

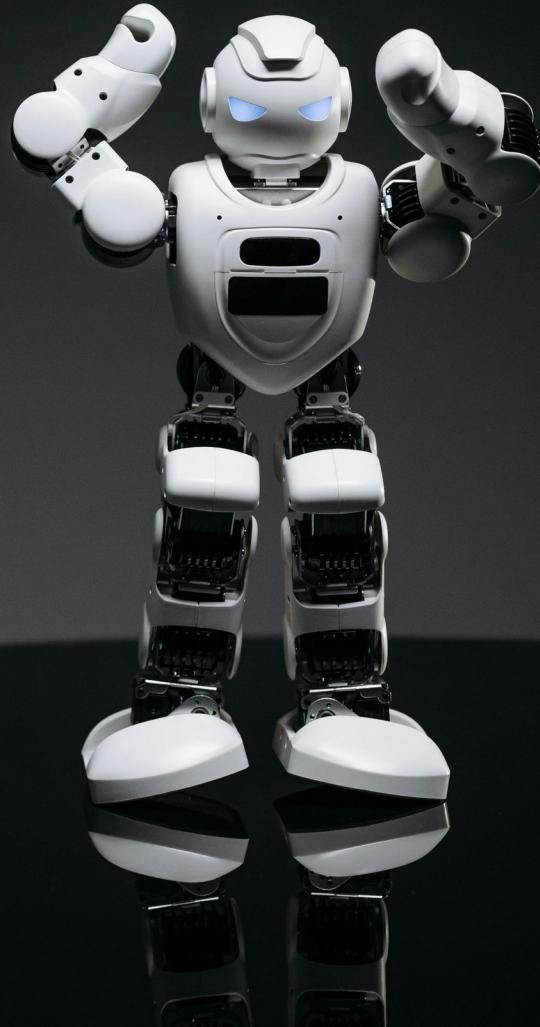
LHIST2532 // Archivage Numérique

IA ET ARCHIVES

Année académique 2025-2026

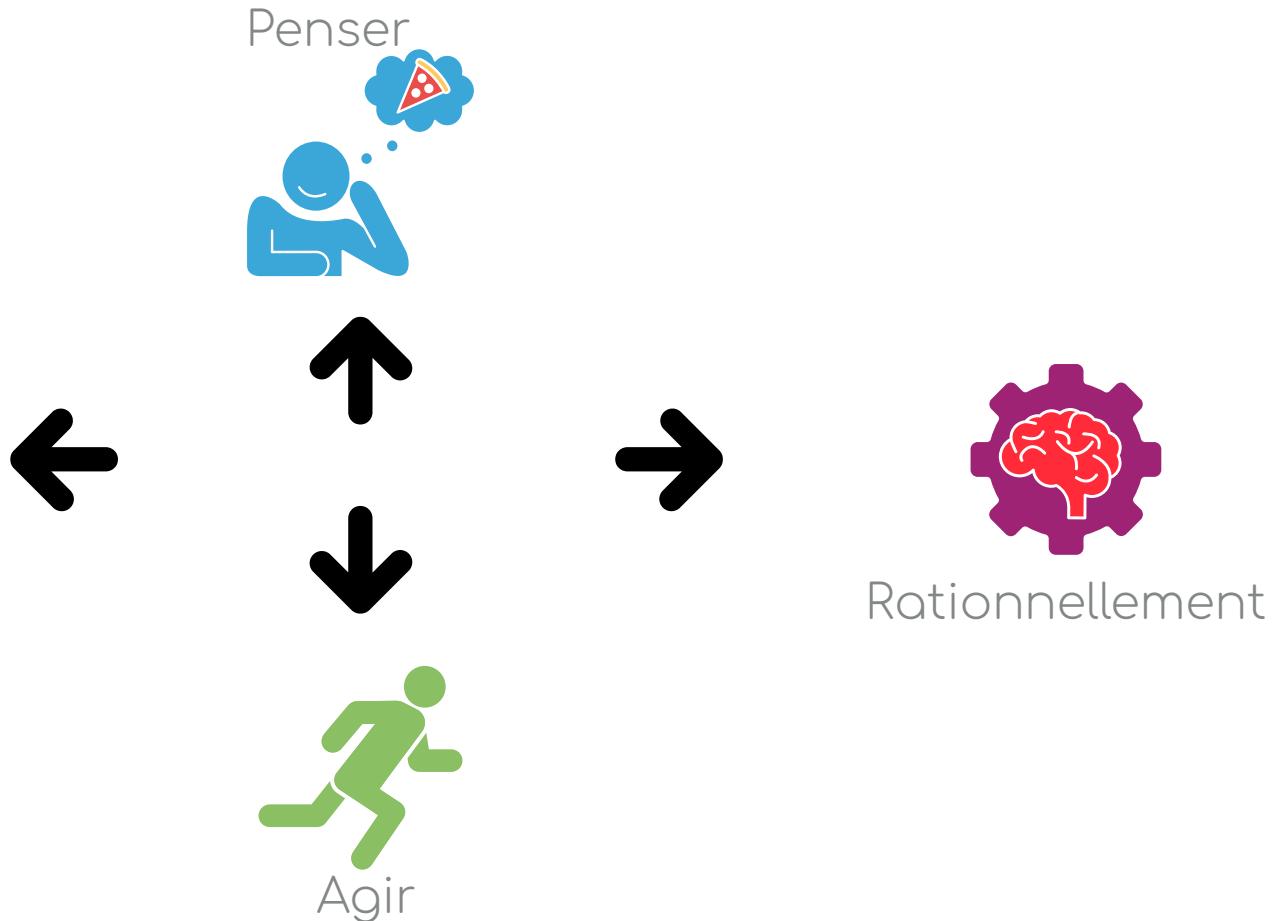
Xavier GILLARD

L'IA qu'est ce que c'est ?
RÉELLEMENT



TENTATIVE DE DEFINITION

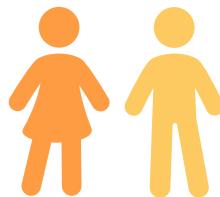
Topologie de Russell Norvig.
Artificial Intelligence a Moder Approach (3rd ed)
Prentice Hall



TENTATIVE DE DEFINITION

Topologie de Russell Norvig.
Artificial Intelligence a Moder Approach (3rd ed)
Prentice Hall

Activités qu'on associe au fait de penser (planning, prise de décision, ...)

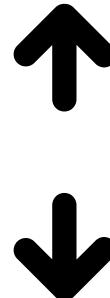


Humainement

Activités pour lesquelles les humains sont meilleurs (peinture, poésie, ...)

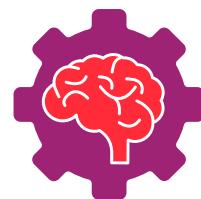


Penser



Agir

Étude de ce qui nous permet de penser et agir rationnellement (logique appliquée)

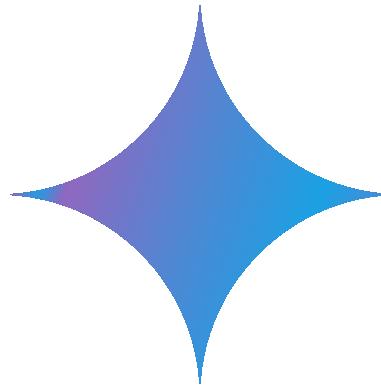
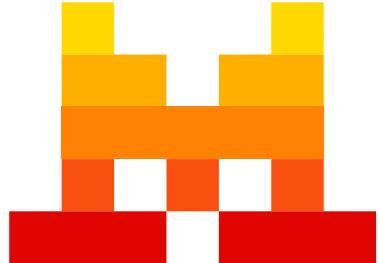


Rationnellement

Systèmes doués d'intelligence

L'IA C'EST TOUT CE QUI
PERMET DE RÉSOUDRE DE
PROBLÈMES TROP DIFFICILES
POUR UN ORDINATEUR

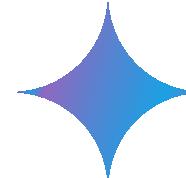
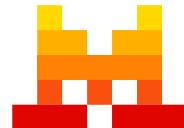
DANS LA PRESSE



CE DONT ON VA PARLER

Traitement Automatique de la Langue Naturelle (TALN)
⇒ Natural Language Processing (NLP)

avec de **Grands Modèles de Langage (LLM)**



TERMINOLOGIE

virtuo *vər-tü-əs* |'vər-chə-wəs| *ad.*
or showing virtue and esp.
2 : CHASTE — vir·tu·ous·ly
vir·u·lent |'vir-ə-lənt, 'vir-yə-lənt| *adj.*
: highly infectious *(a ~ generally fatal)*
: marked by a rapid, severe, deadly course *(a ~ disease)*
extremely poisonous or venomous *(a ~ poison)*
ious **3 : full of malice : MALICIOUS**
vir·u·lence |-ləns| *n.* — **vir·u·len-**
vi·rus |'vī-rəs| *n.*, *pl.* **vi·rus·es** [*L.*,
poisonous emanation] **1 : any of a group of**
agents that have an outer protein shell around a core
can grow and multiply in living cells

TERMINOLOGIE



Algorithmme

Une suite d'étapes qui permet d'accomplir un objectif.

(une *recette de cuisine*)



Programme

Ensemble d'instructions écrites dans un *langage de programmation* qui permet à un ordinateur d'accomplir une tâche donnée.

(réalisation d'un algo.)



Modèle

Représentation simplifiée du monde qui sert à faire des prédictions.

LARGE LANGUAGE MODEL

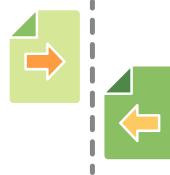
Représentation simplifiée de la
langue visant à faire des
prédictions à propos de texte

TYPES DE PREDICTIONS



Classification

- Quel sujet ?
- Quelle émotion ?



Comparer

- Similarité entre deux textes
- Déetecter les doublons



Retrouver de l'information

- Personnes, lieux, dates, ...
- Réponse à une question



Corriger

- Corriger un texte
- Compléter des trous

EXEMPLE

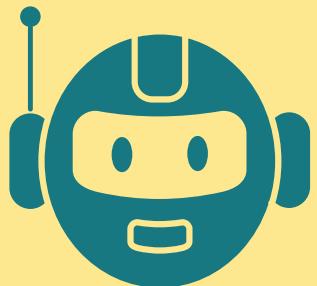


Après une soirée à la CASA, il est plus difficile de me [MASQUE] le vendredi matin

EXEMPLE

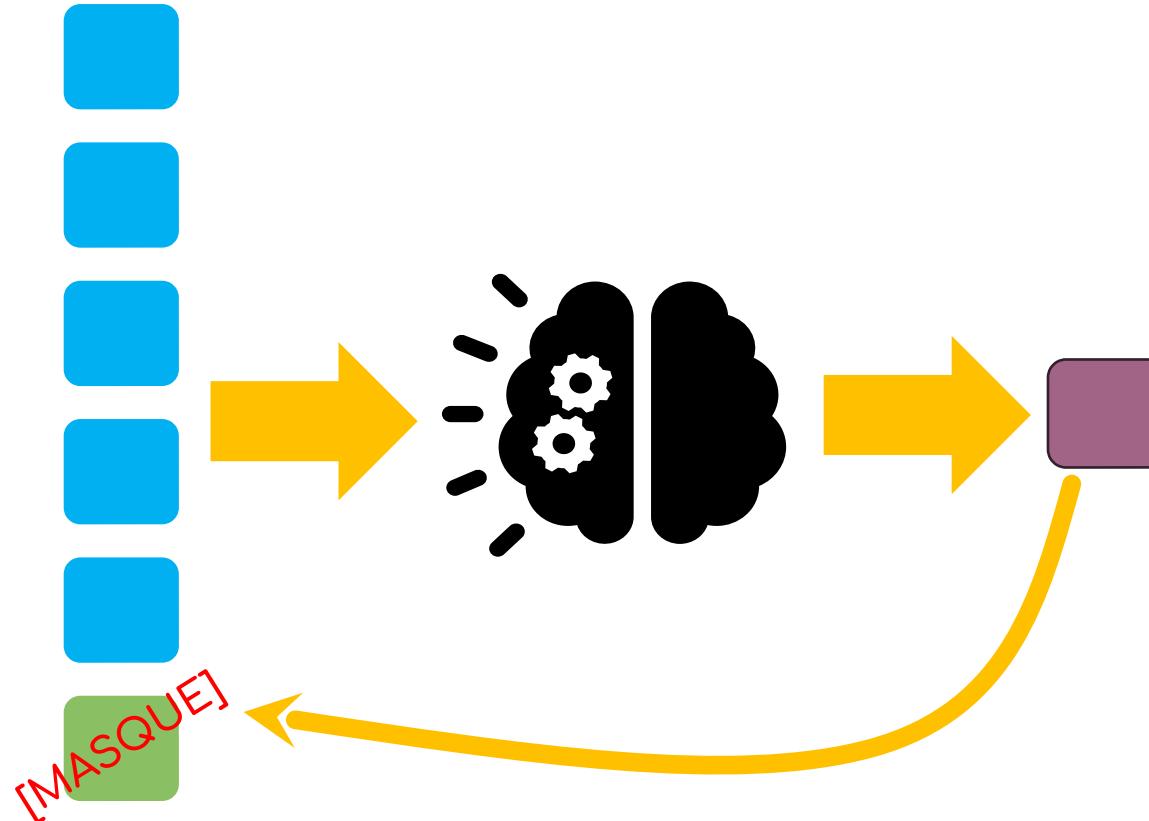


Après une soirée à la CASA, il est plus difficile de me **[MASQUE]** le vendredi matin



- Lever
- Réveiller
- Bouger

ET POUR GENERER DU TEXTE ?



**LES GPT NE FONT JAMAIS QUE
CONTINUER UN TEXTE QUI
LEUR EST DONNÉ (LE PROMPT)
UNE PRÉDICTION À LA FOIS**

Fondamentalement, ils appliquent tous le même algorithme*

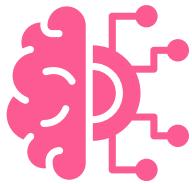
*Les modèles diffusion fonctionnent un peu différemment

GENERATIVE PRETRAINED TRANSFORMERS (GPT)



TRANSFORMERS

Attention is all you need
A. Vaswani et al
NeurIPS 2017



Architecture de Réseau de Neurones Profond
(Deep Neural Network)

Opérations **mathématiques** uniquement



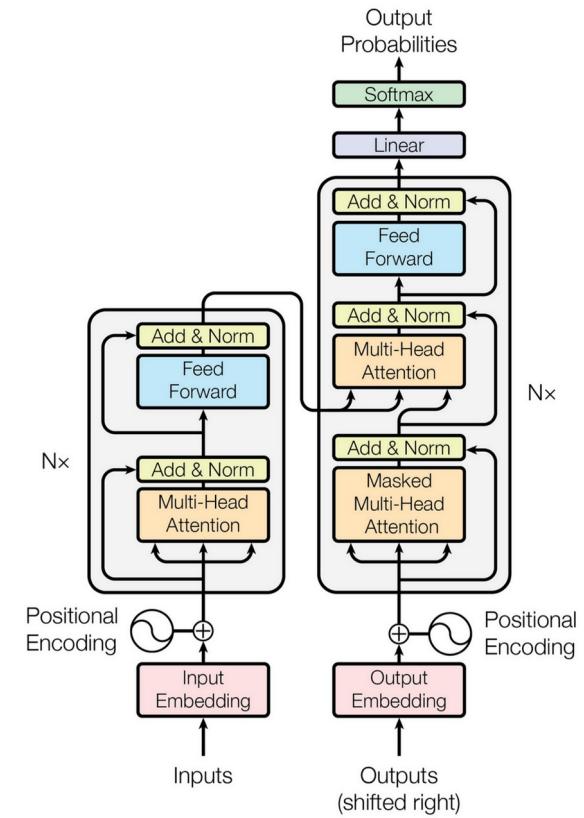
Architecture

Quelles opérations
math et dans quel
ordre



Poids

'Connaissances'
encodées dans le
modèle



**SI ON NE FAIT QUE DES
MATH, D'OU VIENNENT
LES MOTS ?**



TOKENISATION

Les **machines** ne connaissent pas de mots, juste un **ensemble fini de nombres**. Il faut donc utiliser un **vocabulaire fini** pour **transformer le texte** en une suite de nombres

EXEMPLE



Il est difficile de se lever le vendredi matin

EXEMPLE



Il est difficile de se lever le vendredi matin

il	00	lever	08
est	01	le	09
di	02	ve	10
##ffi	03	##nd	11
##ci	04	##red	12
##le	05	##i	13
de	06	mat	14
se	07	##in	15

EXEMPLE



Il est difficile de se lever le vendredi matin

Il est di ##ffi ##ci ##le de se lever le ve ##nd ##red ##i mat ##in

00 01 02 03

04

05

06

07

08

09

10

11

12

13

14

15



UNE MISCONCEPTION A OUBLIER

SOUVIENS-TOI IL Y A 5 SLIDES



Architecture

Quelles opérations
math et dans quel
ordre



Poids

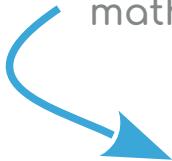
'Connaissances'
encodées dans le
modèle

SOUVIENS-TOI IL Y A 5 SLIDES



Architecture

Quelles opérations
math et dans quel
ordre



**Fixé d'avance,
une fois pour toutes**



Poids

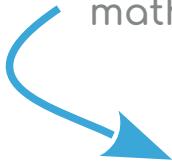
'Connaissances'
encodées dans le
modèle

SOUVIENS-TOI IL Y A 5 SLIDES



Architecture

Quelles opérations
math et dans quel
ordre



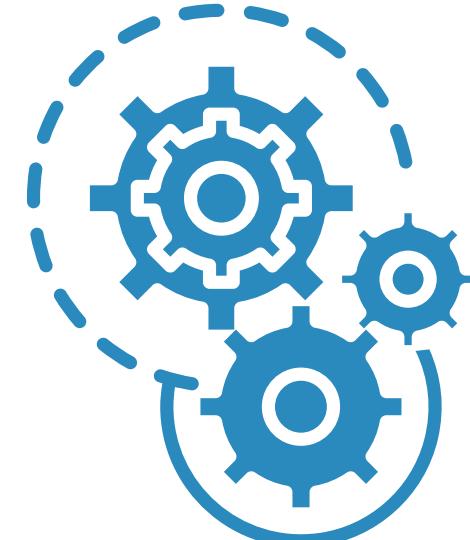
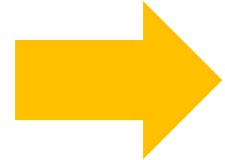
**Fixé d'avance,
une fois pour toutes**



Poids

'Connaissances'
encodées dans le
modèle

ENTRAINEMENT

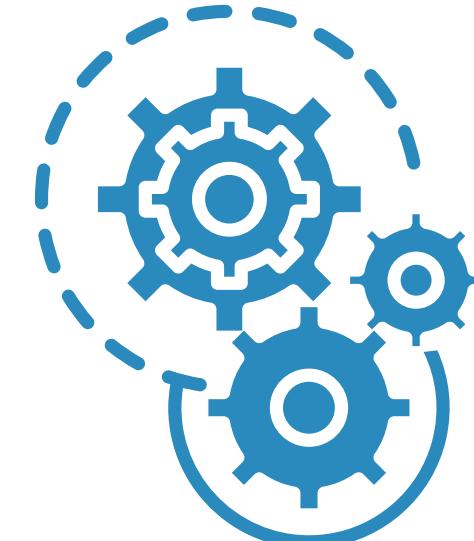
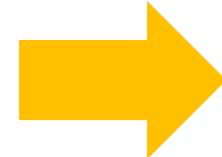


Valeur attendue

08

lever

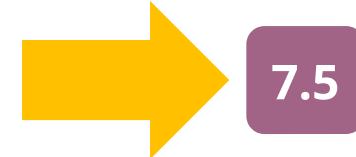
ENTRAINEMENT



Valeur attendue

08

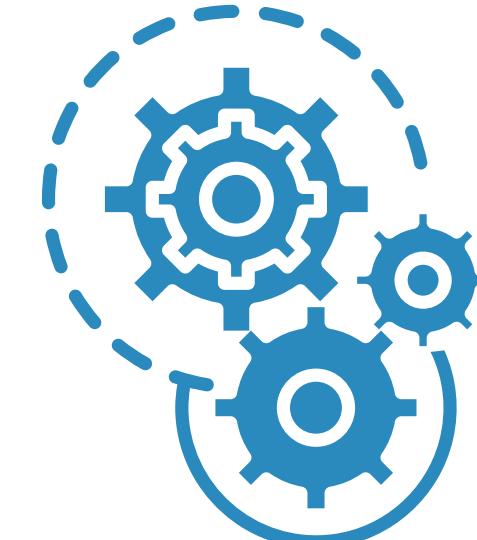
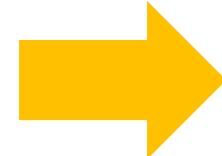
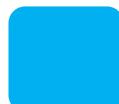
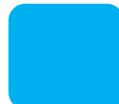
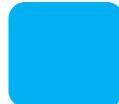
lever



7.5

Prédiction

ENTRAINEMENT



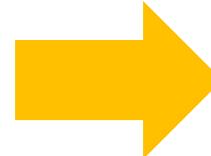
Valeur attendue

08

lever



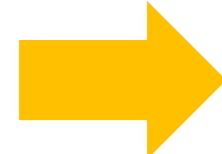
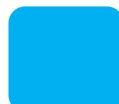
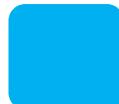
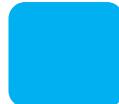
Erreur



7.5

Prédiction

ENTRAINEMENT



Valeur attendue

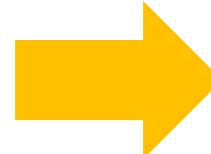
08

lever



Erreur

7.5



Prédiction

SOUVIENS-TOI IL Y A 5 SLIDES



Poids

'Connaissances'
encodées dans le
modèle

**ENTRAINER UN MODÈLE =
AJUSTER LES POIDS**

**UNE FOIS L'ENTRAINEMENT
TERMINÉ, UN MODÈLE
N'APPREND PLUS RIEN**

POUR CLARIFIER ENCORE



Entrainement



Inférence

POUR CLARIFIER ENCORE



Entrainement

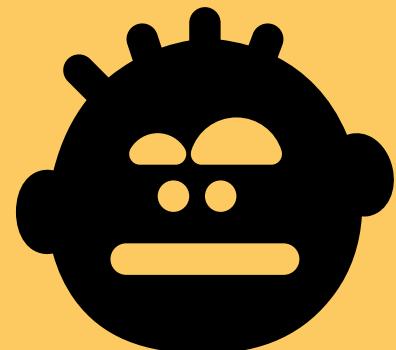
*2^e phase de la vie d'un modèle.
Utilisé pour faire des prédictions*

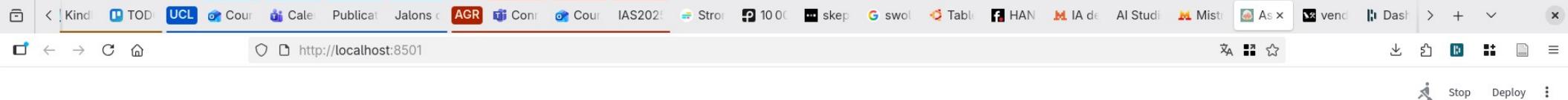


Inférence



**POURTANT CHATGPT
ME CONNAÎT SUPER BIEN**





Ask Agatha



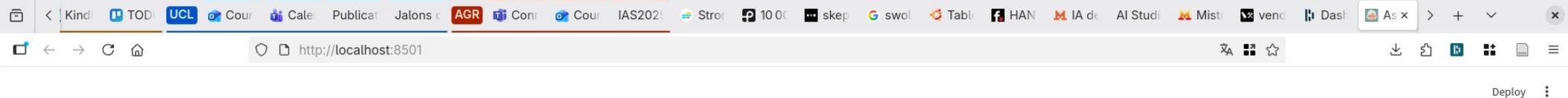
Mon arrière grand mere m'a beaucoup parlé de sa copine chrispina avec qui elle a passé les premières années de sa vie à Aardooie pendant la WW1. Elles allaient toutes les deux chez la même gardienne d'enfant et sont restées amies toute leur vie. J'aimerais prendre contact avec ses arriere petits enfants à elle si elle en a eu. Peux tu m'aider ?



I want to find the details of a birth record where the first name is Chrispina in Aardooie between 1910 and 1920.

Le bon roi dagobert avait-il sa culotte à l'envers ?





Bonjour,

Je comprends que vous cherchez à retrouver les arrière-petits-enfants de Chrispina, une amie de votre arrière-grand-mère qui vivait à Ardoorie pendant la Première Guerre mondiale.

Malheureusement, je n'ai pas pu trouver de scan numérisé du certificat de naissance de Chrispina Maria Ameye pour l'année 1916 à Ardoorie. Cependant, j'ai trouvé des informations sur son acte de naissance.

Chrispina Maria Ameye est née le 28 novembre 1916 à Ardoorie. Ses parents étaient Camille Alfred Ameye et Herminie Marie Verougstraete. Ils se sont mariés à Ardoorie le 25 mai 1916.

Pour continuer votre recherche, vous pourriez essayer les étapes suivantes :

1. **Recherche dans les registres paroissiaux** : Parfois, les registres paroissiaux contiennent des informations supplémentaires ou différentes de celles des registres civils.
2. **Consultation des archives locales** : Vous pourriez contacter les archives locales ou les services d'état civil d'Ardoorie pour voir s'ils ont des informations supplémentaires ou des registres non numérisés.
3. **Recherche généalogique en ligne** : Il existe plusieurs sites web et forums de généalogie où vous pourriez poster une demande d'information. Des personnes ayant des arbres généalogiques publics pourraient avoir des informations sur la famille Ameye.
4. **Réseaux sociaux et groupes de généalogie** : Rejoindre des groupes de généalogie sur les réseaux sociaux pourrait également être utile. Vous pourriez trouver des personnes qui ont des liens avec la famille Ameye ou qui pourraient vous aider dans votre recherche.
5. **Consultation des recensements** : Les recensements peuvent fournir des informations sur les membres de la famille et leurs descendants. Vous pourriez trouver des recensements en ligne ou dans les archives locales.

Malheureusement, sans accès à des scans numérisés ou à des informations supplémentaires, il est difficile de poursuivre la recherche directement. Cependant, en suivant ces étapes, vous

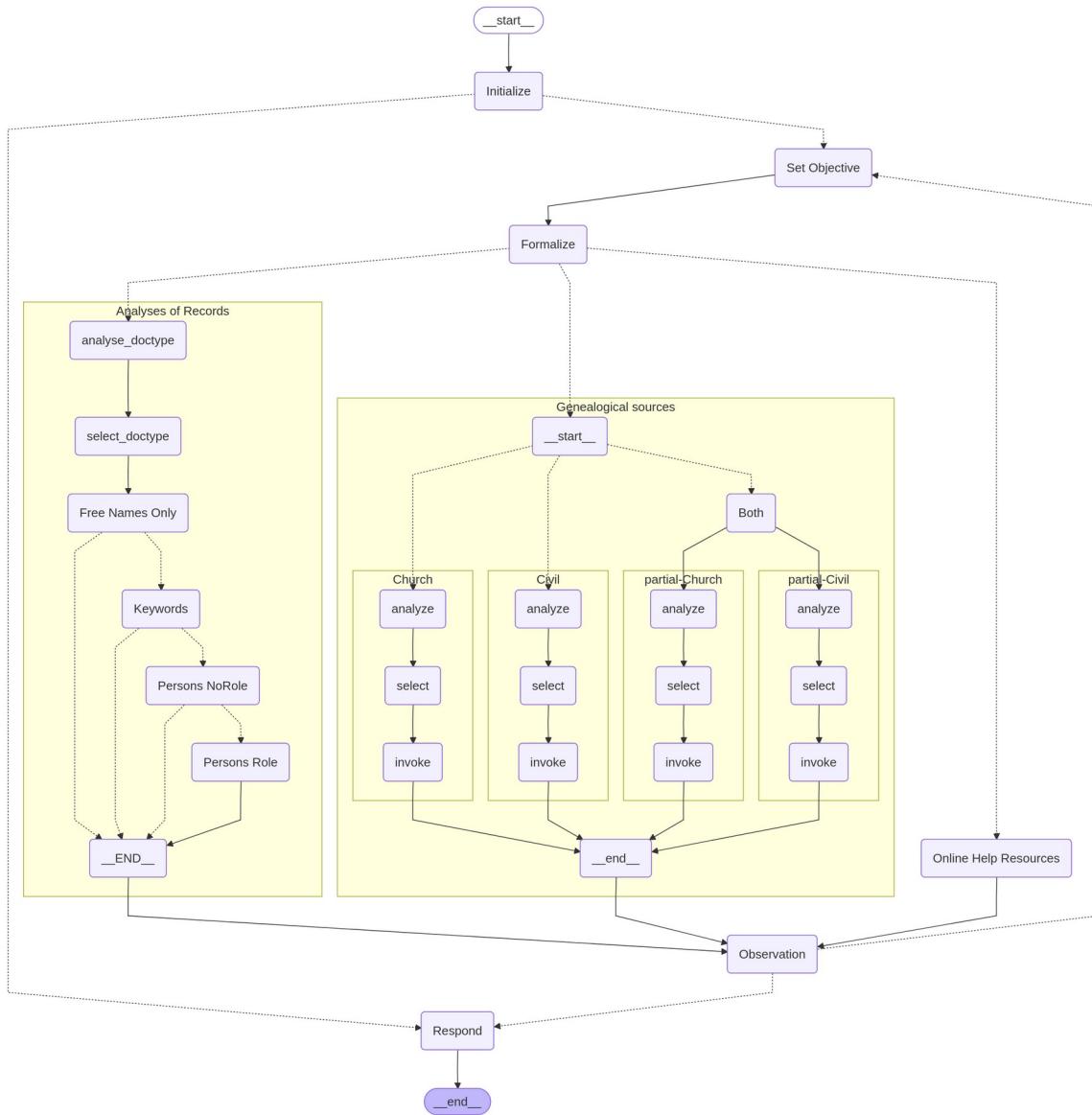
C'EST MAGIQUE
NON ?



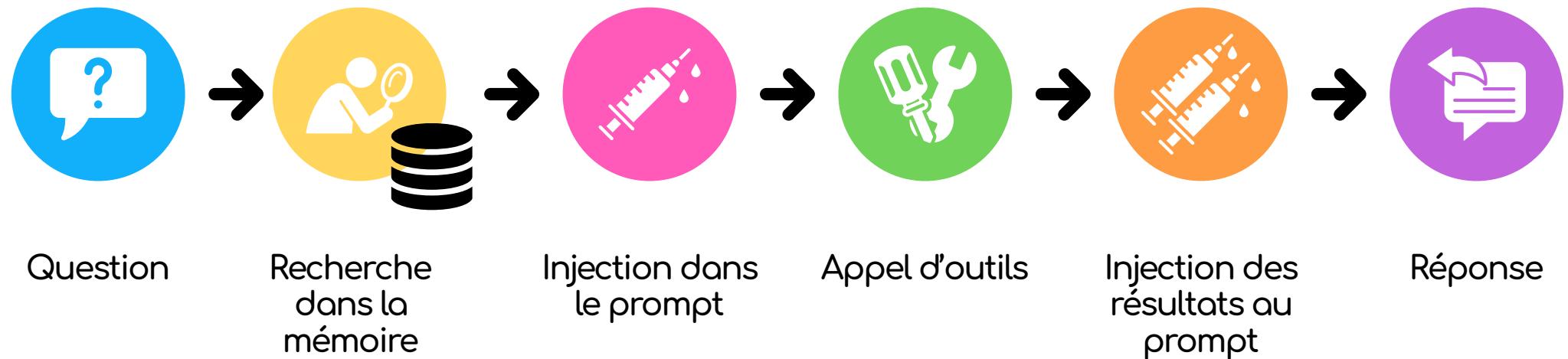
LA RÉALITÉ N'EST
PAS EXACTEMENT
CELLE-LÀ



SYSTÈMES COMPLEXES INTERACTIONS INVISIBLES



SI LE SYSTEME SE RAPPELLE DE VOUS



A CHAQUE ÉTAPE
LE SYSTÈME
PEUT (VA)
SE TROMPER



WRONG
WAY

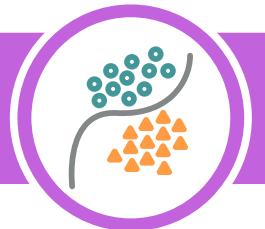
QUELLES ERREURS ?



Hallucinations

Texte vraisemblable
mais factuellement faux

Erreurs de classification



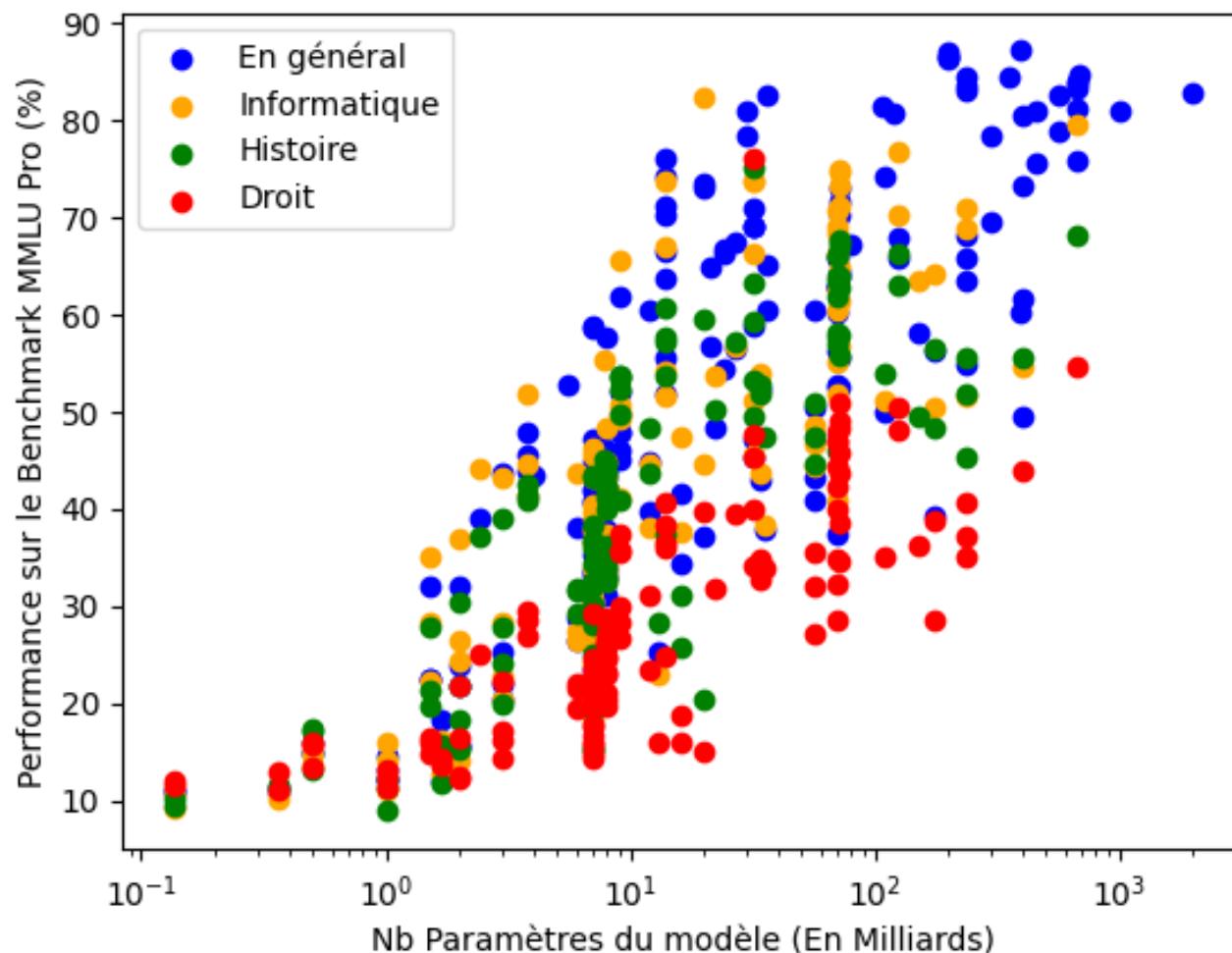
Mauvais outil
Mauvaise préférence
Donnée existante non-retrouvée
Donnée non-pertinente



Biais

Dû aux données d'entraînement
Induit par le prompt

COMME SI CA NE SUFFISAIT PAS



COMME SI CA NE SUFFISAIT PAS

 Différentes tailles de modèle

90

En général

70

60

 Corrélation taille - performance

Performance sur

30

 La performance dépend du sujet

10^{-1}

10^0

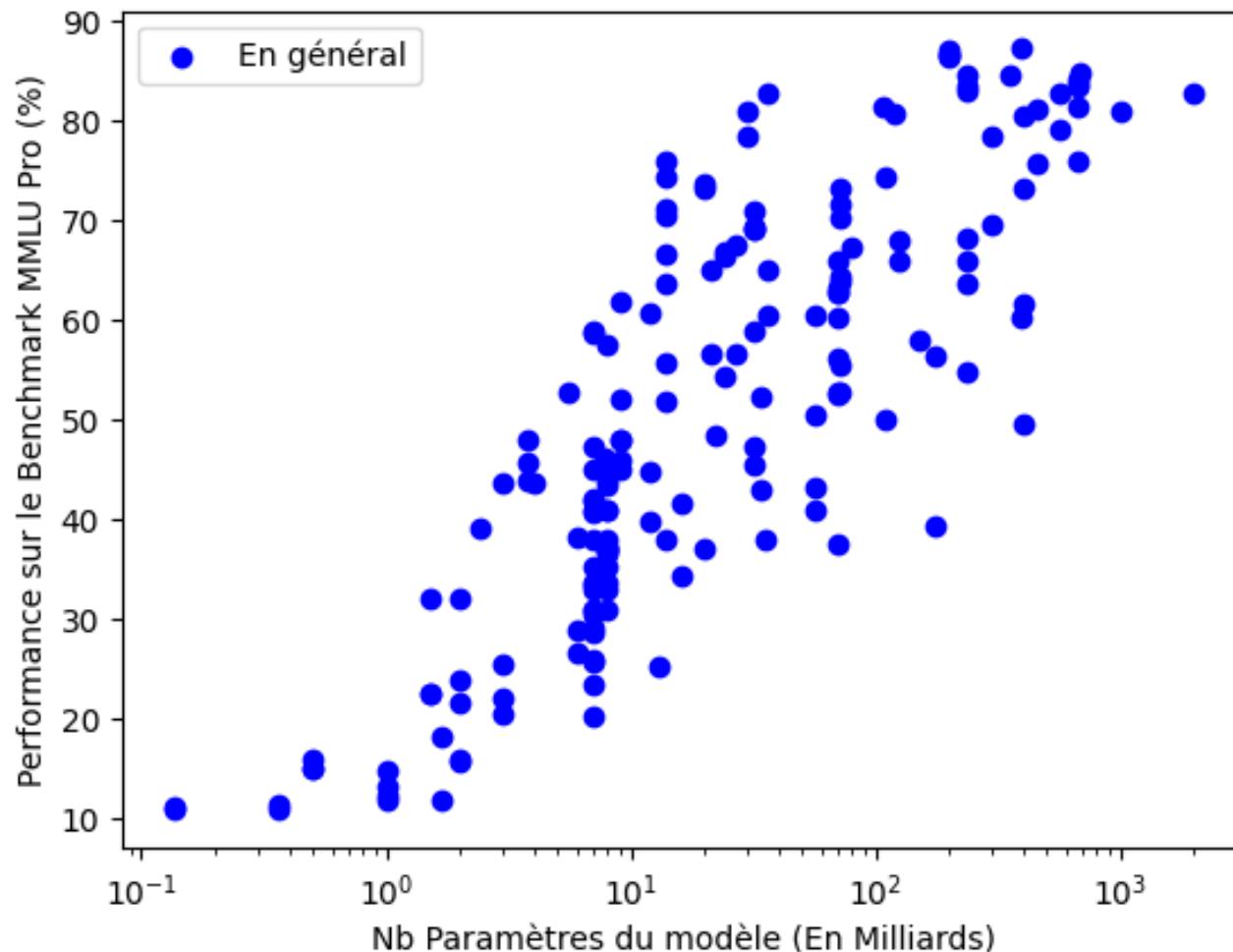
10^1

10^2

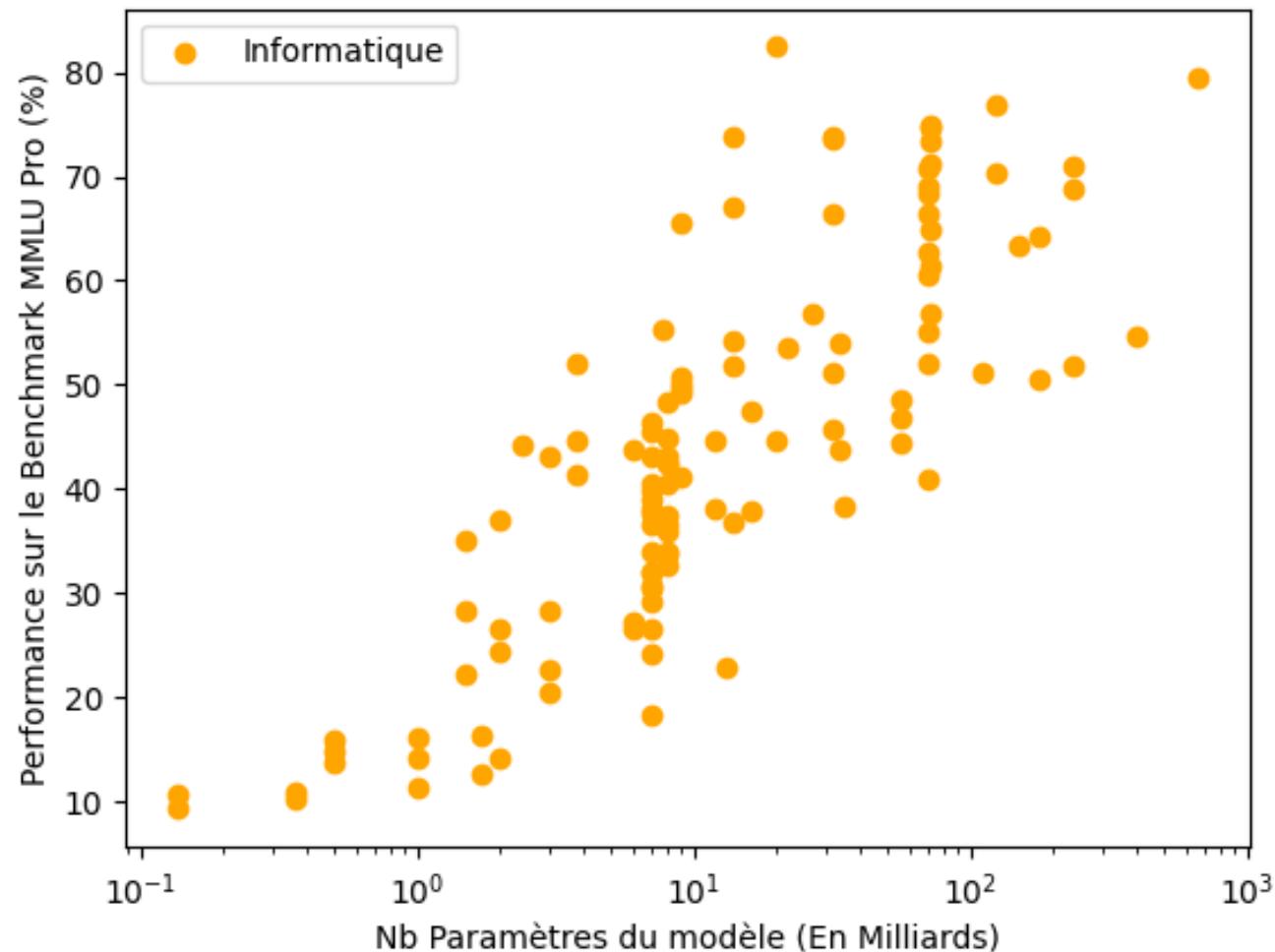
10^3

Nb Paramètres du modèle (En Milliards)

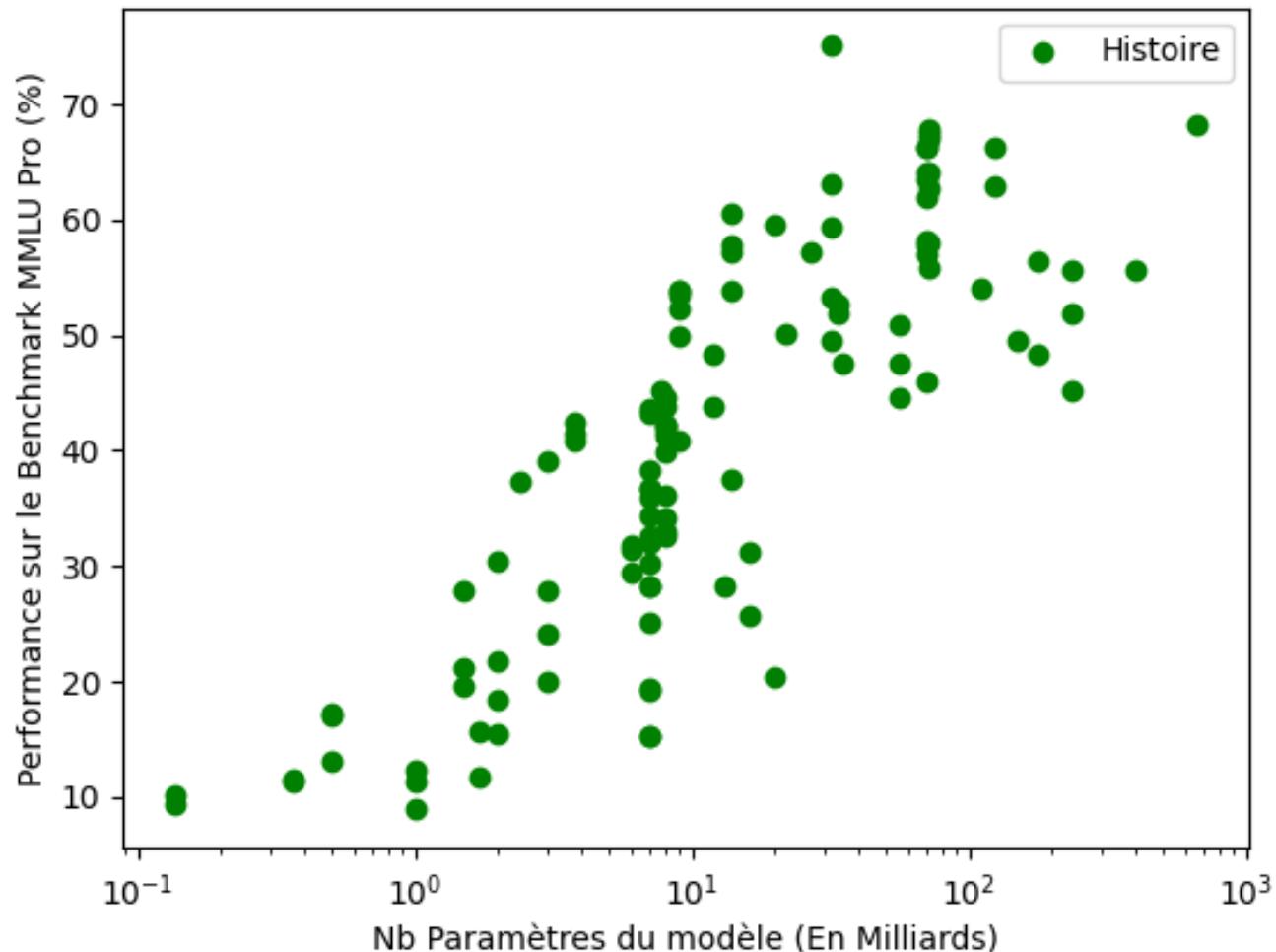
COMME SI CA NE SUFFISAIT PAS



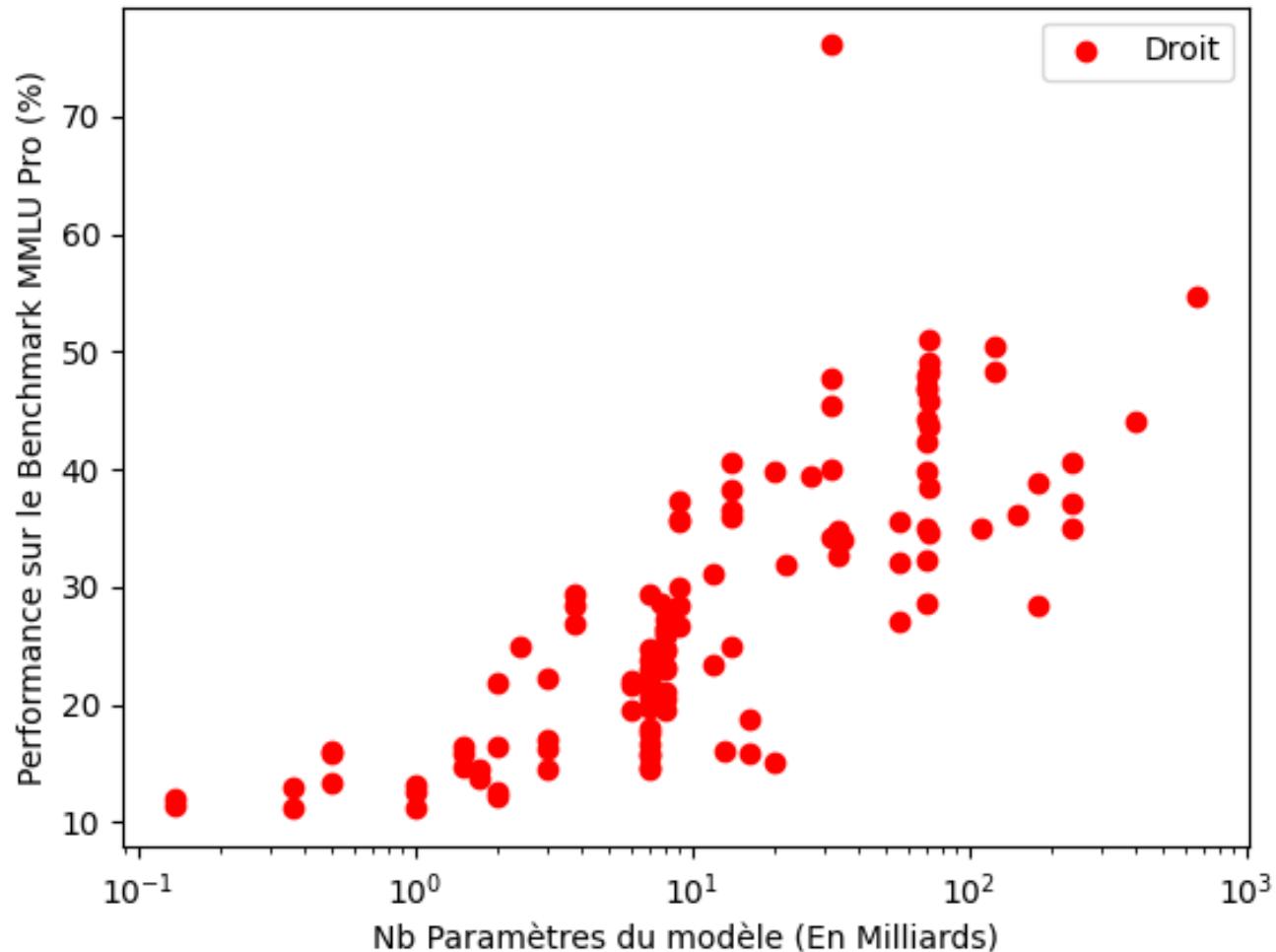
COMME SI CA NE SUFFISAIT PAS



COMME SI CA NE SUFFISAIT PAS



COMME SI CA NE SUFFISAIT PAS



**LA TAILLE DU MODÈLE
DÉTERMINE LE COÛT DE SON
UTILISATION**

MATERIEL SPECIFIQUE



CPU

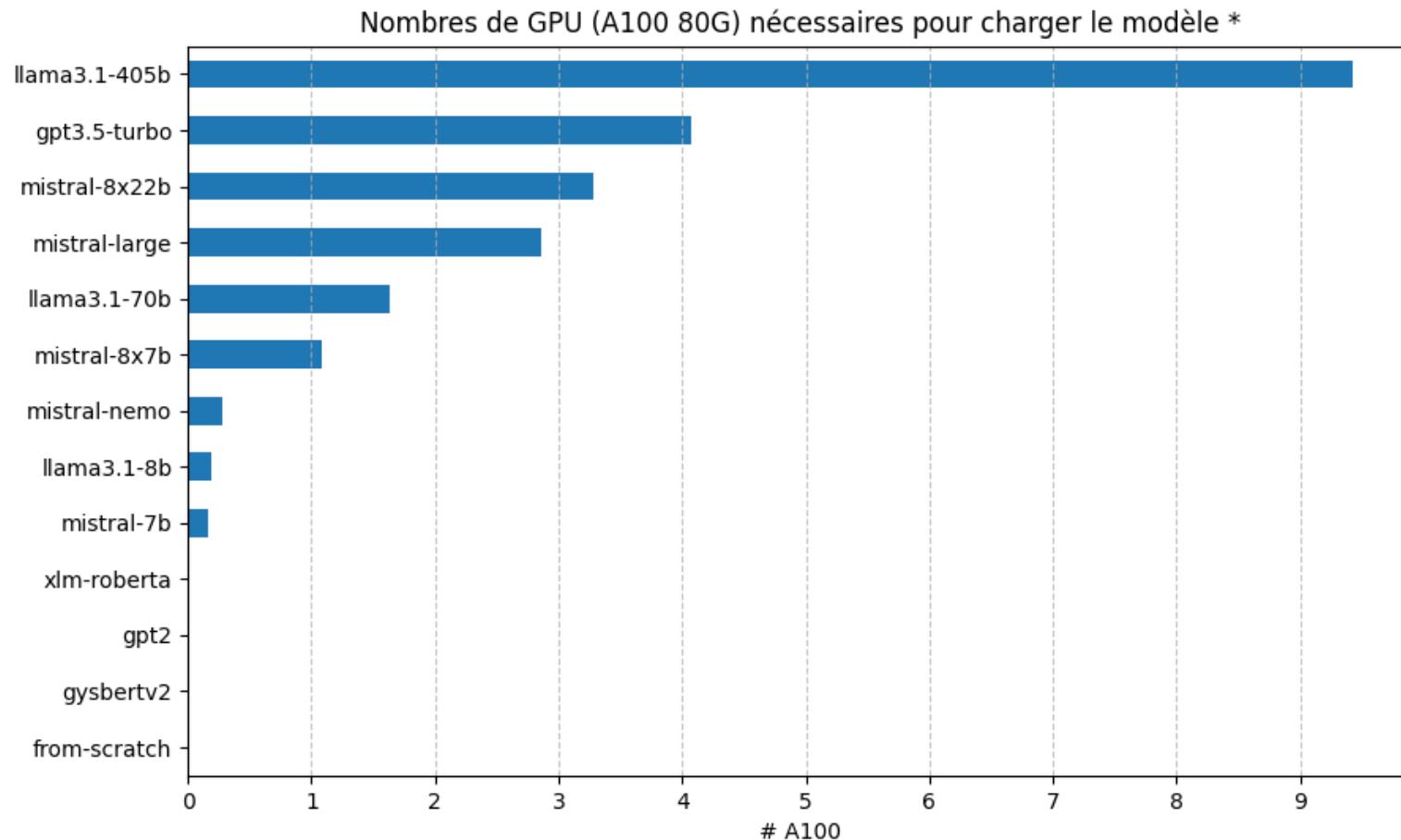
Très rapide
Versatile
Une chose à la fois

GPU

Moins rapide
Ne peut pas tout faire
Énormément de calcul en une fois



COMBIEN CA COUTE ?



CHAQUE GPU COUTE
ENVIRON 20 000 EUR



QUE FAIRE POUR QUE ÇA MARCHE QUAND MÊME ?

→ Du prompt engineering ←

5 PILIERS DU PROMPT ENG.

Prompt Engineering for
Generative AI: Future-Proof
Inputs for Reliable AI Outputs
James Phoenix, Mike Taylor
Oreilly, 2024



Direction

Indications
très précises



Format

Quel résultat
obtenir



Exemples

Montrer ce
qu'il faut faire



Qualité

Contrôler la
qualité des
résultats



Diviser

Allez-y pas à
pas. Surtout si
c'est complexe

5 PILIERS DU PROMPT ENG.

Prompt Engineering for
Generative AI: Future-Proof
Inputs for Reliable AI Outputs
James Phoenix, Mike Taylor
Oreilly, 2024



Qualité

Contrôler la
qualité des
résultats

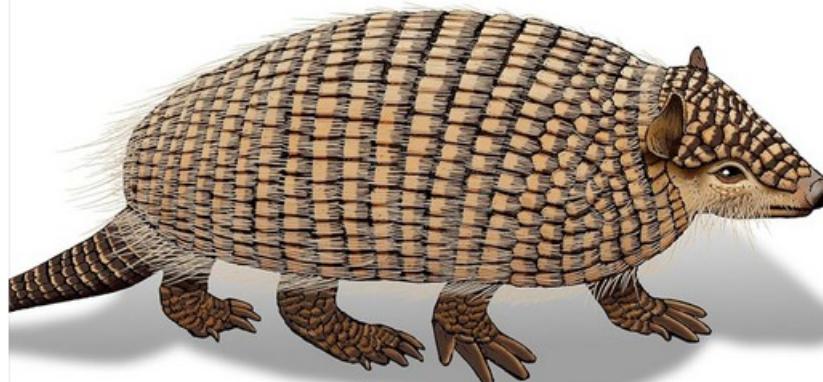
Diviser

Allez-y pas à
pas. Surtout si
c'est complexe

O'REILLY®

Prompt Engineering for Generative AI

Future-Proof Inputs for Reliable AI Outputs



James Phoenix
& Mike Taylor



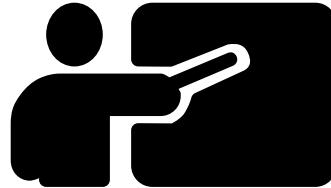
**QUEL INTÉRÊT ?
JE SUIS ARCHIVISTE
PAS INGÉNIEUR**

DEUX UTILITES ESSENTIELLES



**Des super-pouvoirs
pour votre pratique**

Par Ex : Automatisation de
versements



**La pédagogie sera de
plus en plus nécessaire**

- * Comprendre et expliquer les hallucinations.
- * Restaurer la confiance

**ENFILEZ VOS
HOODIES, ON VA
METTRE ÇA EN
PRATIQUE...**

