université PARIS-SACLAY

IA GÉNÉRATIVES: DES OUTILS ET DES DÉFIS

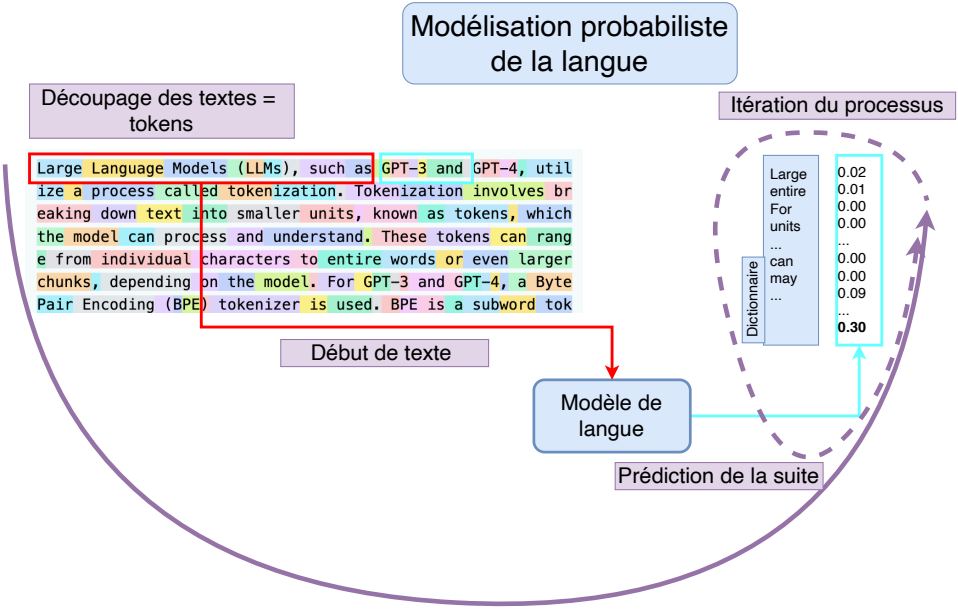
Jeudi 10 avril 2025
Assemblée Générale d'AgroParisTech Alumni
Agros & IA : Regards croisés sur une révolution en marche

Vincent Guigue
https://vguigue.github.io

MIA PARIS-SACLAY
EKINOCS

AGROPARISTECH
Institut des Sciences et Industries du Vivant et de l'Environnement

Les modèles de langue en 5 tableaux



Les modèles de langue en 5 tableaux

Pré-entrainement
Compléter des phrases

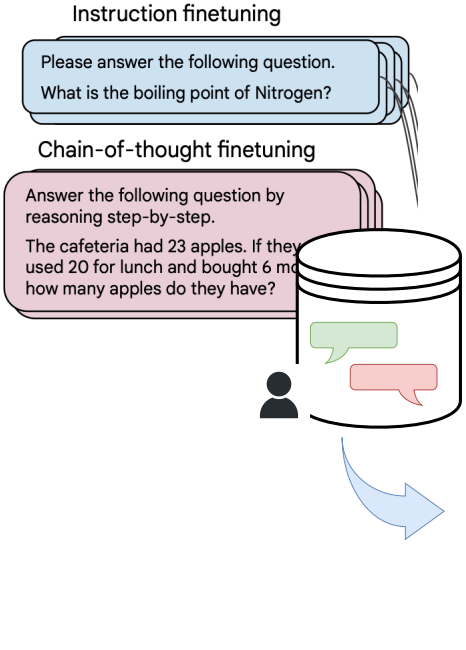
1963, he was was assassinated in Dallas ...

Modèle de langue

JFK died in

Acquisition de:
grammaire
entités
connaissances
raisonnement (?)

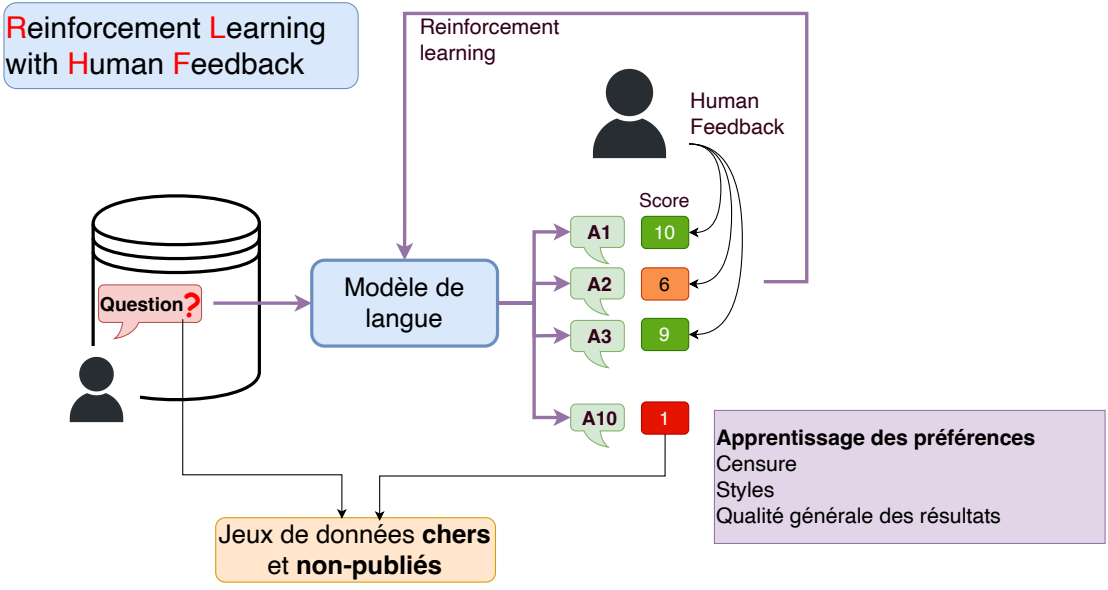
Les modèles de langue en 5 tableaux



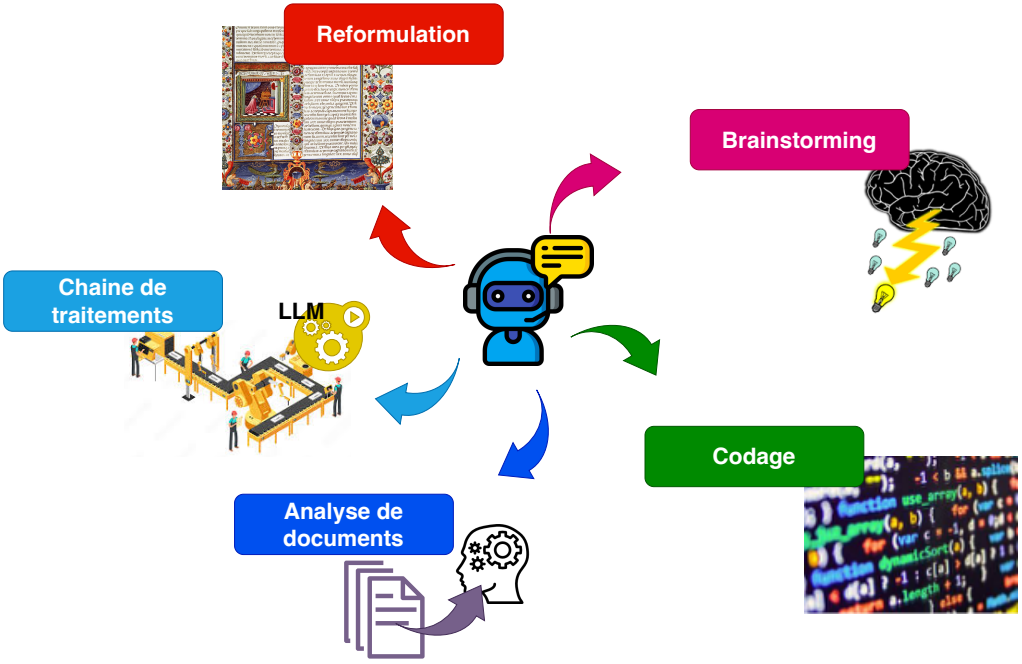
Spécialisation sur des tâches

Acquisition de:
Capacité de répondre à une question
Suivre un dialogue
Connaissances physiques
Bases de raisonnement

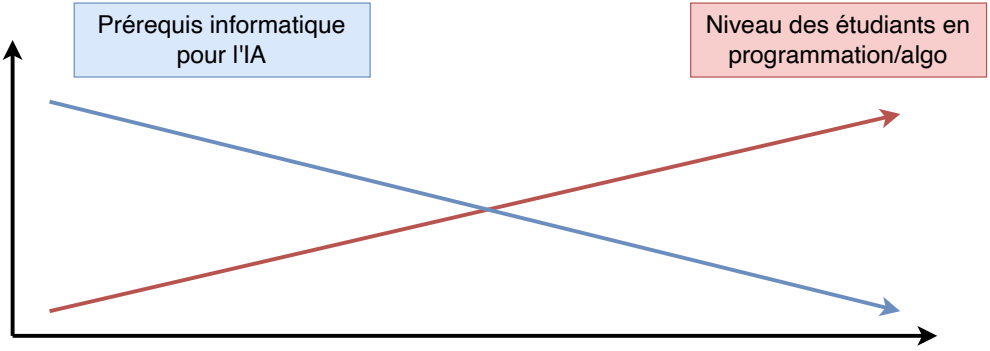
Les modèles de langue en 5 tableaux



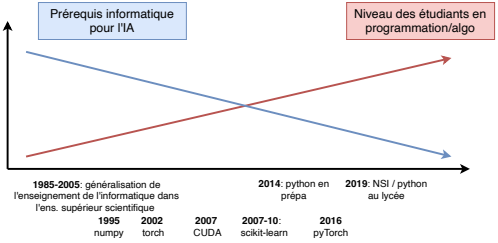
Les principaux usages en 5 tableaux



Accès à l'IA: la croisée des chemins



Accès à l'IA: la croisée des chemins



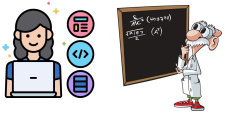
3 Niveaux d'accès à l'IA:

- Exploiter un chatbot... de manière **optimale** & **responsable**
- Utiliser les outils, manipuler des données
- Développer des outils

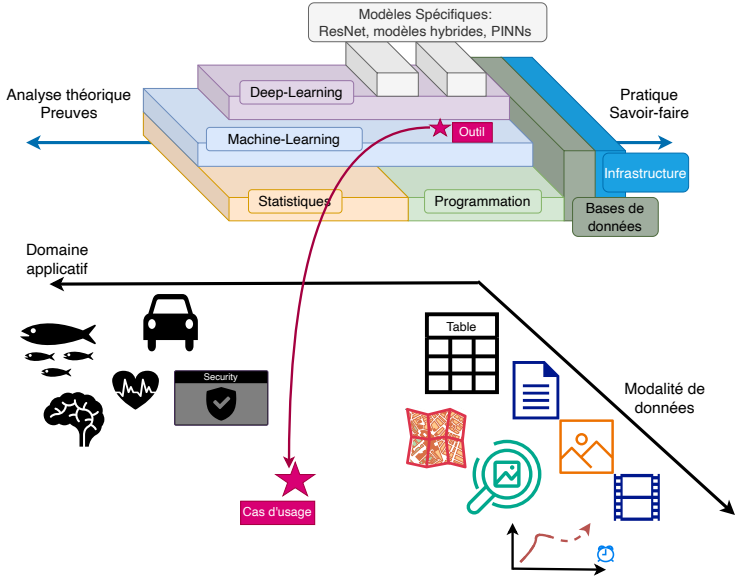
1

2

3



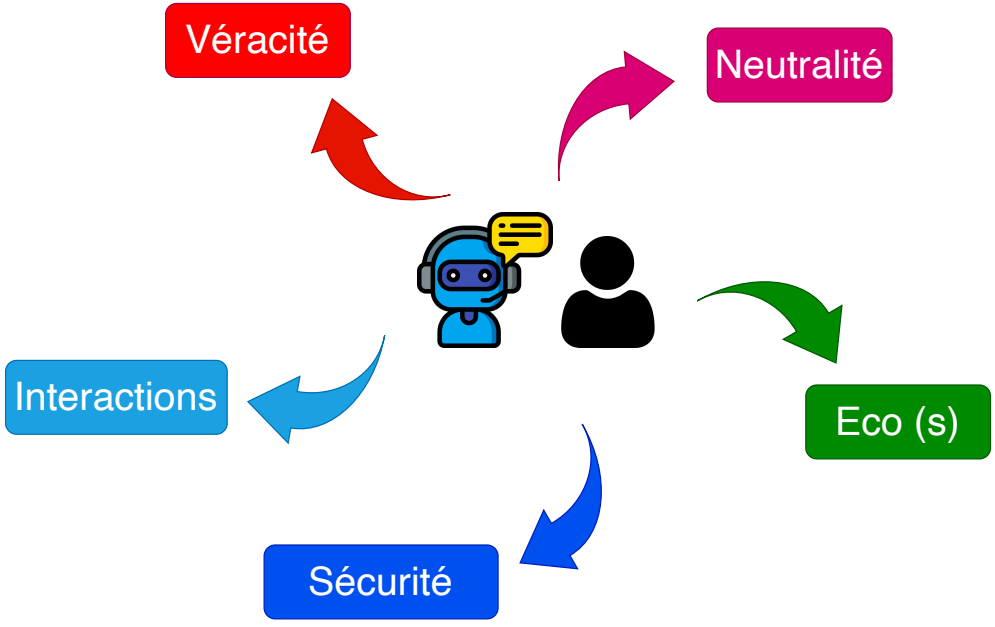
Palette de compétences



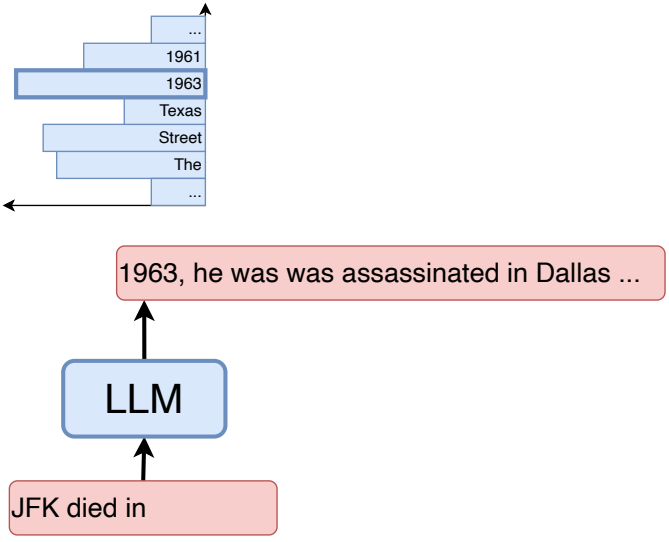
AGROPARISTECH
Institut des Sciences et Industries du Vivant et de l'Environnement

1A: python + stats
2A: mod. linéaires, machine learning
3A: Spécialisation IODAA, deep learning, texte, image, bio-info

Les défis de l'IA



Les défis de l'IA: la véracité



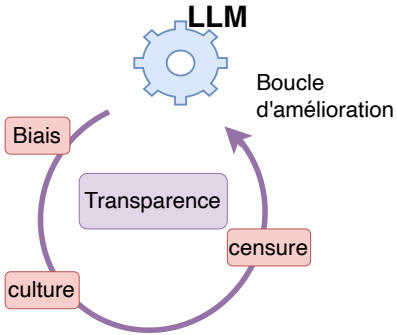
- Véracité $\neq$ vraisemblance
  $\Rightarrow$ hallucinations
- Confiance / filtrage
- Stabilité
- Evaluation

$\Leftrightarrow$ fréquence des informations:

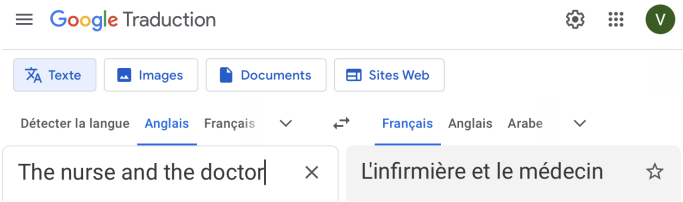
- grammaire
- fait historique
- évènement ponctuel
- ...

Les défis de l'IA: neutralité et transparence

- Sélection des données (sources, pondérations)
- Transformation des données (combinaison, filtrage)
- Architecture (coûts, optim., contraintes)
- Post-traitement (Alignement des LLMs)



Les biais des données sont amplifiés par les algos



La censure / alignement... pose aussi des problèmes

Les défis de l'IA: écologie / économie

Ecologie

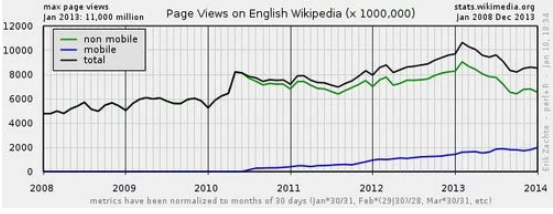
- Requête LLM = 10/(100x) une requête Google
- Ratio coûts *vs* gains
- Risque du techno-solutionisme (agriculture de précision, puit carbone, ...)

Economie

- Quel coût/souveraineté/perrenité pour les projets en IA?
- IA & fracture sociale: réduction ou augmentation?
- Quel modèle économique pour les producteurs de contenus?

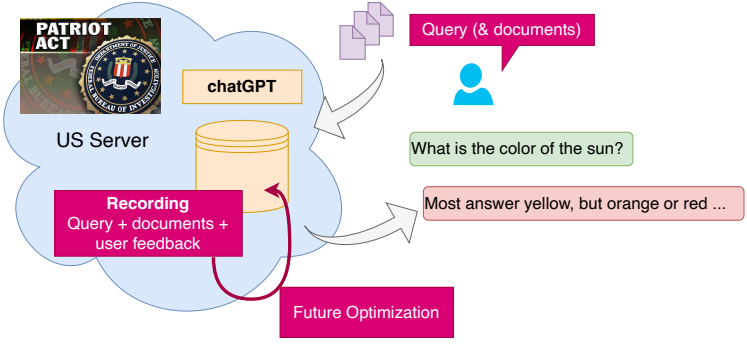
Google's Knowledge Graph Boxes: killing Wikipedia?

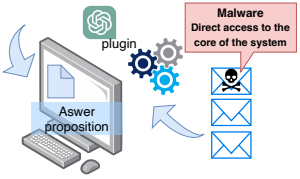
by Gregory Kohs

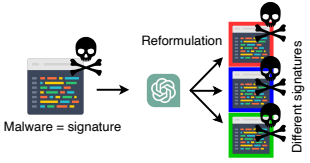


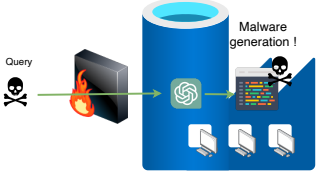
Les défis de l'IA: la sécurité et la vie privé

IA = (souvent) fuite de données

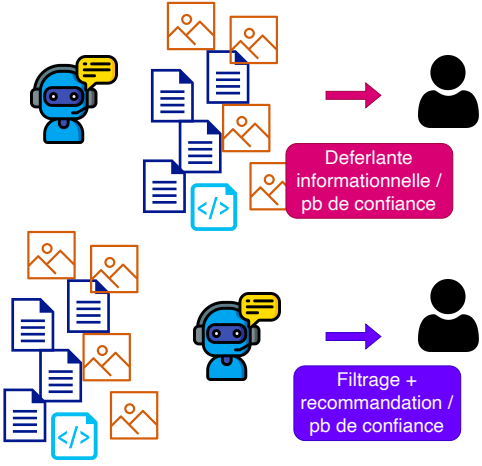








Les défis de l'IA: les interactions homme-machine



- Interface commerciale
- Interface RH
- Interface médicale
- Interface pédagogique
- Interface informationnelle

$\Rightarrow$ Des opportunités et des risques : biais, stigmatisation, inégalité de traitement, désinformation, harcellement

Des outils à double tranchant

L'humain augmenté... Mais au profit de qui?
Quid du remplacement de l'homme par la machine?

Comment aborder la question de l'éthique ?

Médecine

1. **Autonomie :** le patient doit pouvoir prendre des décisions éclairées.
2. **Bienfaisance :** obligation de faire le bien, dans l'intérêt des patients.
3. **Non-malfaisance :** éviter de causer du tort, évaluer les risques et les bénéfices.
4. **Justice :** équité dans la distribution des ressources et des soins de santé.
5. **Confidentialité :** confidentialité des informations des patients.
6. **Vérité et transparence :** fournir une information honnête, complète et compréhensible.
7. **Consentement éclairé :** obtenir le consentement libre et éclairé des patients.
8. **Respect de la dignité humaine :** traiter tous les patients avec respect et dignité.

Intelligence Artificielle

1. **Autonomie :** les humains contrôlent le processus
2. **Bienfaisance :** dans l'intérêt de qui ? Utilisateur + GAFAM...
3. **Non-malfaisance :** humains + environnement / durabilité / usages malveillants
4. **Justice :** accès à l'IA et égalité des chances
5. **Confidentialité :** qu'en est-il du modèle économique de Google/Facebook ?
6. **Vérité et transparence :** la tragédie de l'IA moderne
7. **Consentement éclairé :** des cookies aux algorithmes, savoir quand on interagit avec une IA
8. **Respect de la dignité humaine :** Alignement / censure nécessaire mais qui pose des questions

Comment aborder la question de l'éthique ?

Médecine

1. **Autonomie :** le patient doit pouvoir prendre des décisions éclairées.
2. **Bienfaisance :** obligation de faire le bien, dans l'intérêt des patients.
3. **Non-malfaisance :** éviter de causer du tort, évaluer les risques et les bénéfices.
4. **Justice :** équité dans la distribution des ressources et des soins de santé.
5. **Confidentialité :** confidentialité des informations des patients.
6. **Vérité et transparence :** fournir une information honnête, complète et compréhensible.
7. **Consentement éclairé :** obtenir le consentement libre et éclairé des patients.
8. **Respect de la dignité humaine :** traiter tous les patients avec respect et dignité.

Intelligence Artificielle

1. **Autonomie :** les humains contrôlent le processus
2. **Bienfaisance :** dans l'intérêt de qui ? Utilisateur + GAFAM...
3. **Non-malfaisance :** humains + environnement / durabilité / usages malveillants
4. **Justice :** accès à l'IA et égalité des chances
5. **Confidentialité :** qu'en est-il du modèle économique de Google/Facebook ?
6. **Vérité et transparence :** la tragédie de l'IA moderne
7. **Consentement éclairé :** des cookies aux algorithmes, savoir quand on interagit avec une IA
8. **Respect de la dignité humaine :** Alignement / censure nécessaire mais qui pose des questions

Chaine des compétences et de la souvenraineté

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