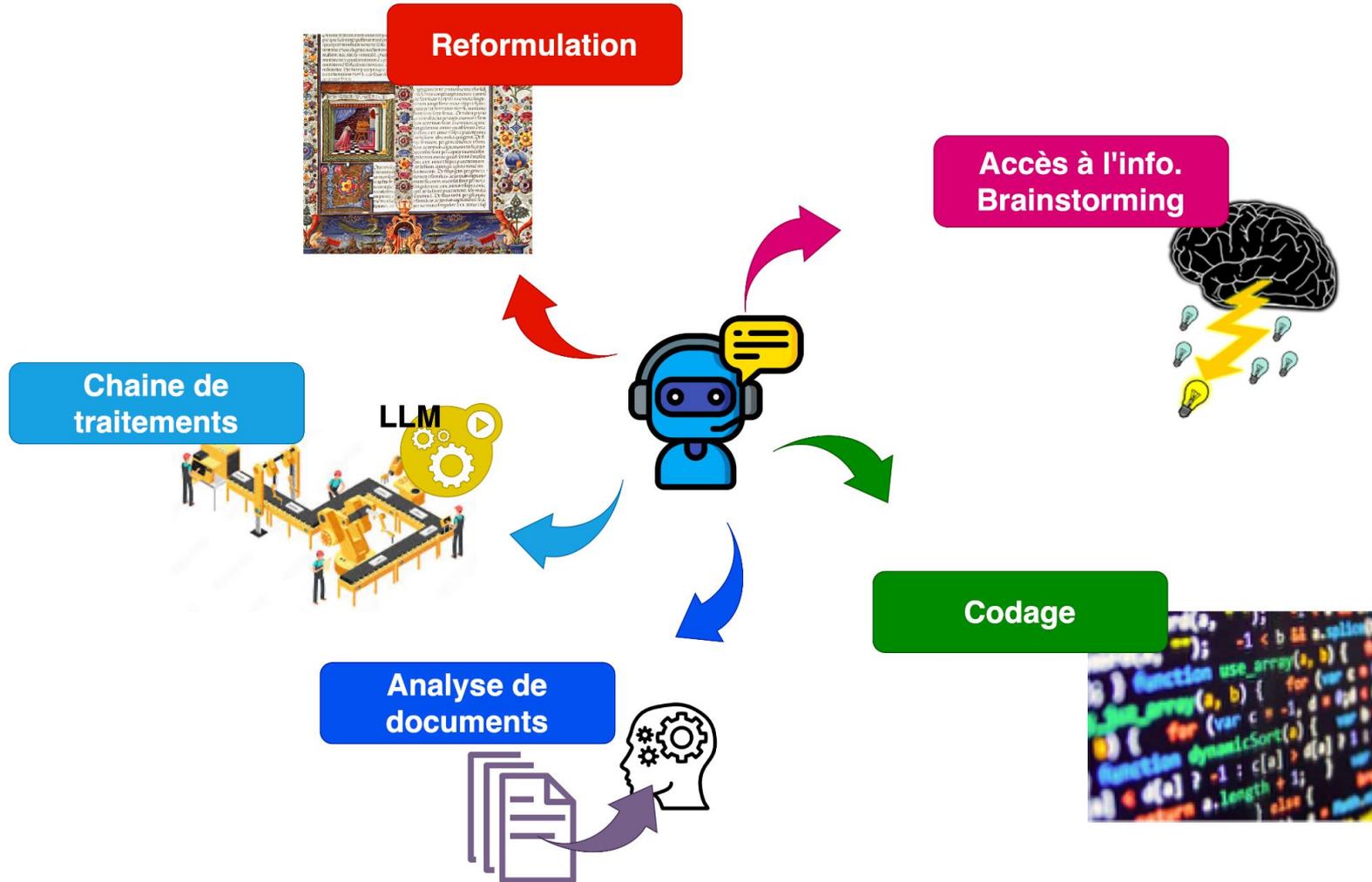
=> **Raffiner** un modèle de langue sur une **tâche spécifique**

- Utiliser des modèles plus petits
- S'adapter à un vocabulaire spécifique
- Maximiser les performances

(3) Raffiner l'architecture

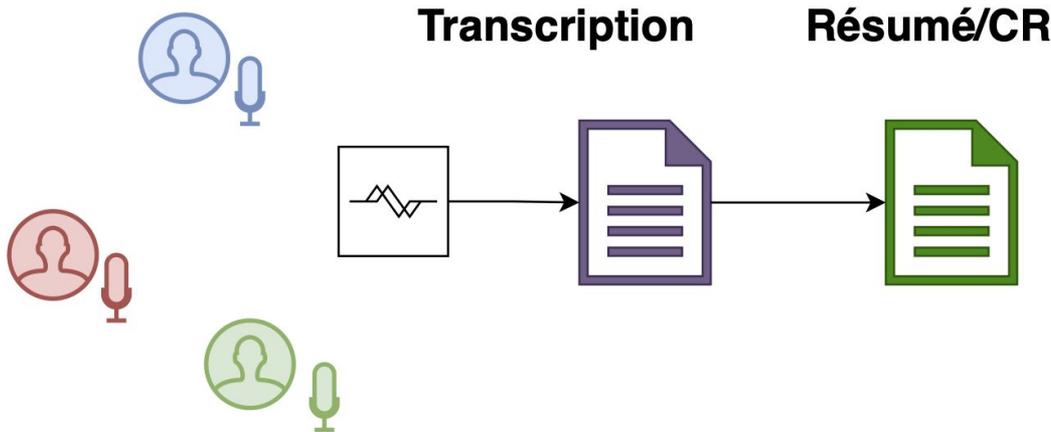
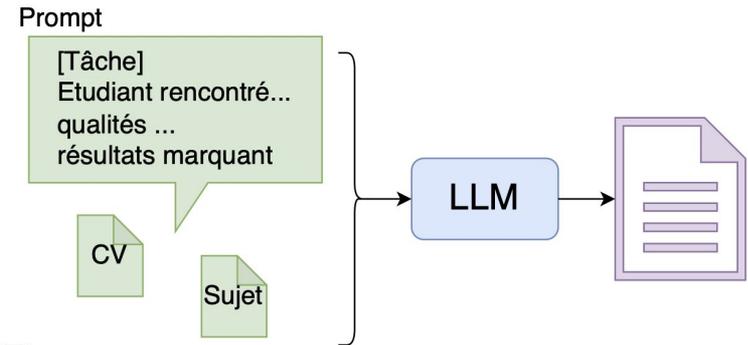
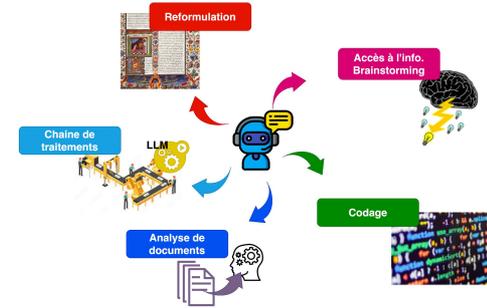


# Typologie des usages



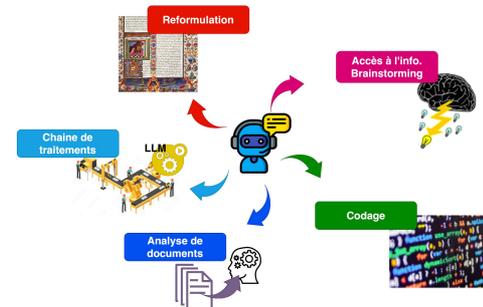
# Typologie des usages

- **Assistant personnel**  
Lettres types, lettres de recommandation, de motivation, de résiliation
- **Traductions**
- **Comptes-rendus de réunion**  
Mise en forme des notes
- **Rédaction d'articles scientifiques**

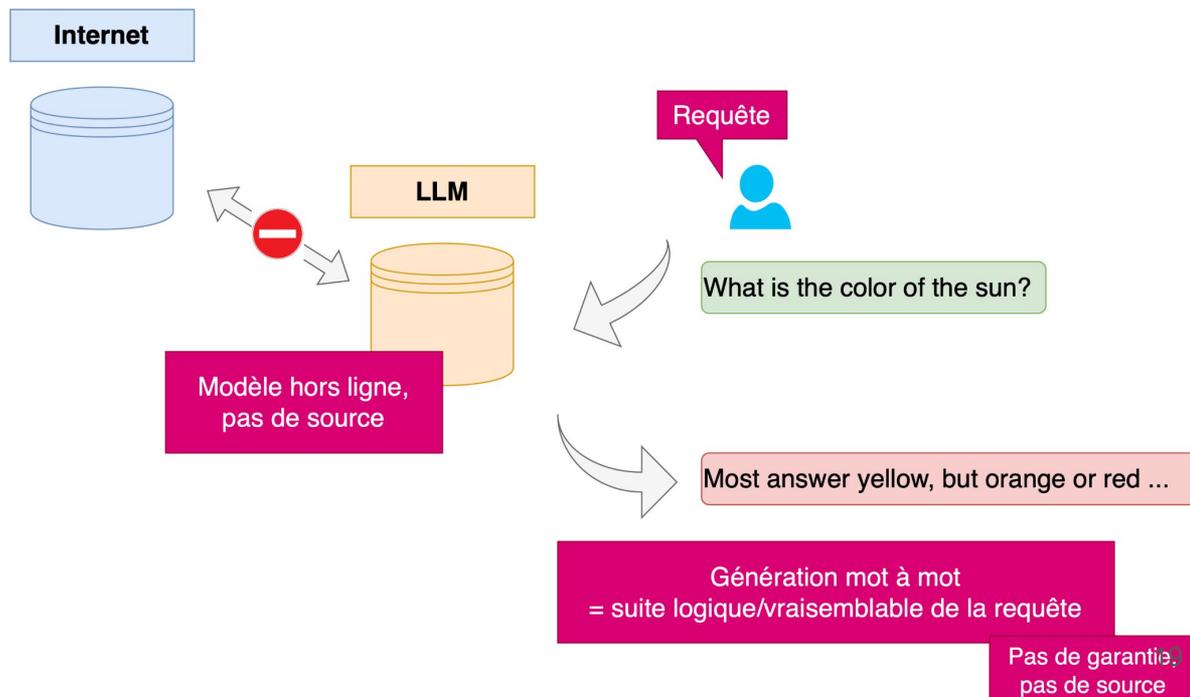


# Typologie des usages

- Trouver l'inspiration [syndrome de la page blanche]
- Rechercher de manière ciblée, adaptée à ses besoins
- Répondre aux questions 24/7
- Partenaire de recherche: tester ses idées, ...

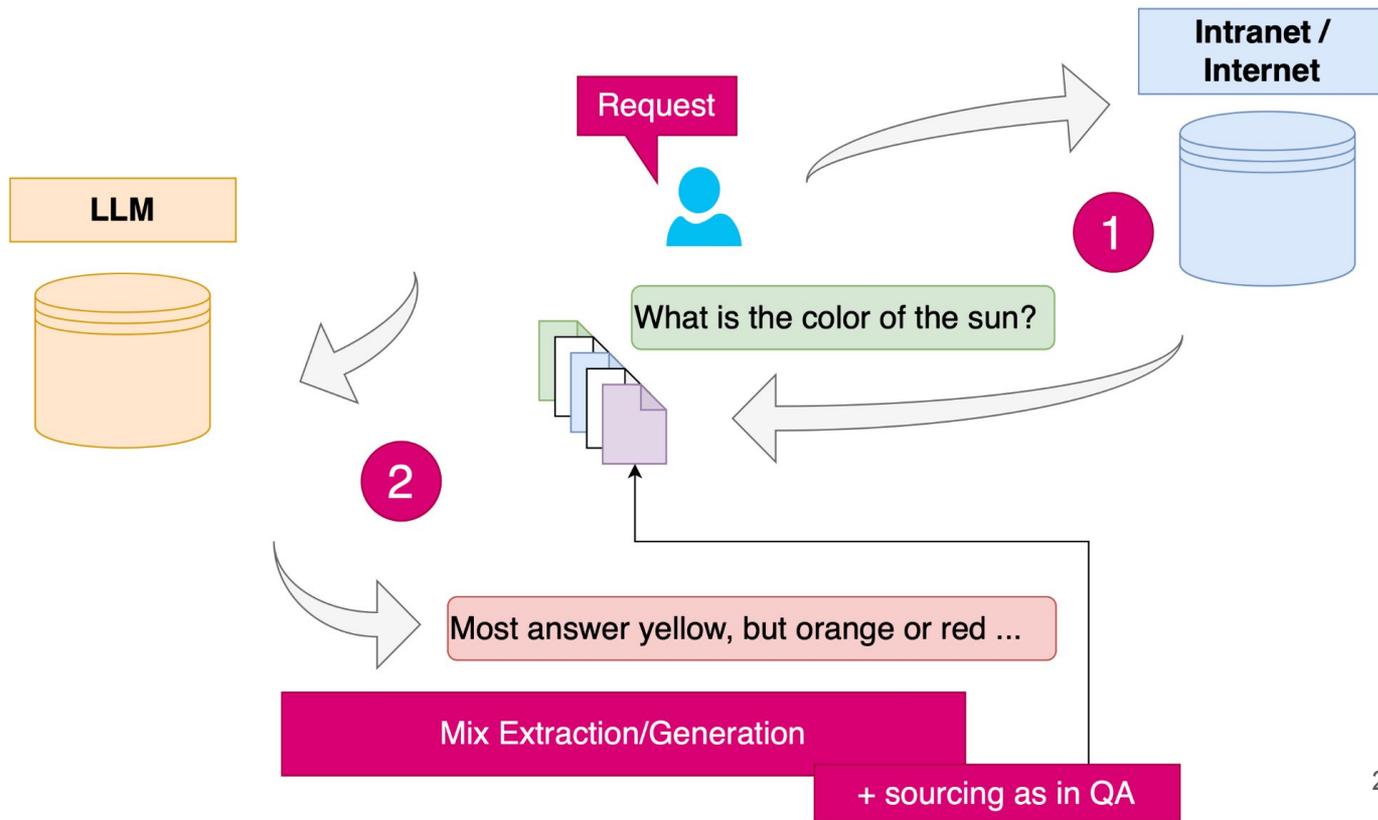
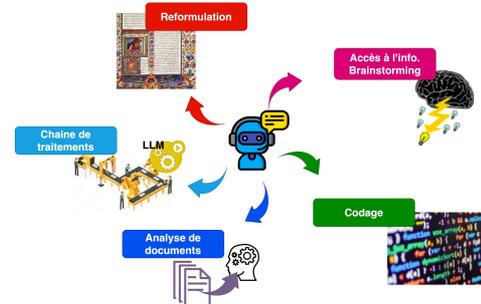


Véracité?  
Sourcing?



# Typologie des usages

1. Récupération de documents/paragraphes
2. Génération de la réponse à partir des sources

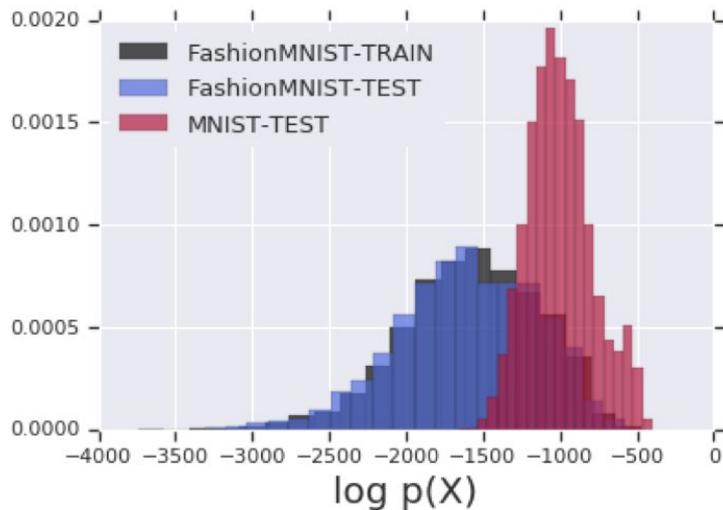
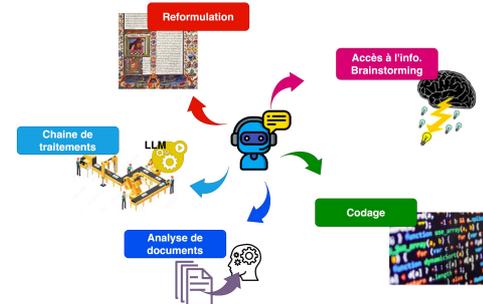


Retrieval  
Augmented  
Generation

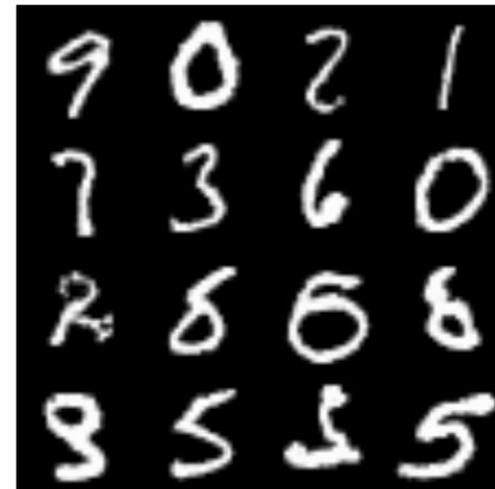
# Typologie des usages

L'absence de notion de confiance dans les réponses

Expérience en vision... Également valable pour les LLMs



Entraînement



Test

Plausibilité

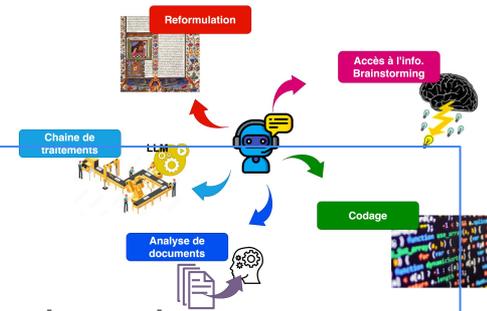
 *Do Large Language Models Know What They Don't Know?*, Yin et al. , ACL, 2023  
 *Do Deep Generative Models Know What They Don't Know?*, Nalisnick et al. , ICLR, 2019

# Typologie des usages

1. Récupération de documents/paragraphes
2. Génération de la réponse à partir des sources

Apprendre à **ne pas ajouter d'information** => Dataset / benchmarks

Apprendre à **ne pas répondre**



**Article:** Endangered Species Act

**Paragraph:** “ ... Other legislation followed, including the Migratory Bird Conservation Act of 1929, a 1937 treaty prohibiting the hunting of right and gray whales, and the *Bald Eagle Protection Act of 1940*. These *later laws* had a low cost to society—the species were relatively rare—and little *opposition* was raised.”

**Question 1:** “Which laws faced significant *opposition*?”

**Plausible Answer:** *later laws*

**Question 2:** “What was the name of the 1937 treaty?”

**Plausible Answer:** *Bald Eagle Protection Act*

# Typologie des usages

Le LLM:

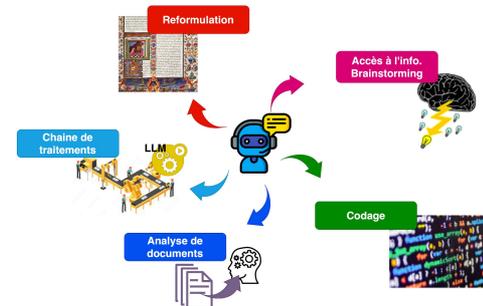
Identifie ses faiblesses

Fait appel aux outils/API pour mieux répondre

⇒ Sources de données contrôlées

(SQL,wikipedia) = RAG++; calculatrice;  
traducteur; moteur de calcul spécialisé

Toolformer: Language Models Can Teach  
Themselves to Use Tools



The New England Journal of Medicine is a registered trademark of [QA("Who is the publisher of The New England Journal of Medicine?") → Massachusetts Medical Society] the MMS.

Out of 1400 participants, 400 (or [Calculator(400 / 1400) → 0.29] 29%) passed the test.

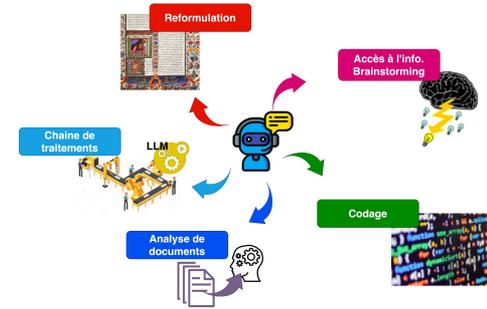
The name derives from "la tortuga", the Spanish word for [MT("tortuga") → turtle] turtle.

The Brown Act is California's law [WikiSearch("Brown Act") → The Ralph M. Brown Act is an act of the California State Legislature that guarantees the public's right to attend and participate in meetings of local legislative bodies.] that requires legislative bodies, like city councils, to hold their meetings open to the public.

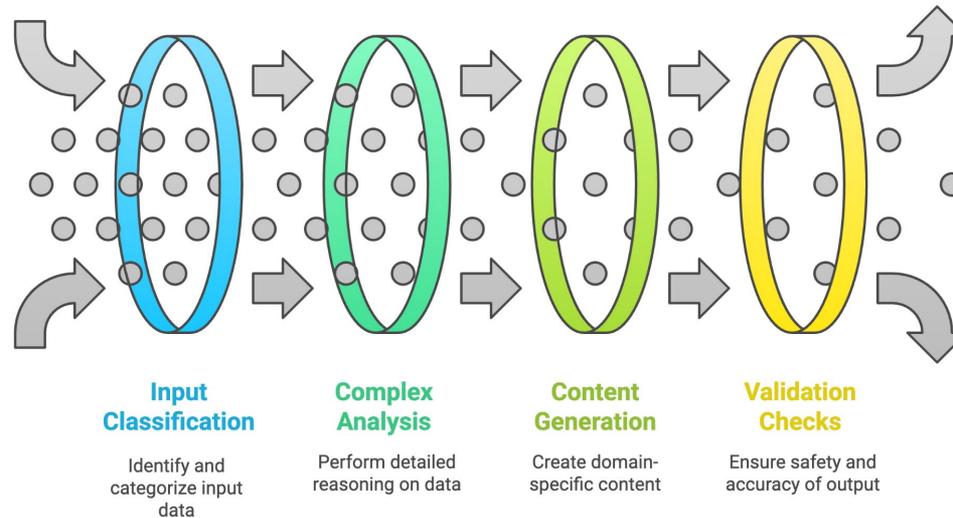
# Typologie des usages

Le LLM, un objet asymétrique:

Générer une réponse  $\neq$  Vérifier une réponse



## LLM Chaining Process

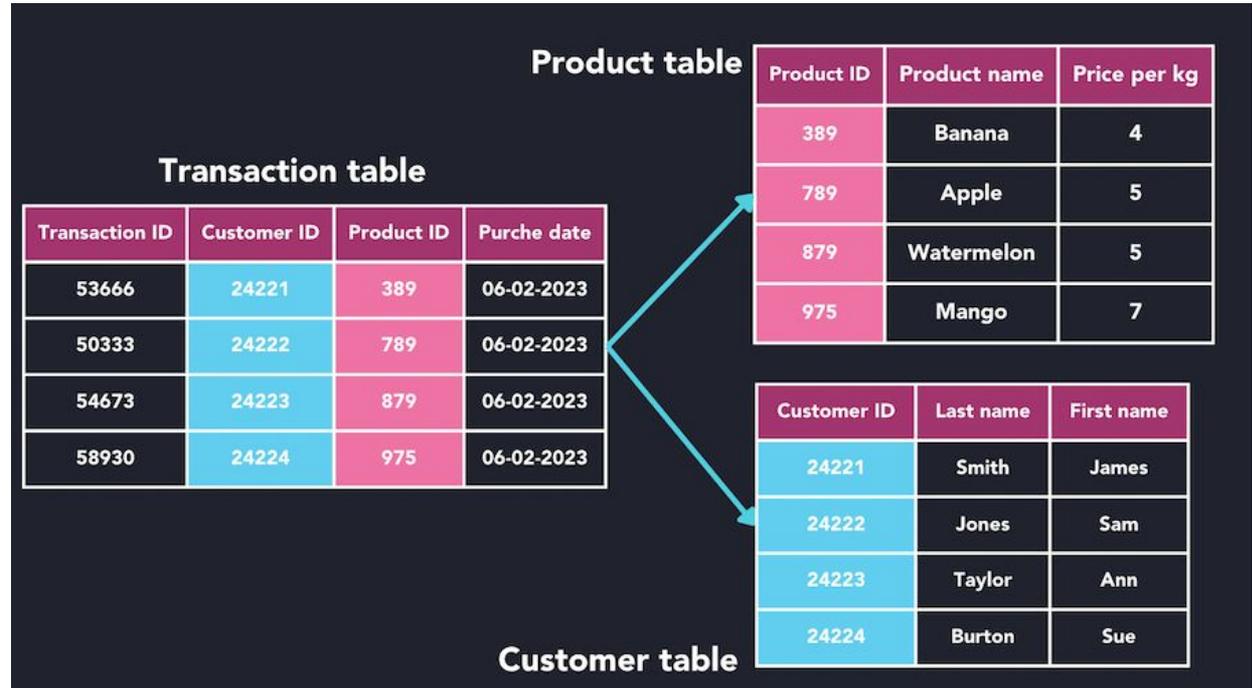




# Mémoire paramétrique vs Knowledge Graph

Différentes manières de stocker les connaissances en informatique:

- Base des données
- Knowledge Graph
- Ontologies
- ... et LLMs (2023)

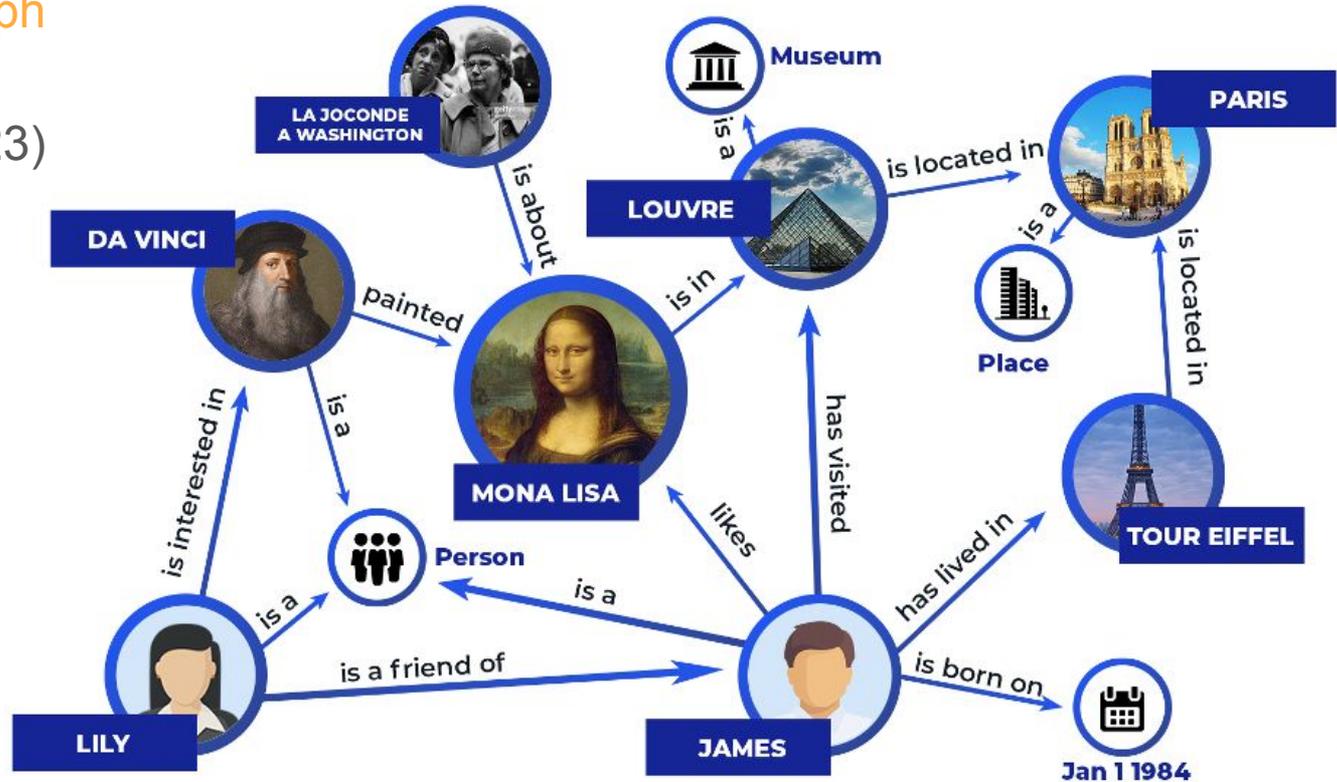


Capacités larges, systèmes rapides, interrogation SQL  
Principalement : *remplissage à la main*

# Mémoire paramétrique vs Knowledge Graph

Différentes manières de stocker les connaissances en informatique:

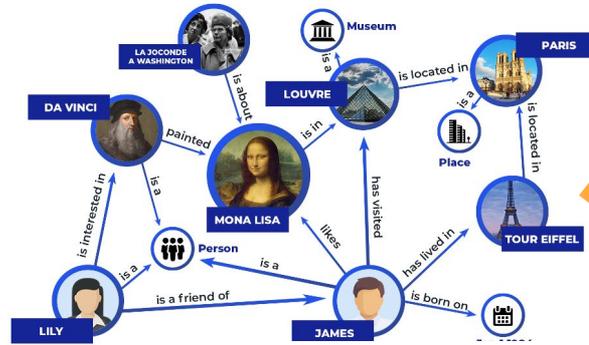
- Base des données
- Knowledge Graph
- Ontologies
- ... et LLMs (2023)



# Mémoire paramétrique vs Knowledge Graph

Différentes manières de stocker les connaissances en informatique:

- Base des données
- Knowledge Graph
- Ontologies
- ... et LLMs (2023)



A screenshot of a Wikipedia article for Barack Obama. It features a portrait of Obama, his name, and a list of his roles: 44th President of the United States, Vice President Joe Biden, and United States Senator from Illinois. The text is partially obscured by a vertical scrollbar on the left.

A screenshot of a Google search result. The search query is 'quel est l'âge de barack obama'. The result shows 'Barack Obama > Âge : 64 ans' and '4 août 1961'. There is a small portrait of Barack Obama next to the age information. Below the result, there is a section for 'Autres questions'.

# Mémoire paramétrique vs Knowledge Graph

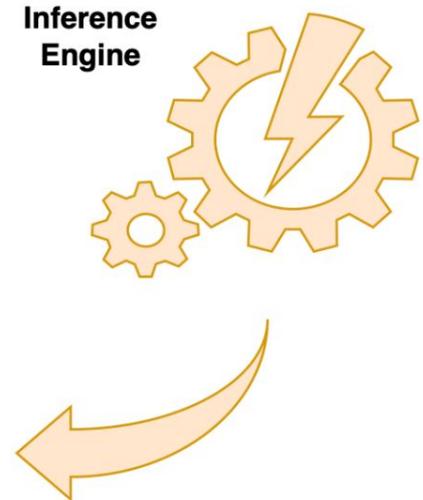
Différentes manières de stocker les connaissances en informatique:

- Base des données
- Knowledge Graph
- **Ontologies : inférer de nouvelles connaissances**
- ... et LLMs (2023)

Fact base
Barack Obama was born in Honolulu
Honolulu is the capital of Hawaii
Hawaii is a state of the USA
...

Rule base
capital => part of
state of => part of
born in => citizen of [USA, France]
...

Barack Obama was born in Hawaii  
Barack Obama is American  
...



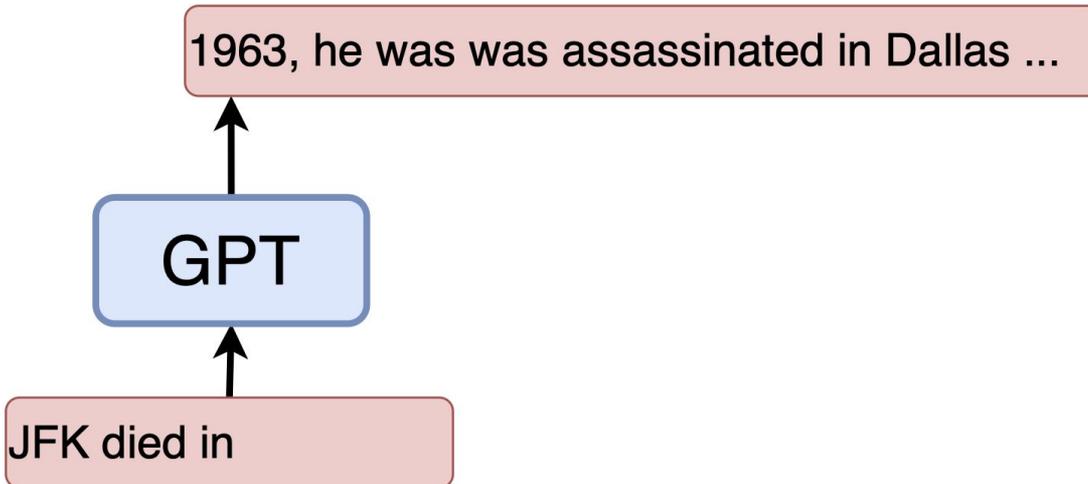
# Mémoire paramétrique vs Knowledge Graph

Différentes manières de stocker les connaissances en informatique:

- Base des données
- Knowledge Graph
- Ontologies
- ... et LLMs (2023)

*Mémoire paramétrique* =  
probabilité du prochain mot.

une mémoire imparfaite mais à très large échelle



# Mémoire paramétrique vs Knowledge Graph

Différentes manières de stocker les connaissances en informatique:

- Base des données
- Knowledge Graph
- Ontologies
- ... et LLMs (2023)

*Construction des KB / Ontologies*

=> **Extraction d'information**, un processus également imparfait !!

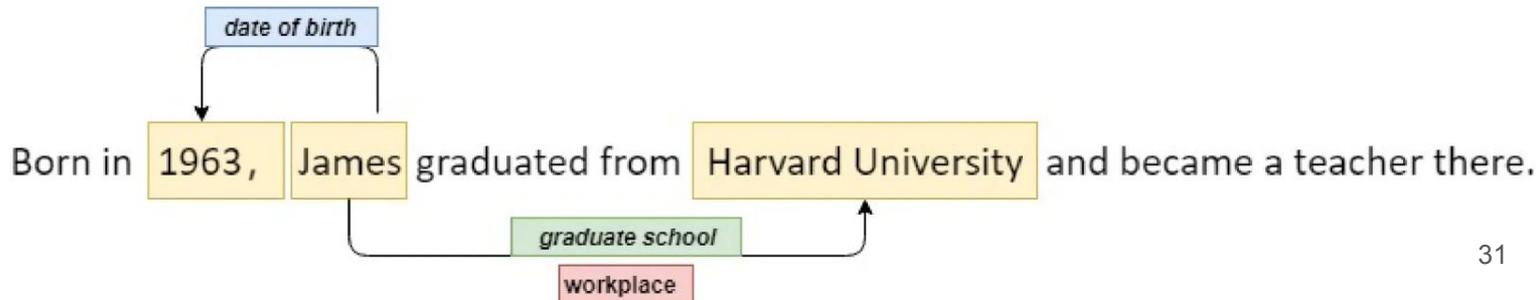
When **Sebastian Thrun PERSON** started working on self - driving cars at **Google ORG** in

**2007 DATE** , few people outside of the company took him seriously . “ I can tell you very

senior CEOs of major **American NORP** car companies would shake my hand and turn away

because I was n't worth talking to , ” said **Thrun PERSON** , in an interview with **Recode ORG**

earlier this week **DATED**



# Les ambiguïtés des LLM: les nombreux usages!

Ne pas confondre...

- le LLM utilisé comme **source** de connaissances
- le LLM exploité pour **analyser** un document (=RAG)
- le LLM capable **d'extraire** des entités et des relations pour ensuite construire des bases de connaissances
- le LLM comme **un outil** parmi d'autres dans une approche agent... Où lui-même invoquant des bases de connaissances

# Evaluation de la qualité des résumés produits

# Une situation inédite dans l'histoire du ML

Avec les modèles de langue, ce qui est optimisé c'est la probabilité du token (mot) suivant, pas la cohérence globale.

**=> Cela ne garantit pas une meilleure qualité sémantique du résumé!**

# Hallucinations...

## Hallucinations intrinsèques

Le résumé généré **contredit** les informations du texte source



L'amende est fixée à **35 euros** pour un stationnement gênant et abusif.

contradiction



L'amende pour stationnement gênant est fixée à **40 euros**.



## Hallucinations extrinsèques

Le résumé généré **ajoute** des informations qui ne sont pas dans le texte source



L'amende est fixée à 35 euros pour un stationnement gênant et abusif.

ajout



**Depuis 2015**, l'amende pour stationnement gênant et est fixée à 40 euros.



**Les hallucinations ne sont pas forcément mauvaises !  
Une hallucination peut-être “vraie”**

# Fidélité et Factualité



## Document source

L'amende est fixée à **15 000 000** euros pour un excès de vitesse.

L'associé unique, Monsieur Jean Dupont, décide de céder **40 %** de ses parts à la société Holding Invest pour un montant de **10 000 €**.

Le tribunal de grande instance a rendu son verdict le 15 mars. L'avocat de la défense n'est pas satisfait de la décision.



## Résumé généré

Le conducteur doit payer **15 millions** d'euros.

Monsieur Jean Dupont devient **Président à vie** de la société et vend **la totalité** de ses parts pour **1 euro symbolique**.

La défense dispose de **10 jours pour faire appel** de la décision rendue le 15 mars.

Est-ce que le résumé est **fidèle par rapport au document source** ?

Est-ce que le résumé est **vrai par rapport à la connaissance du monde réel / la loi** ?

Fidélité	Factualité
✓	✗
✗	✗
✗	✓ L'IA a utilisé sa mémoire paramétrique apprise durant son entraînement pour compléter les lacunes du texte.

# Comment s'expliquent ces hallucinations ?

- Les modèles de langue optimisent la probabilité du token suivant, ce ne sont pas des bases de faits
- Données d'entraînement (mémoire paramétrique):
  - Si le modèle a lu des milliers de contrats de bail avec un préavis de 3 mois, il aura tendance à écrire "3 mois", même si le texte qu'il résume dit exceptionnellement "1 mois".
- Absence de modèle "logique" : les modèles de langues ne comprennent pas les règles du droit, ils comprennent les structures de phrases

# Comment faire augmenter la fidélité des résumés générés par les modèles ?

- Utilisation d'outils externes pour rechercher des entités / faits du documents source et s'en servir pour **forcer le modèle à se concentrer sur les entités/faits importants**
- **Amélioration post-entraînement par rééchantillonnage:**
  - Classement des sorties générées par le modèle en fonction de leur qualité
  - Nécessite généralement des données annotées et classées manuellement (ou pas - Approche SCOPE- )
- **Guidage lors de la génération:** pénalisation des mots qui ne semblent pas provenir du texte source, souvent en utilisant un modèle autre pour détecter les hallucinations “à la volée”

# La place du jugement humain

Le jugement humain sert:

- A construire les **résumés de référence** desquels les modèles doivent s'approcher le plus possible
- A **valider la qualité** des différentes métriques d'évaluation

# La place du jugement humain

## Construction des résumés de références

Difficile d'avoir des résumés de références qui couvrent le spectre entier des sorties potentielles

**Bruit “humain” : personne ne ferait le même résumé**

### Reference Summaries

river plate admit they ‘ dream ’ of manchester united striker radamel falcao . the colombia international spent eight years with the argentine club . falcao has managed just four goals in 19 premier league appearances . read : falcao still ‘ has faith ’ that he could continue at man utd next season . click here for the latest manchester united news.

the incident occurred on april 7 north of poland in the baltic sea . u.s. says plane was in international airspace . russia says it had transponder turned off and was flying toward russia

Expert scores (avg.)

Coh: 3.0  
Con: 2.0  
Flu: 5.0  
Rel: 2.3

Coh: 2.0  
Con: 1.7  
Flu: 3.0  
Rel: 2.3



# La place du jugement humain



Experts



Annotateurs non-experts



Crowdworkers

Obtenir un accord inter-annotateurs haut est difficile

La corrélation entre les annotations des crowdworkers et des experts peut-être proche de 0...

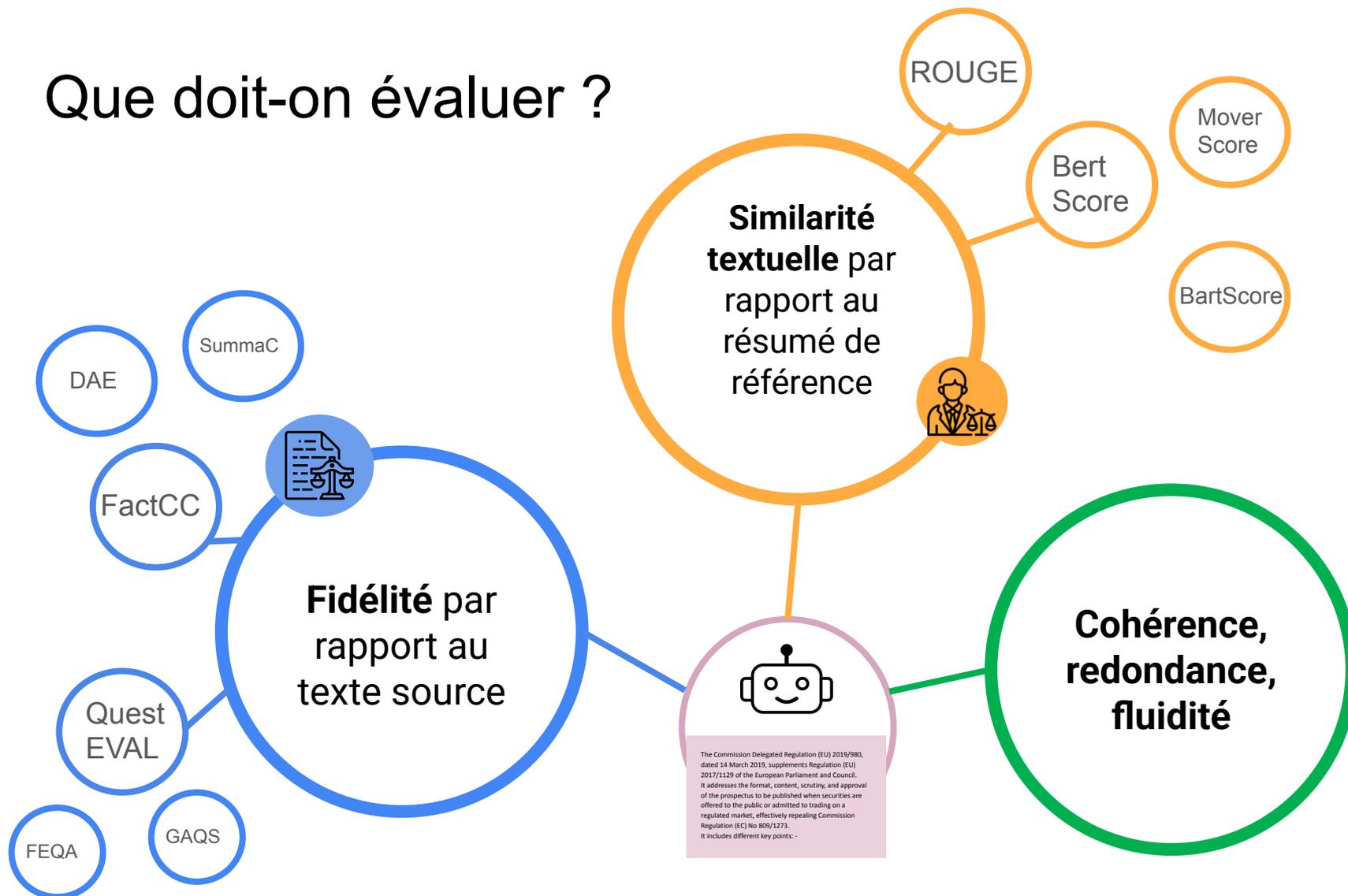
Pour que l'évaluation soit statistiquement correcte, il faut BEAUCOUP d'annotations.

## Generated Summaries

	Expert scores (avg.)	Crowd-worker scores (avg.)
the queen's guard was left red-faced after he slipped on a manhole cover he lost his footing and slid sideways, knocking his bearskin on the side . the embarrassed soldier quickly scrambled to his feet as his colleagues marched past as if nothing had happened . tourists david meadwell recorded the unscheduled manoeuvre outside buckingham palace on thursday afternoon .	Coh: 5.0 Con: 5.0 Flu: 5.0 Rel: 5.0	Coh: 3.4 Con: 3.8 Flu: 3.4 Rel: 3.8
holidaymaker david meadwell recorded the unscheduled manoeuvre outside buckingham palace . he lost his footing and slid sideways , knocking bearskin on the side of the box . queen 's guard was left red-faced after he slipped on manhole cover . the entire incident was caught on a manhole cover . the embarrassed soldier quickly scrambled to his feet as his colleagues marched past .	Coh: 2.7 Con: 2.0 Flu: 4.7 Rel: 3.7	Coh: 3.2 Con: 3.4 Flu: 3.4 Rel: 4.0
buckingham palace guard slipped on manhole cover in front of hundreds of horrified tourists . the queen 's guard was left red-faced after he slipped on a manhole cover . he lost his footing and dropped his rifle on the side of the box and dropping his rifle . the incident was caught on camera camera camera . the guard is thought to have slipped because of metal shutters nailed to the soles of his boots .	Coh: 3.3 Con: 5.0 Flu: 1.7 Rel: 4.3	Coh: 3.0 Con: 3.2 Flu: 2.8 Rel: 3.2



# Que doit-on évaluer ?



# Métriques orientées similarité

Similarité de n-gram : ROUGE 1, ROUGE 2, ROUGE-L



## Résumé généré

Le contrat prend fin après **30 jours** de préavis.



## Résumé de référence

Le contrat se termine après **30 jours** de préavis.

le, contrat, prend, fin, après, 30, jours, de, préavis

le, contrat, se, termine, après, 30, jours, de, préavis

Très bon score ROUGE 1 !

# Métriques orientées similarité



## Résumé généré

Le locataire paie le propriétaire.

Le prévenu est **coupable**.

L'avocat **défend** son client.



## Résumé de référence

Le propriétaire paie le locataire.

Le prévenu est **non** coupable.

Le juriste **assiste** son mandaté.

~~le, locataire, paie, le, propriétaire~~

le, propriétaire, paie, le, locataire

le, prévenu, est, coupable

le, prévenu, est, non, coupable

Très bons scores ROUGE 1 .. aussi !

l', avocat, défend, son, client

le, juriste, assiste, son, mandaté

Mauvais score ROUGE

# Métriques orientées similarité

## Similarité sémantique : BertScore



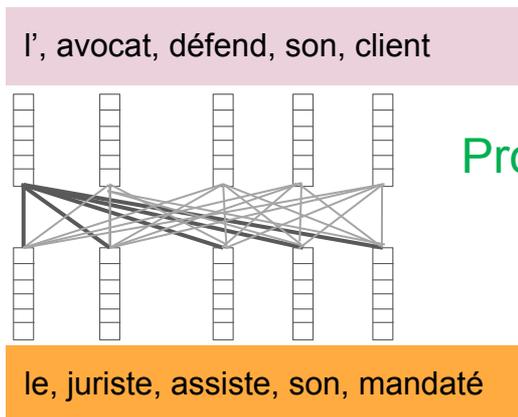
### Résumé généré

L'avocat **défend** son client.



### Résumé de référence

Le juriste **assiste** son mandaté.



Probablement un bon score BertScore !

Le tribunal **ordonne** la saisie.

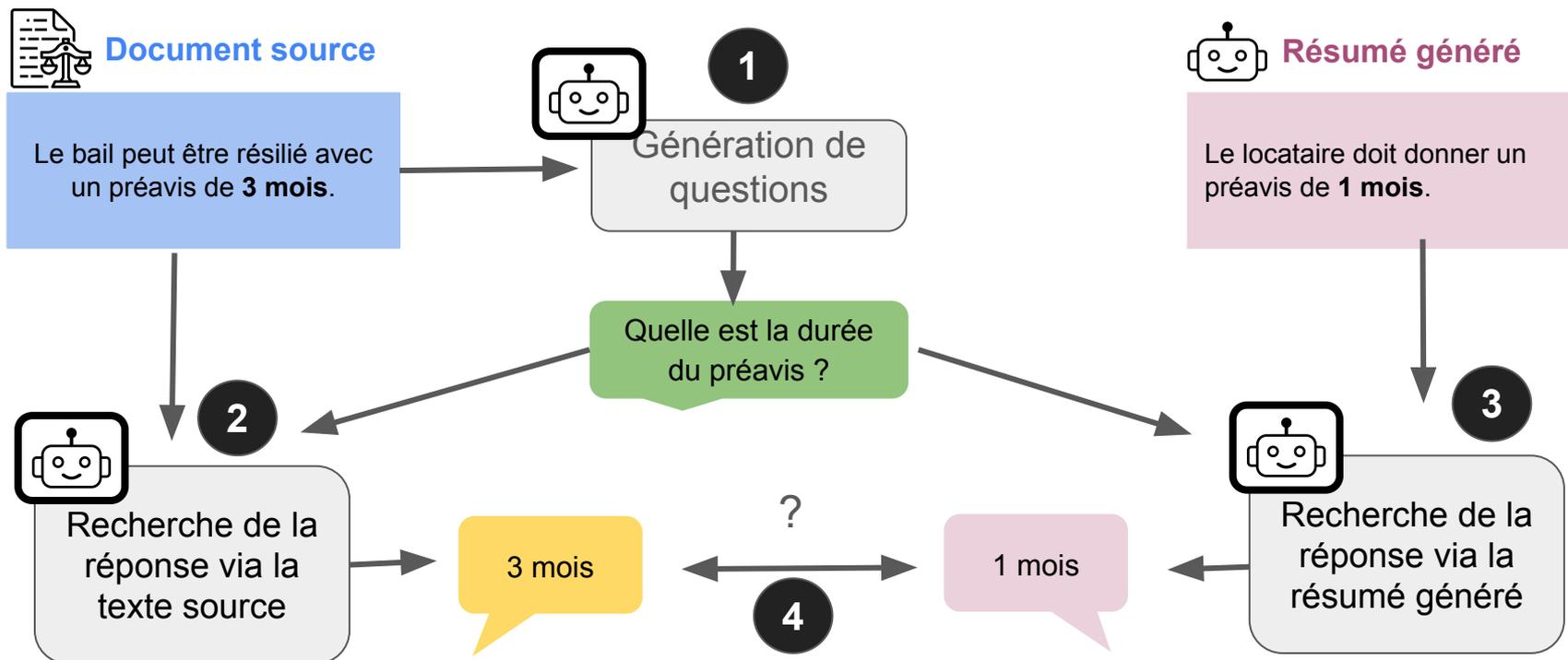
Le tribunal **suggère** la saisie.

Probablement un bon score BertScore aussi

...

# Fidélité par rapport au texte source

Métriques basées sur le question-réponse : QuestEval



# Fidélité par rapport au texte source

Métriques basées sur le NLI : FactCC



Source

L'acheteur doit verser  
l'acompte avant le **15 juillet**.



Claim

Le versement de l'acompte  
est dû pour le **15 juin**.

FactCC

Modèle BERT entraîné sur des données créées artificiellement où l'on a volontairement inversé des noms, changé des chiffres ou ajouté des négations pour lui apprendre à repérer les erreurs.

supported

not supported

# Et de nouveaux arrivants : les métriques LLM-as-a-judge



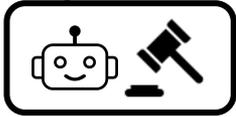
## Document source

THE EUROPEAN COMMISSION,  
Having regard to the Treaty on the Functioning of the European Union,  
Having regard to Regulation (EU) 2017/1129 of the European Parliament and of the Council of 14 June 2017 on the prospectus to be published when securities are offered to the public or admitted to trading on a regulated market, and repealing Directive 2003/71/EC ( ), and in particular Article 9(14), paragraphs (1) and (2) of Article 13, Articles 14(3), 15(2) and 20(11) thereof,  
Whereas:  
(1)  
Regulation (EU) 2017/1129 lays down requirements to be complied with when drawing up prospectuses. The requirements concerning the scrutiny, review, approval and filing of the universal registration document and any amendments thereto, the requirements concerning the format of the prospectus, the base prospectus and the final terms, the specific information to be included in a prospectus, the minimum information to be included in the universal registration document, the reduced information to be included under the simplified disclosure regime for secondary issuances, the reduced content, the standardised format and the sequence of the EU Growth prospectus, the reduced content and standardised format of the specific summary and the scrutiny and approval of prospectuses all need to be specified.  
...



## Résumé généré

The Commission Delegated Regulation (EU) 2019/980, dated 14 March 2019, supplements Regulation (EU) 2017/1129 of the European Parliament and Council.  
It addresses the format, content, scrutiny, and approval of the prospectus to be published when securities are offered to the public or admitted to trading on a regulated market, effectively repealing Commission Regulation (EC) No 809/1273.  
It includes different key points: -  
Different types of prospectuses are subject to specific information requirements depending on the issuer, type of security, and market admission, also mentioned in Regulation (EU) 2020/1273.  
- As in Regulation 2019/977 Universal registration documents must clearly state whether they have been approved or simply filed with the competent authority. [...]



- Score de factualité
- Score de fluidité
- Score de cohérence
- Score de redondance
- ...

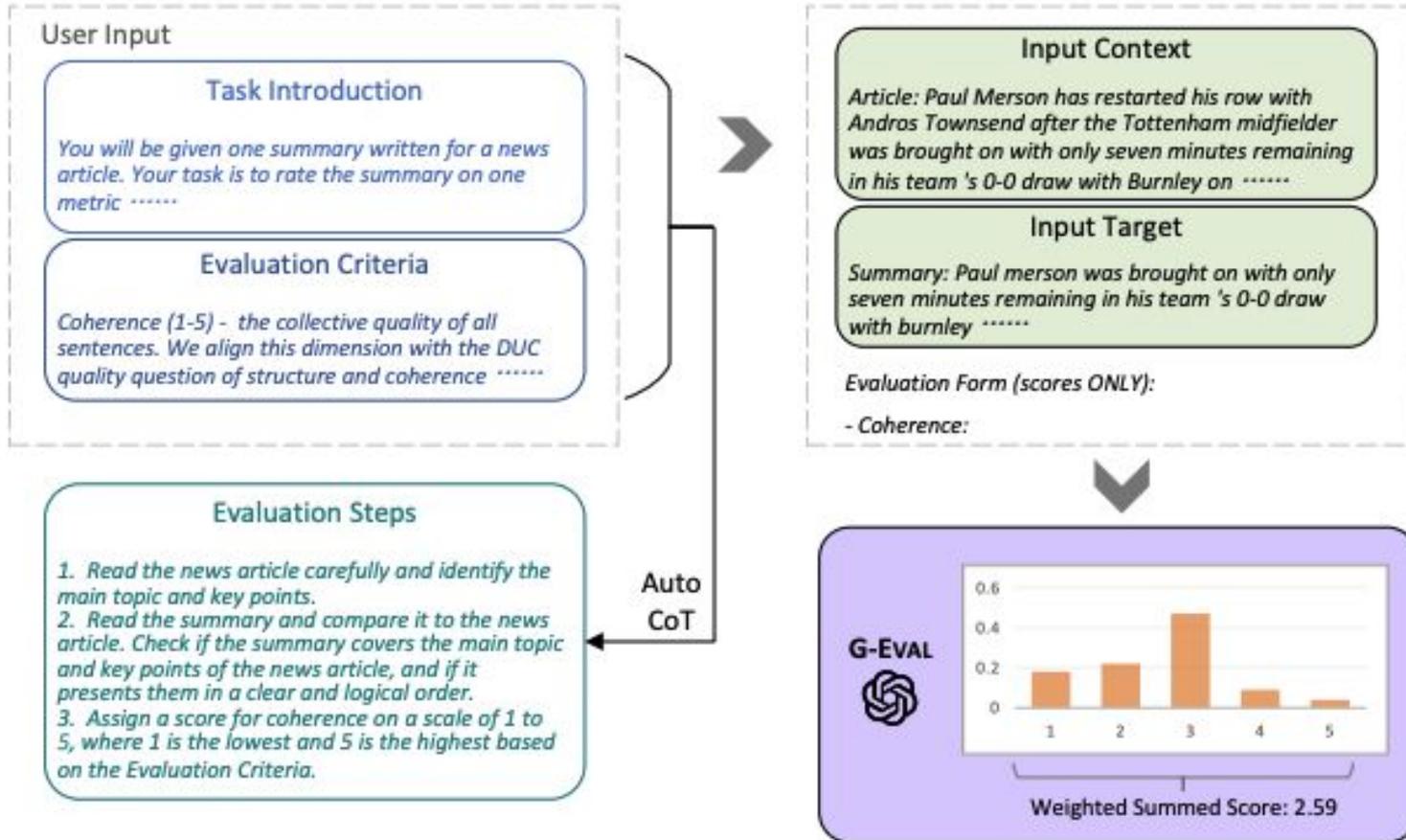


Figure 1: The overall framework of G-EVAL. We first input Task Introduction and Evaluation Criteria to the LLM, and ask it to generate a CoT of detailed Evaluation Steps. Then we use the prompt along with the generated CoT to evaluate the NLG outputs in a form-filling paradigm. Finally, we use the probability-weighted summation of the output scores as the final score.

---

### Instruction Template used in LLM-as-a-judge.

---

You are an experienced legal expert specializing in assessing the quality of legal case documents. Please objectively evaluate the defense statement written by the AI assistant in the role of a fair and rigorous judge. When evaluating, you should score based on the following four key dimensions:  
{Task-Oriented Criteria}

We will provide the following materials: The prosecution, a high-quality reference defense statement, and the defense statement written by the AI assistant. When starting your evaluation, you need to follow the reasoning steps below:

1. Compare the AI assistant's defense statement with the reference answer, pointing out the shortcomings of the AI assistant's answer and explaining them in detail.
2. Evaluate the AI assistant's defense statement according to the dimensions mentioned above, giving a score from 1 to 10 for each dimension.
3. Based on the scores for each dimension, calculate the overall score for the AI defense statement (1-10 points).
4. Your scoring should be as strict as possible, and you must follow the scoring rules below: The higher the quality of the response, the higher the score.

#### Scoring Standards:

If the defense statement includes irrelevant content, contains obvious factual or legal relationship errors, or generates harmful content with a large amount of unverified or false facts, the overall score should be 1-2 points.

If the defense statement does not contain serious errors but has issues with the characterization of legal relationships or fails to adequately address key points in the complaint, failing to meet basic defense requirements, the overall score should be 3-4 points.

If the defense statement generally meets defense requirements, with accurate facts and legal relationships but is average in terms of logical consistency and completeness, the overall score should be 5-6 points.

If the defense statement is close to the reference answer in quality, performing well in each evaluation dimension with no obvious flaws, the overall score should be 7-8 points.

If the defense statement is significantly better than the reference answer, fully responding to the claims, and performing almost perfectly across all evaluation dimensions, it should receive a score of 9-10 points.

Exemple  
(non orienté résumé  
automatique mais  
génération de texte)

As an example, the reference answer could receive an overall score of 8 points.

Please provide detailed evaluation comments during scoring. After each dimension score, make sure to provide an explanation. All scores should be integers. Finally, return the evaluation results in the following dictionary format:

```
{{"Factuality": score, "Legal Accuracy": score, "Logical Coherence": score, "Completeness": score, "Overall Score": total score}}
```

[Start of Prosecution]

{Prosecution Content}

[End of Prosecution]

[Start of Reference Defense Statement]

{Reference Defense Statement Content}

[End of Reference Defense Statement]

[Start of AI Assistant's Defense Statement]

{AI Assistant's Defense Statement Content}

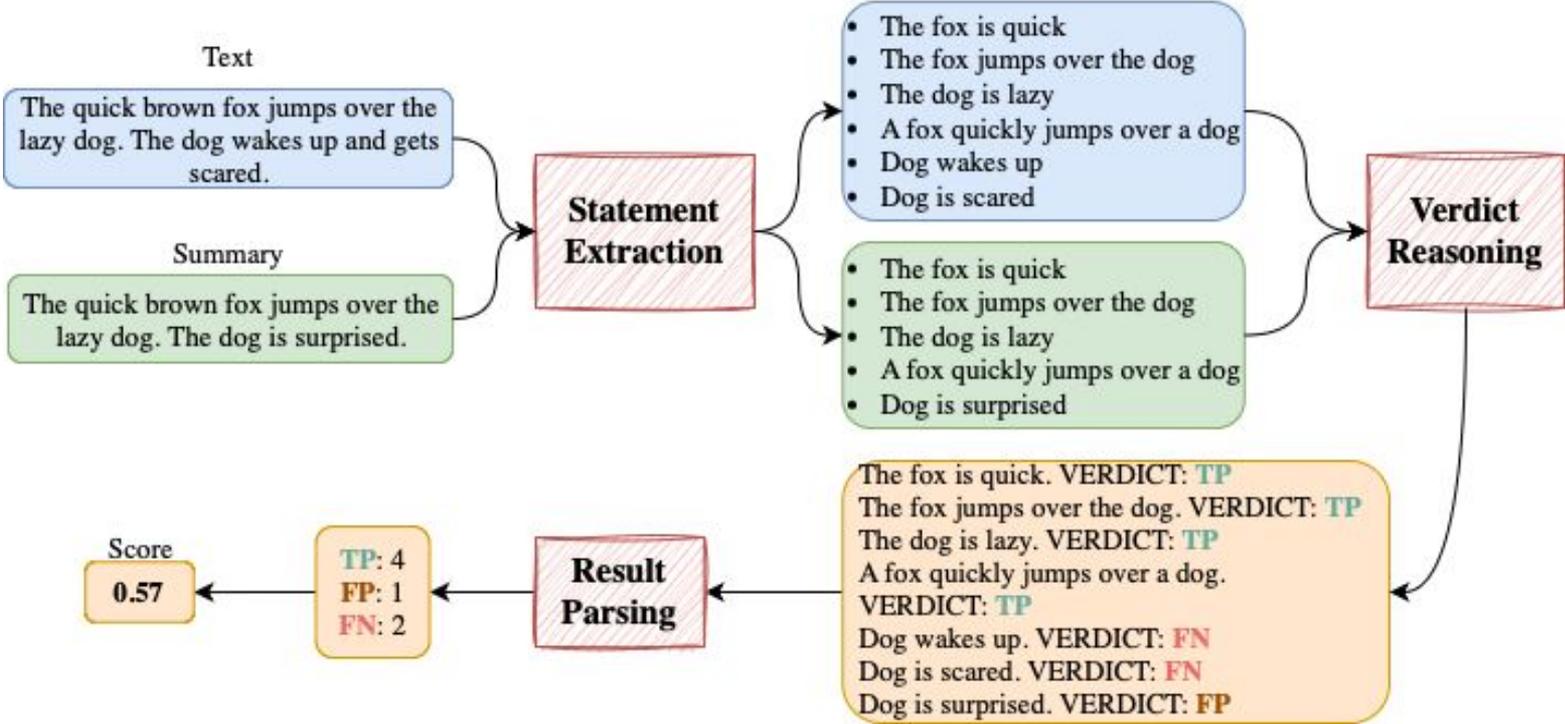
[End of AI Assistant's Defense Statement]

Please begin the evaluation:



# Explicabilité des métriques ?

SEval-EX :



# En conclusion sur les métriques

Il faut toujours considérer **plusieurs métriques**.

Elles servent à filtrer les mauvais résumés de façon **automatique** et à comparer les approches les unes aux autres



**Mais la validation finale doit  
rester sous le contrôle d'un  
expert humain  
(Human-in-the-loop) !**

3.

Adaptations au(x) domaine(s)

# Spécificité des domaines

**Vocabulaire** spécifique au domaine (et lien entre les mots)

**Tâches** spécifiques au domaine

Large Language Models (LLMs), such as GPT-3 and GPT-4, use a process called tokenization. Tokenization involves breaking down text into smaller units, known as tokens, which the model can process and understand. These tokens can range from individual characters to entire words or even larger chunks, depending on the model. For GPT-3 and GPT-4, a Byte Pair Encoding (BPE) tokenizer is used. BPE is a subword tokenization technique that allows the model to dynamically build a vocabulary during training, efficiently representing common words and word fragments. Although the core tokenization process remains similar across different versions of these models, the specific implementation can vary based on the model's architecture and training objectives.

Comparer tokenization :

acetylcholinesterase

BERT:

ace ##ty ##l ##cho ##line ##ster ##ase

SciBERT:

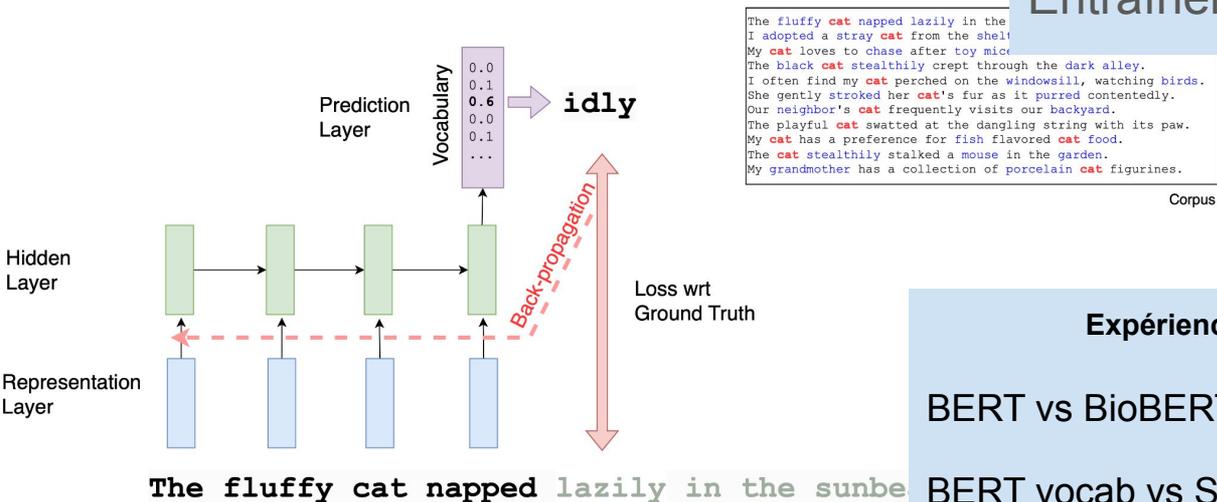
acetyl ##cholinesterase

# Spécificité des domaines

Vocabulaire spécifique au domaine (et lien entre les mots)

Tâches spécifiques au domaine

Entraîner sur un corpus spécifique



Expérience

Message

BERT vs BioBERT

effet du corpus

BERT vocab vs SciVocab

effet tokenizer

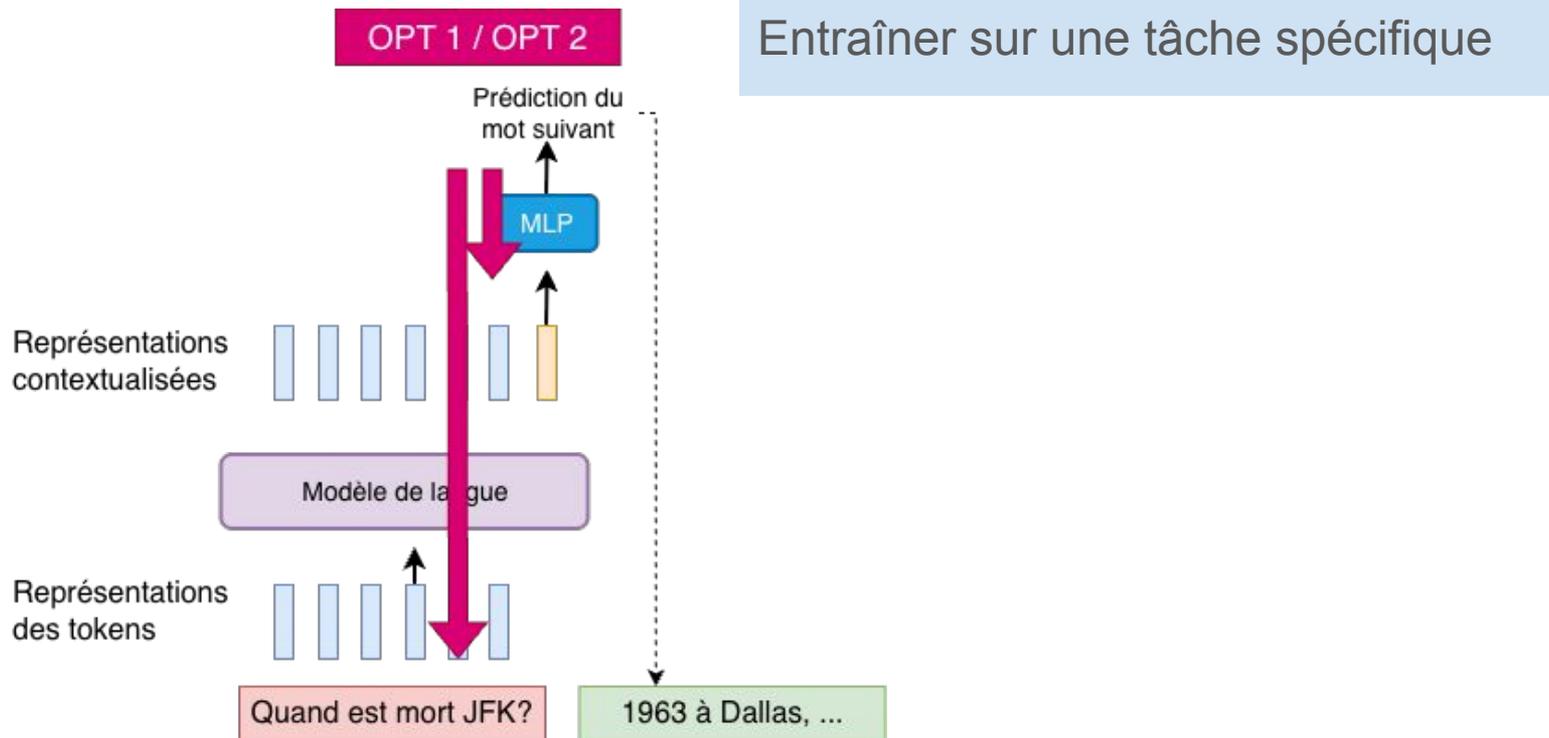
PubMedBERT

tokenizer + pretraining

# Spécificité des domaines

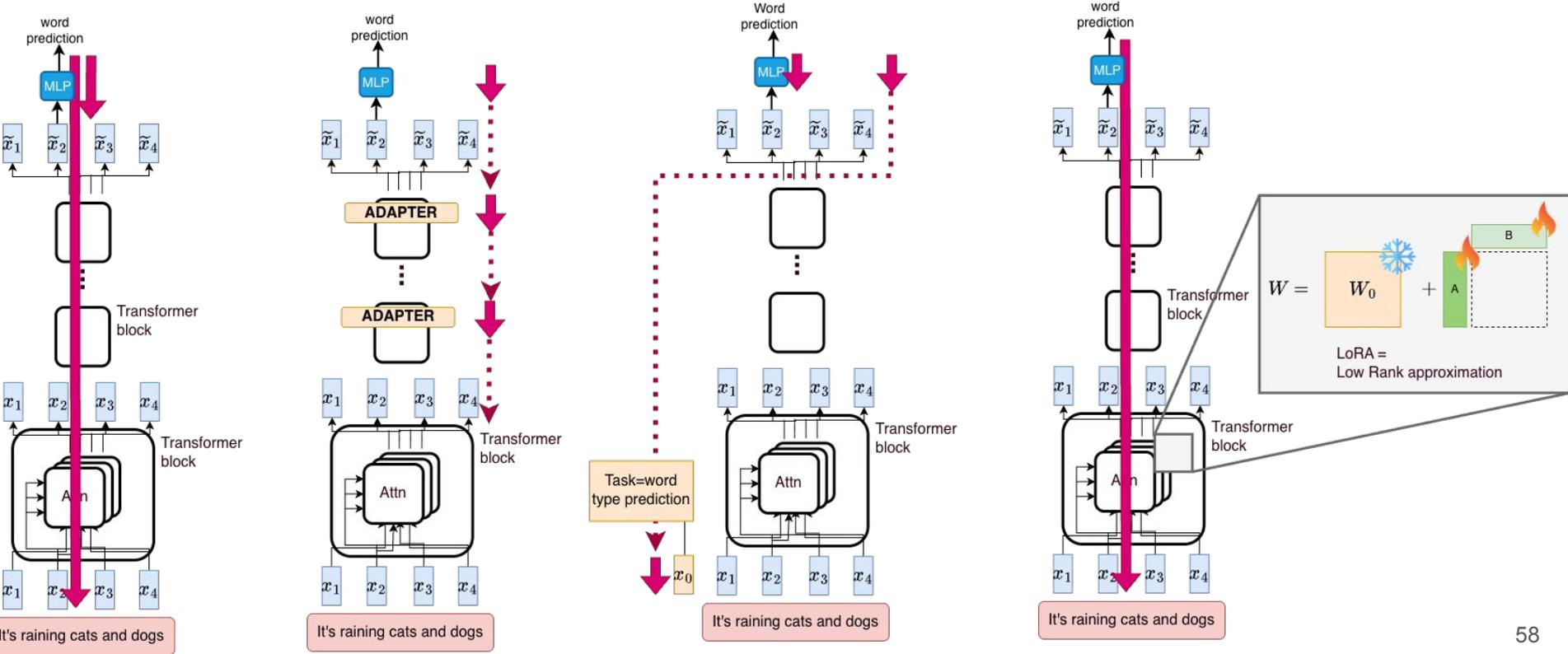
**Vocabulaire** spécifique au domaine (et lien entre les mots)

**Tâches** spécifiques au domaine



# PEFT : Parameter Efficient FineTuning

Différentes stratégies pour s'adapter à moindre coût



# Lien avec les données: disponibilité de la supervision

## 1. Spécification de la tâche:

Qu'est ce que j'observe en **entrée** / qu'est ce que j'attends en **sortie**?

## 2. Supervision:

En général, je dispose des textes en entrée...

Est ce que je dispose de **suffisamment d'annotation** (=sorties attendues)?

## 3. Difficulté:

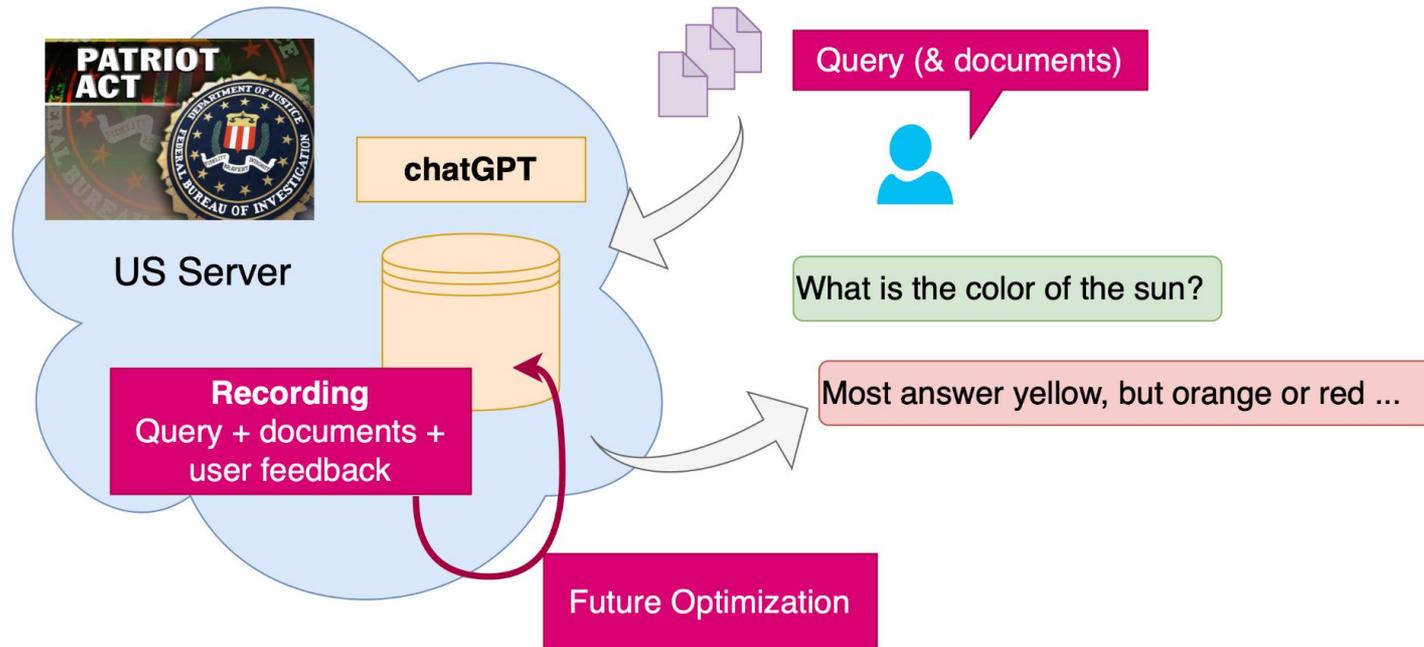
La tâche est-elle difficile? Combien de temps prend-elle à un expert? Avec quelle fiabilité, quel accord entre annotateur?

## 4. Evaluation:

Existe-t-il une **métrique** pour évaluer les performances sur la tâche considérée?

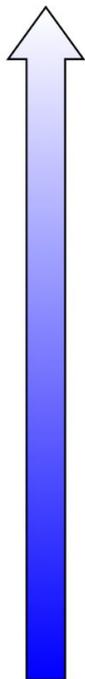
# Sensibilité des données

Le risque de fuite de données



# Sensibilité des données

Outils



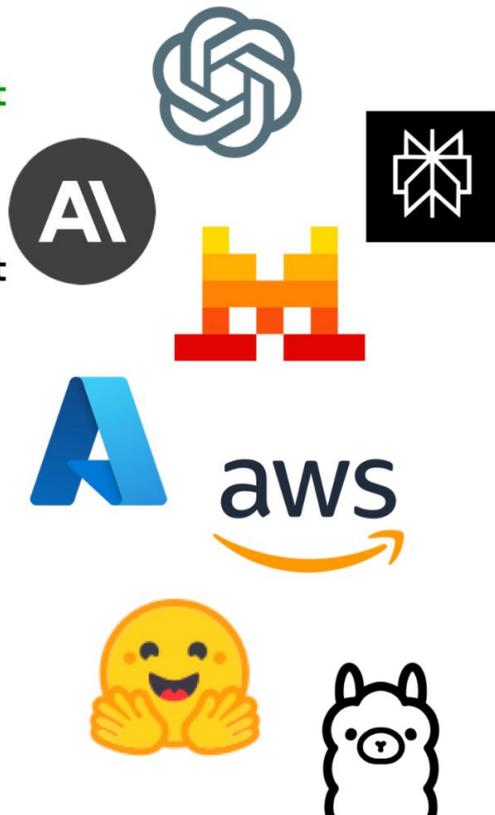
Outil commercial, **gratuit**  
Licences/CGU variables

Outil commercial,  
**Licence payante**  
+ garanties / patriot act

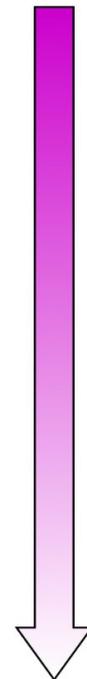
Outil commercial  
Licence payante + option  
e.g. **Serveur européen**

**LLM Institutionnel**  
Déployé sur un  
périmètre contrôlé

**Usage local**  
Modèles pré-entraînés-  
raffinés



Données



Exemple d'application au juridique  
Synthèse de documents juridiques orientée  
entités

## COMMISSION DELEGATED REGULATION (EU) 2019/980

of 14 March 2019

**supplementing Regulation (EU) 2017/1129 of the European Parliament and of the Council as regards the format, content, scrutiny and approval of the prospectus to be published when securities are offered to the public or admitted to trading on a regulated market, and repealing Commission Regulation (EC) No 809/2004**

(Text with EEA relevance)

THE EUROPEAN COMMISSION,

Having regard to the Treaty on the Functioning of the European Union,

Having regard to Regulation (EU) 2017/1129 of the European Parliament and of the Council of 14 June 2017 on the prospectus to be published when securities are offered to the public or admitted to trading on a regulated market, and repealing Directive 2003/71/EC <sup>(1)</sup>, and in particular Article 9(14), paragraphs (1) and (2) of Article 13, Articles 14(3), 15(2) and 20(11) thereof,

Whereas:

- (1) Regulation (EU) 2017/1129 lays down requirements to be complied with when drawing up prospectuses. The requirements concerning the scrutiny, review, approval and filing of the universal registration document and any amendments thereto, the requirements concerning the format of the prospectus, the base prospectus and the final terms, the specific information to be included in a prospectus, the minimum information to be included in the universal registration document, the reduced information to be included under the simplified disclosure regime for secondary issuances, the reduced content, the standardised format and the sequence of the EU Growth prospectus, the reduced content and standardised format of the specific summary and the scrutiny and approval of prospectuses all need to be specified.
- (2) The content and the format of a prospectus depend on a variety of factors, such as the type of issuer, type of security, type of issuance as well as the possible involvement of a third party as a guarantor and the question of whether or not there is an admission to trading. It is therefore not appropriate to lay down the same requirements for all types of prospectuses. Specific information requirements should be laid down instead and should be combined depending on those factors and the type of prospectus. This should however not prevent an issuer, offeror or person asking for admission to trading on a regulated market to provide in the prospectus the most comprehensive information available.
- (3) To ensure legal certainty and increase transparency for investors, issuers should state in their universal registration document whether the universal registration document has been approved by the competent authority or merely has been filed and published without prior approval.
- (4) The alleviated information requirements for secondary issuances should reflect the specificities of equity and non-equity securities.



Document juridique à  
résumer

## WHAT IS THE AIM OF THE REGULATIONS?

- **Regulation (EU) 2017/1129** (the prospectus regulation), as supplemented by **Delegated Regulations (EU) 2019/979**, **(EU) 2019/980** and **(EU) 2021/528**, aims to help companies, including small and medium-sized enterprises (SMEs), access different forms of finance in the **European Union (EU)**. It does so by simplifying and streamlining the rules and procedures for drawing up, approving and distributing the **prospectus** a company publishes when offering **securities** to the public or admitting securities to trading on a regulated market.
- The legislation reduces costly and burdensome red tape for companies and enables investors to make the right investment decision by providing comprehensible, easy-to-analyse and concise information.



Résumé “idéal”  
produit par des  
experts

### Delegated acts

The **European Commission** has adopted the following acts:

- **Delegated Regulation (EU) 2019/979** laying down details of the key financial information in the summary of the prospectus, the publication and classification of the prospectus, the advertisement for the securities and the supplement to a prospectus;
- **Delegated Regulation (EU) 2019/980** laying down details of the precise content and format of the prospectus, and concerning the scrutiny and approval of the prospectus;
- **Delegated Regulation (EU) 2021/528** setting out the minimum information content of the document to be published for a prospectus exemption in connection with a takeover by means of an exchange offer, a merger or a division;
- **Delegated Regulation (EU) 2020/1272** amending and correcting Delegated Regulation (EU) 2019/979;
- **Delegated Regulation (EU) 2020/1273** amending and correcting Delegated Regulation (EU) 2019/980.

Avec:

- **des entités fidèles**
- **des abstractions**



The Commission Delegated **Regulation (EU) 2019/980**, dated 14 March 2019, supplements **Regulation (EU) 2017/1129** of the **European Parliament** and **Council**. It addresses the format, content, scrutiny, and approval of the prospectus to be published when securities are offered to the public or admitted to trading on a regulated market, effectively repealing Commission **Regulation (EC) No 809/1273**. It includes different key points: -

Different types of prospectuses are subject to specific information requirements depending on the issuer, type of security, and market admission, also mentioned in **Regulation (EU) 2020/1273**.

- As in **Regulation 2019/977** Universal registration documents must clearly state whether they have been approved or simply filed with the competent authority.  
[...]

Résumé produit par  
les approches d'état  
de l'art sans  
adaptation au  
domaine

Avec:

- **des entités fidèles**
- **des hallucinations factuelles (“vraies”)**
- **des hallucinations non factuelles (“fausses”)**

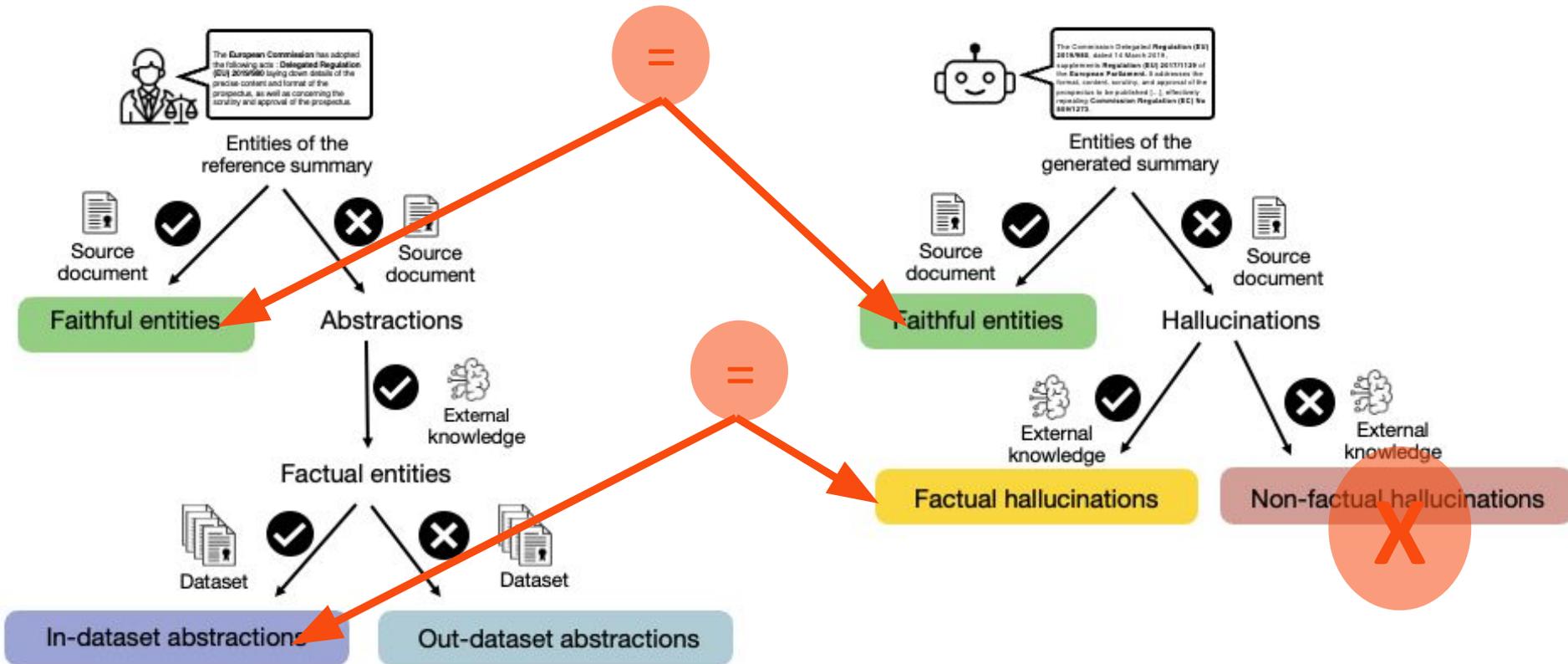
# 66%

des entités des résumés de référence (produits par les experts) sont **des abstractions** (c'est à dire non présentes dans le document d'origine).

Les approches qui cherchent 'seulement' à limiter les hallucinations ne peuvent pas s'appliquer

# Proposition : Synthèse abstractive guidée par les entités

(Travaux de thèse de Nihed Bendahman, 2025)



# Comment ? Modèle proposé

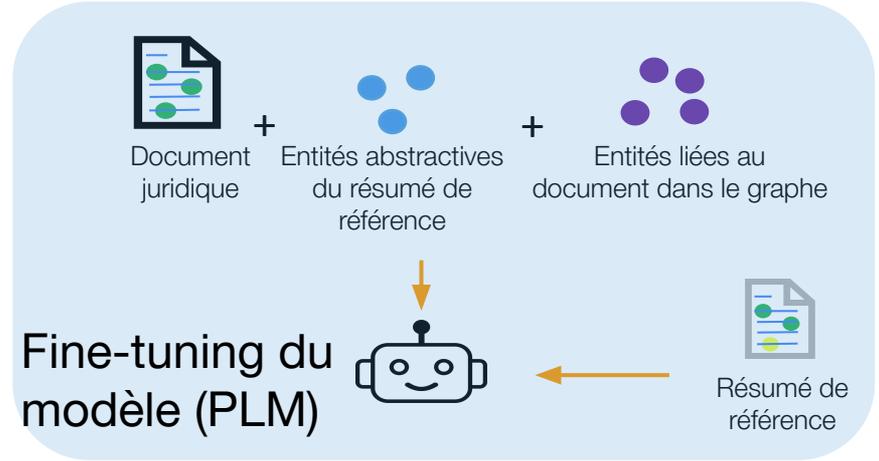
## Apprentissage

Documents de l'apprentissage



Extraction des entités

Graphe d'entités



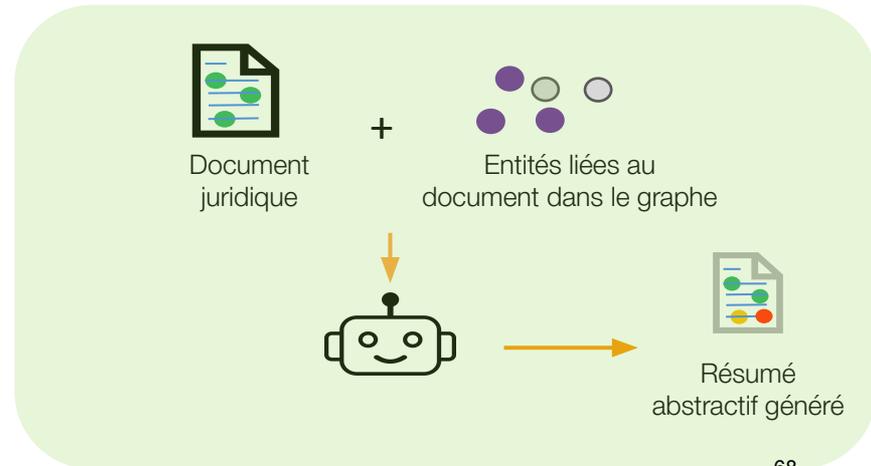
## Inférence

Documents du test



Extraction des entités

Graphe d'entités





# Quelques résultats

Modèle	Faithfulness ↑	Factual Hallucination ↑	Non Factual Hallucination ↓	Bert-Score ↑	Fact-CC ↑	ROUGE ↑	
Modèle de l'état de l'art → ECC	0,21	0,07	0,70	0,84	0,36	0,22	
Oracle	0,21	0,11	0,60	0,86	0,39	0,26	
Différents modes d'entraînement	Setting 0	0,17	0,04	0,73	0,83	0,27	0,20
	Setting 1	0,15	0,05	0,78	0,86	0,34	0,23
	Setting 2	0,20	0,08	0,68	0,87	0,39	0,21
	Setting 3	0,20	0,08	0,68	0,86	0,40	0,24

Métriques orientées entités



Plus d'hallucinations factuelles



Moins d'hallucinations non-factuelles

# Quelques extensions et travaux en cours

Plutôt que d'injecter des entités dans le modèle

=> injecter des **triplets relationnels** ?

Il nous faut donc les triplets....

=> annotation du corpus EurLex en triplets

Résultats:

**Meilleurs!** y compris sur les métriques non-orientés triplets

**Text** : **Regulation 2018/1046** of the European Parliament and of the Council of 18 July 2018 on the financial rules applicable to the general budget of the Union, **amending Regulation 1296/2013, Regulation 1301/2013, Regulation 1303/2013, Regulation 1304/2013, Regulation 1309/2013, Regulation 1316/2013, Regulation 223/2014, Regulation 283/2014, and Decision 541/2014/eu** and **repealing Regulation 966/2012** (OJ L 193, 30.7.2018, p. 1).

**Annotation Process**

- Click the Add button to start the annotation process.
- Double-click or highlight the part of the text corresponding to an entity or an action. Please select the 3 elements of the triplet in the following order: Entity 1 → Action → Entity 2.
- You can see the current selection items and cancel it at anytime.
- Once the 3 element of a selection are made, it will automatically be added to the Selected Triples section. If a triplet needs to be removed it can be done there too.
- Once the annotation is finished, please click on the "Download" button to download the annotations or the "Next Document" button to download the annotations and move to next document.

Home Go to the document translation

**Entity Highlighter**

Previous regulation 2019/631 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL---of 17 April 2019---setting CO2 emission performance standards for new passenger cars and for new light commercial vehicles, and repealing regulation 443/2009 regulation 510/2011 Next

1/17 Add Download

**Selected Triples**

p1, regulation 2019/631, repealing, regulation 443/2009	Remove
p1, regulation 2019/631, repealing, regulation 510/2011	Remove

# Quelques extensions et travaux en cours

Utilisation de LLM plutôt que de PLM fine-tunés

---

---

**Base Prompt (Zero-shot Learning) :**

You are a legal expert specialized in generating summaries of legal documents from EUR-Lex.

**Task :** Generate a summary of the legal document provided below.

**Summary specifications :**

- The summary must mention the legal entities present in the document (directives, regulations, decisions, resolutions, recommendations, opinions)
- The summary must specify the relations between legal entities (e.g., to amend, to repeal, to correct, to supplement, to extend, to recast, to replace) to form relational triples.
- The summary must be coherent, fluent, and factually accurate

**Document to summarize :**

{document}

**Final instructions :**

Generate a summary that captures the essential information from the document while preserving references to legal entities and their relations.

---

Pour le moment, **pas d'avancée notable**

Problème de prompt ?

**En conclusion**

# La souveraineté en IA

## Construction du modèle de base

### Maitrise des données

- Collecte/équilibrage
- Nettoyage

### Entrainement

- Puissance machine (milliers de GPU)
- Architecture/recherche ML

## Raffinement du modèle

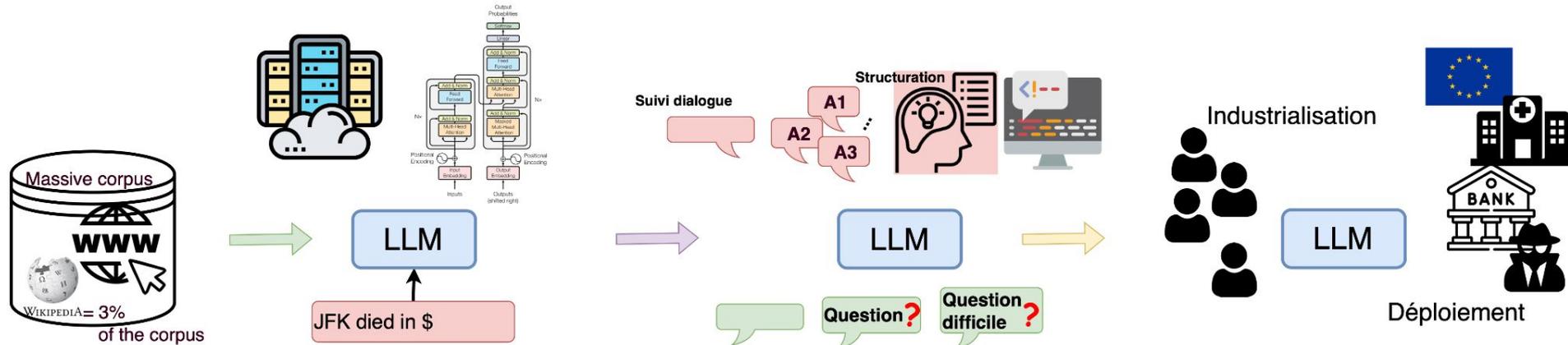
### Maitrise & construction des données

- interactions humaines +++
- prix des données
- spécialisation à la demande

## Exploitation du modèle

### Optimisation / Limitation du coût

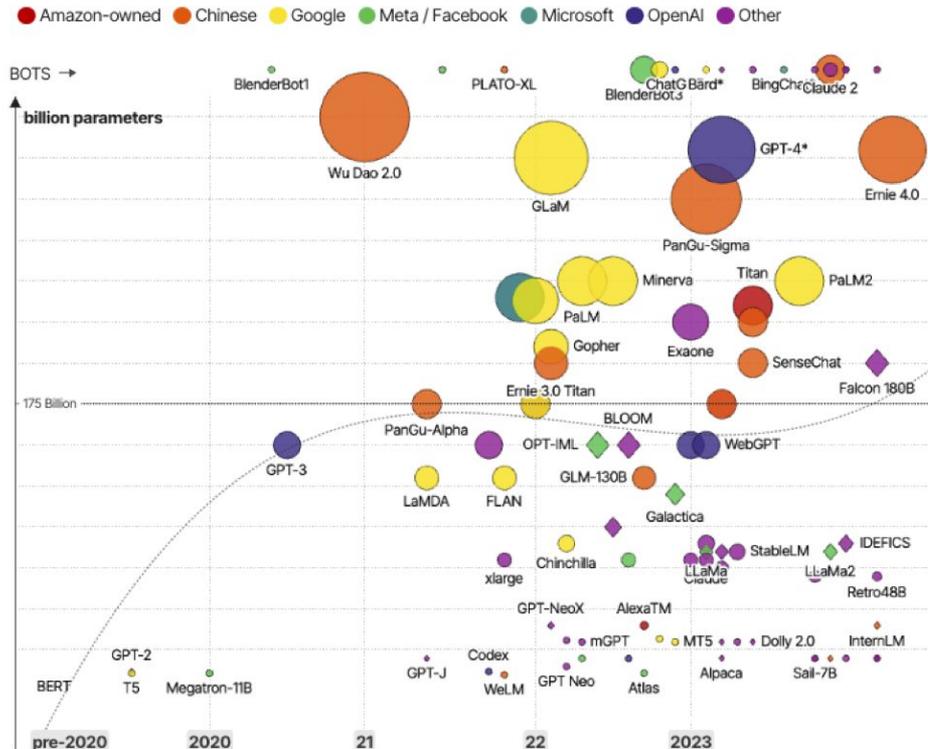
- compétence MLOps
- déploiement local



# Des solutions chères

Des capacités qui émergent avec la taille des modèles

## The Rise and Rise of A.I. Large Language Models (LLMs) & their associated bots like ChatGPT



## # Paramètres

1998	LeNet-5	= 0,06M
2011	Senna	= 7,3M
2012	AlexNet	= 60M
2017	Transformer	= 65M / 210M
2018	ELMo	= 94M
2018	BERT	= 110M / 340M
2019	GPT-2	= 1 500M
2020	GPT-3	= 175 000M
2025	Llama-4	= 2 000 000M

# Les risques à utiliser un système d'IA

Biais cognitif	Effet sur l'utilisateur	IA génératives	IA & diag. médical
Biais d'automatisation	Acceptation passive des résultats (réduction de la vérification)	Élevé	Très élevé
Biais d'autorité	Surestimation de l'autorité (et de la fiabilité) de l'IA	Élevé	Très élevé
Surconfiance / mauvaise calibration de la confiance	Utilisation inappropriée de l'IA dans des situations où elle est peu fiable	Très élevé	Élevé
Biais de confirmation	Renforcement d'hypothèses erronées et réduction de l'exploration d'alternatives	Modéré	Élevé
Décharge cognitive	Diminution de l'analyse critique et de la métacognition	Élevé	Très élevé
Effet de cadrage. Illusion de profondeur explicative	Interprétation biaisée du niveau de certitude ou de risque (style de la réponse, présentation de chiffres)	Très élevé	Modéré
Biais d'anthropomorphisme	Attribution de capacités humaines (confiance++)	Très élevé	Modéré