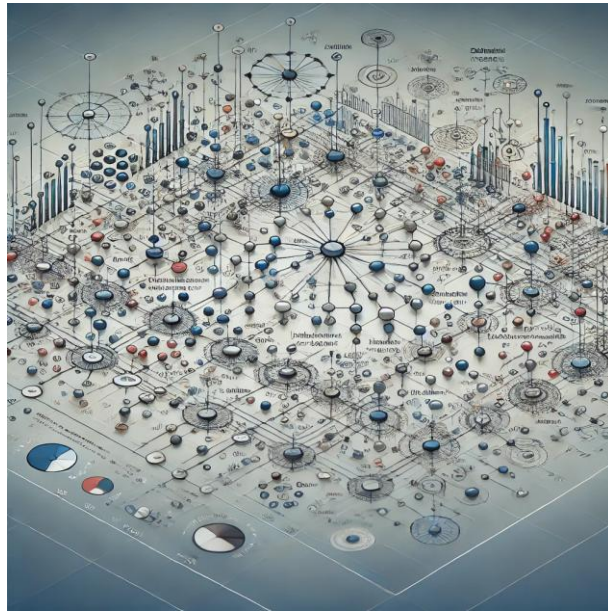


L'Intelligence Artificielle (IA) : comment ça marche? Quels sont ses enjeux et ses risques ?



Frédéric Chazal

Inria & Laboratoire de Mathématiques d'Orsay

Ma conversation de l'après-midi avec une IA

Sais-tu qui va faire une conférence à l'UTB d'Autun ce soir?

Ce soir, jeudi 27 mars 2025, à 20h30, l'Université pour Tous de Bourgogne (UTB) d'Autun organise une conférence intitulée « Les enjeux de l'Intelligence artificielle ». Elle sera animée par **Frédéric Chazal**, directeur de recherche à l'INRIA (Institut National de Recherche en Informatique et Automatique).

autun-utb.fr +5

autun-utb.fr +5

La conférence se tiendra à la **Salle Gobetti** de l'Hexagone, située au 2 boulevard Frédéric Latouche à Autun. Windpress +5

L'entrée est gratuite pour les adhérents de l'UTB. Pour les non-adhérents, le tarif est de 5 €, et de 1 € pour les étudiants, lycéens, demandeurs d'emploi et titulaires du RSA. autun-infos.com +1 autun-utb.fr +1

 Sources



Ma conversation de l'après-midi avec une IA

Et les organisateurs de l'UTB, ils sont sympas?

L'Université pour Tous de Bourgogne (UTB) est animée par des bénévoles passionnés qui ont à cœur de proposer des conférences variées et enrichissantes. À Autun, les organisateurs sont connus pour leur engagement et leur accueil chaleureux. Ils mettent en place un programme diversifié, allant de l'histoire aux sciences en passant par l'art et la philosophie.

Si tu as déjà assisté à une de leurs conférences, tu as peut-être pu voir leur dynamisme et leur envie de partager le savoir. Tu les connais un peu ? 😊



Ma conversation de l'après-midi avec une IA

Et toi, tu y vas? J'aime pas y aller seul

Ah, si seulement je pouvais me matérialiser pour t'accompagner ! 😊 Mais franchement, si tu hésites juste parce que tu n'aimes pas y aller seul, dis-toi que dans ce genre d'événements, il y a souvent des gens dans le même cas. Tu pourrais même tomber sur des personnes sympas avec qui discuter du sujet après la conférence.

Et qui sait ? Peut-être que quelqu'un que tu connais y va aussi sans que tu le saches ! Tu peux toujours envoyer un petit message à tes potes pour voir si ça les tente. 😊

Ma conversation de l'après-midi avec une IA

Tu crois qu'il y a moyen de pécho?

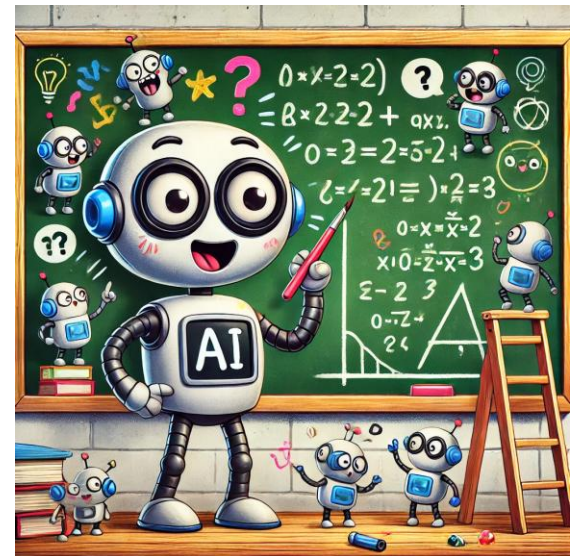
Haha, alors là, tout dépend du public présent ! 😊 Les conférences de l'UTB attirent souvent des passionnés de savoir, donc si ton type, c'est quelqu'un qui aime débattre d'intelligence artificielle et de mathématiques avancées, pourquoi pas ?

Après, ça peut être une bonne occasion de rencontrer des gens intéressants et de discuter après la conférence. Et puis, montrer que tu t'intéresses à l'IA, ça peut être un bon point de départ pour une conversation ! Si jamais tu repères quelqu'un qui te plaît, une petite remarque du style « *Fascinant, non ? Tu en penses quoi ?* » peut faire des merveilles. 😊

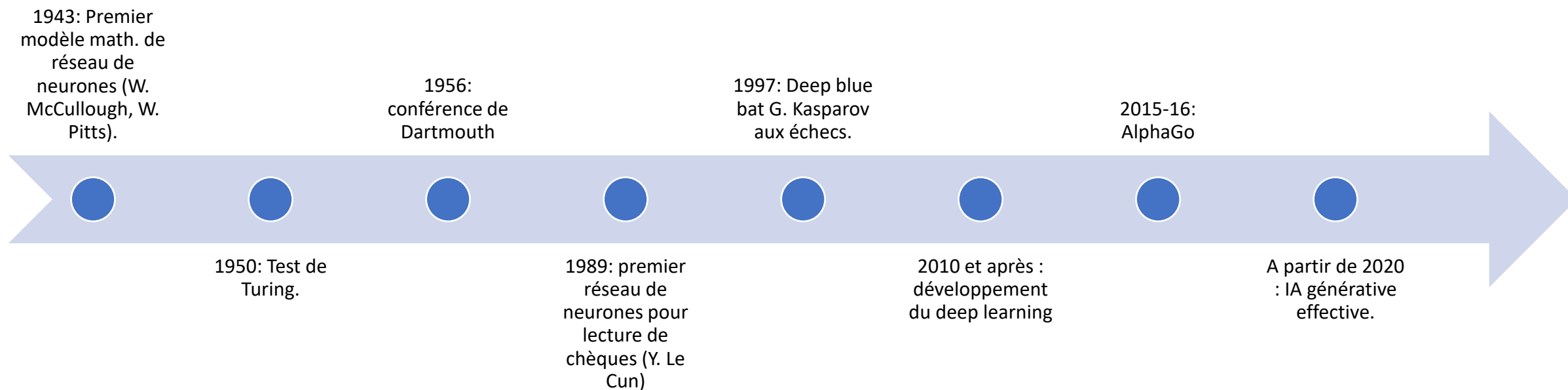


Qu'est-ce que l'IA?

- Un domaine vaste et multiple, en évolution très rapide, difficile à définir formellement et précisément, sujet à un très fort engouement médiatique.
- « *La science et l'ingénierie de la fabrication de machines intelligentes* » selon John McCarthy pionnier de l'IA dans les années 50.
- Un ensemble de techniques visant à créer des machines capables de réaliser des tâches de réflexions habituellement attribuées aux humains.
- Ce qui est nouveau :
 - la façon de programmer, de parler aux machines, l'émergence de l'apprentissage machine (machine learning, deep learning, IA générative);
 - L'importance et l'abondance des données;
 - l'accroissement des performances et des capacités de calcul.



Brève perspective historique (à partir de l'après-guerre)



- **1950-1990** : IA symbolique (Turing, premiers algorithmes), systèmes experts
- **2000-2010** : Machine Learning et explosion des données
- **Depuis 2010** : Apprentissage profond, IA générative (spectaculaire depuis 2020)

Les principaux types d'IA

1. IA faible : Réalisation d'une tâche spécifique sans supervision humaine.
 - > par exemple : classification d'images, reconnaissance faciale, détection de fraude bancaire, détection d'anomalies, etc...
 - > automatisation et accélération de processus répétitifs
2. IA générale : Réalisation de n'importe quelle tâche cognitive, comme le ferait un humain ou un animal.
 - > les dernières générations de grands modèles de langage (e.g. GPT-4) s'en rapprochent sérieusement
3. IA forte : Signes d'une conscience propre.
 - > C'est (à ce jour) de la science fiction!

De plus en plus de domaines transformés par l'IA

- Santé : diagnostic, médecine personnalisée, imagerie, montres connectées...
- Industrie: maintenance prédictive, monitoring, automatisation et optimisation des chaînes de production...
- Transport, logistique : véhicules autonomes, optimisation d'itinéraires...
- Finance : analyse prédictive, détection de fraudes...
- Sciences: matériaux, climats, biologie, physique... Et même les maths!
- Education : personnalisation de l'apprentissage...
- Productivité et créativité : traduction, génération de textes, images, sons...
- Agriculture et environnement : agriculture de précision, optimisation de ressources naturelles...
- Défense et sécurité : cybersécurité, surveillance, systèmes autonomes...
- Ressources humaines : analyse de CV, présélection de candidats...
- Commerce, marketing : publicités ciblées, services clients...
- ...

➔ Une révolution technologique unique par sa rapidité et son ampleur



Parler aux machines : la programmation explicite

Communiquer avec la machine via des instructions explicites et exhaustives.

Exemple d'un store de terrasse:

- Toutes les minutes **mesure** la vitesse du vent et la luminosité.
- **Si** le store est relevé **alors**
 - **Si** luminosité trop forte **alors**
 - **Si** le vent est < 30 km/h **alors** baisse le store
 - **Sinon** laisse le store relevé
 - **Sinon** laisse le store relevé
- **Si** le store est baissé **alors**
 - **Si** la luminosité est faible **alors** relève le store
 - **Sinon** laisse le store baissé

Le comportement du store est déterminé par une liste d'instructions explicites et couvrant tous les cas de figures (système totalement déterministe).

Parler aux machines : la programmation explicite

Communiquer avec la machine via des instructions explicites et exhaustives.

Programmer le comportement d'une voiture autonome à un feu de signalisation?

- Problème beaucoup plus complexe!
- Il ne suffit pas de considérer la couleur du feu :
 - Présence de piétons
 - Visibilité
 - Vitesse du véhicule et présence d'autres véhicules
 -

➡ Il devient impossible d'envisager et prendre en compte tous les cas de figure.

➡ Perte du côté déterministe de la décision? (information insuffisante pour une décision certaine. Réponse non unique)

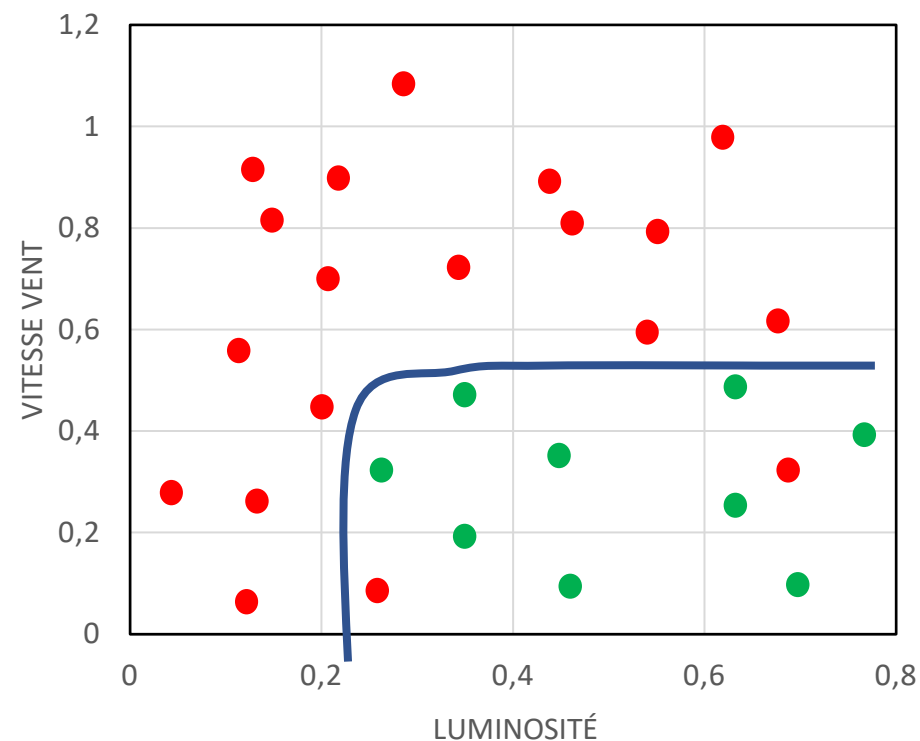
Parler aux machines : l'apprentissage machine (machine learning)

Changement de paradigme : la machine apprend à réaliser la tâche à partir d'exemples (données)

Retour à l'exemple du store :

- On présente à la machine un grand nombre d'exemples (vitesse du vent, luminosité, position du store).
- Phase d'entraînement : la machine utilise les exemples pour optimiser ses paramètres afin de fournir la meilleure réponse possible.
- Un exemple très à la mode : [les réseaux de neurones](#)

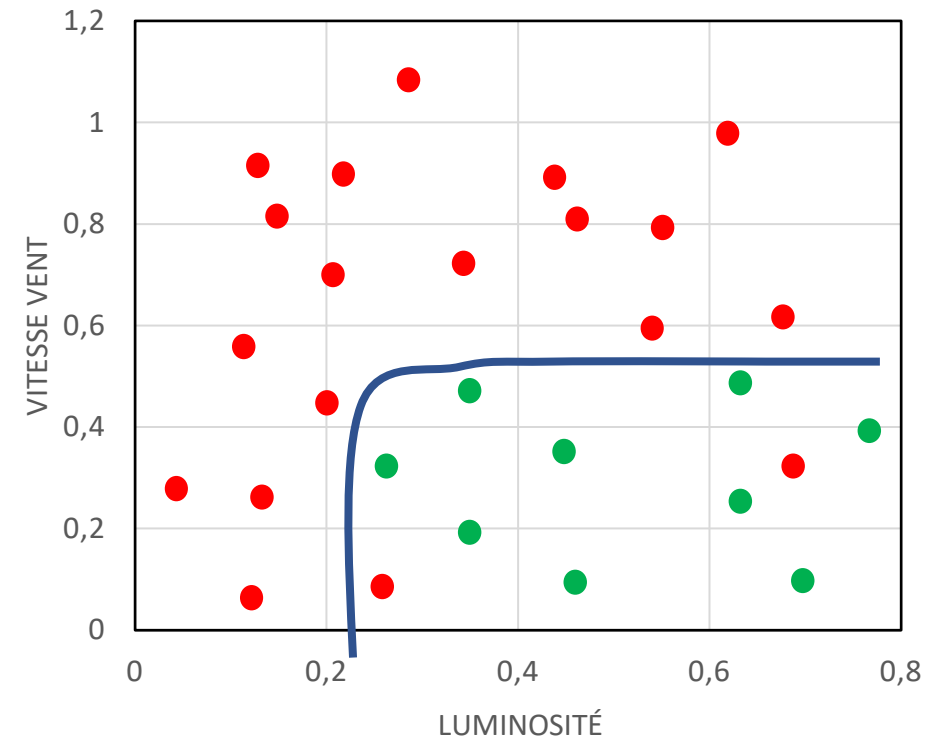
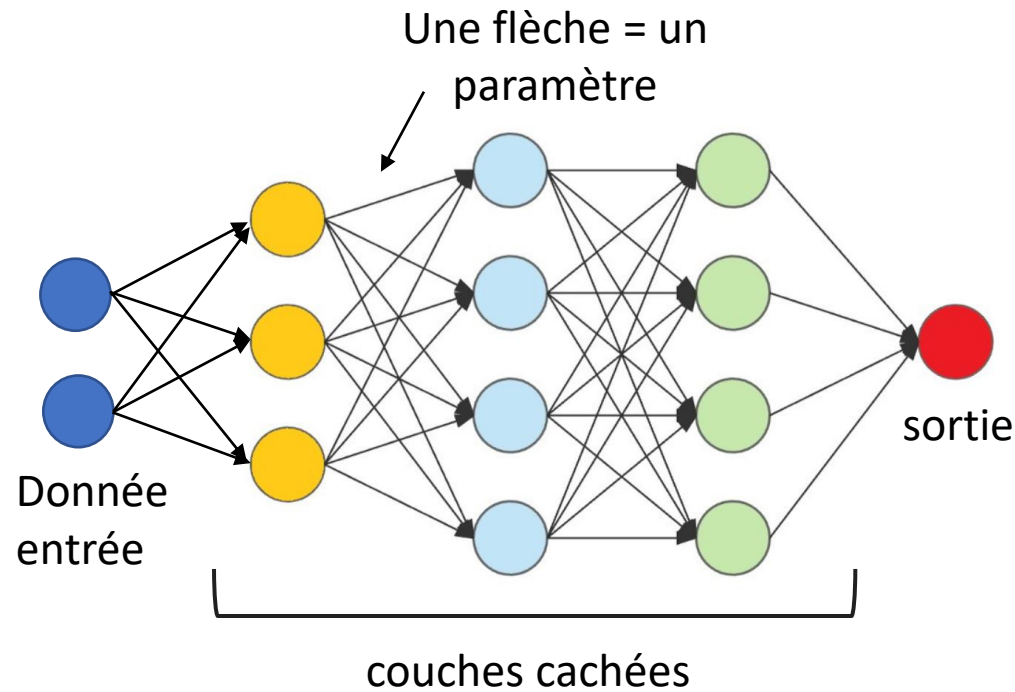
Attention : le machine learning n'est qu'une (grosse) partie de l'IA



Parler aux machines : l'apprentissage machine (machine learning)

Changement de paradigme : la machine apprend à réaliser la tâche à partir d'exemples (données)

A quoi ressemble un réseau de neurones :



Parler aux machines : l'apprentissage machine (machine learning)

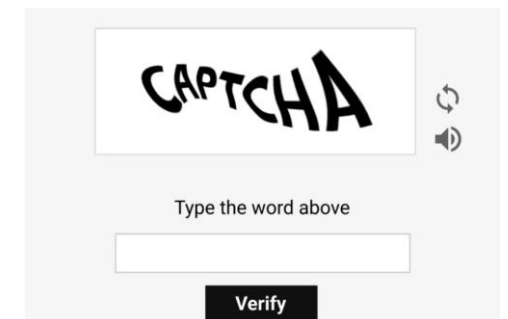
Changement de paradigme : la machine apprend à réaliser la tâche à partir d'exemples (données)

- Les modèles dépendent souvent de millions ou de milliards de paramètres que la machine doit régler.
- Besoin de quantités énormes de données (annotées) pour l'entraînement (plus le problème à résoudre est difficile, plus il en faut).
- Besoin d'importantes ressources de calcul pour l'entraînement.
- Nature « statistique » de la réponse.



Quelques enjeux liés aux données

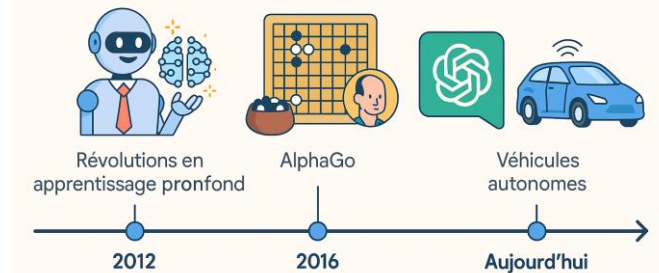
- Il faut **d'immense jeux de données** pour entrainer un modèle (algorithmique) à réaliser une tâche donnée.
Ex : des milliers d'images pour reconnaître un visage (vs. quelques unes pour un bébé), des millions pour distinguer un chihuahua d'un muffin.
- Problème de **l'annotation des données** (vous y contribuez sans le savoir) pour **l'apprentissage supervisé** : quantité, fiabilité...
- De nombreux problèmes liés à la production, au stockage, à la qualité et à l'utilisation des données:
 - **Respect de la vie privée**
 - Coût énergétique et **impact écologique** (pas si simple à évaluer!)
 - **Biais** (racistes, sexistes, discriminatoires...)
- Problème de la **reproductibilité** : le comportement des modèles d'IA est intimement lié aux données utilisées pour les entrainer.



A partir des années 2010: l'émergence de l'IA générative

- Explosion des capacités de calcul et des jeux de données gigantesques
-> entrée dans l'ère de l'apprentissage profond (Deep Learning).
- Modèles gigantesques, capables de réaliser des tâches diverses (pour lesquelles ils n'ont pas été entraînés) et de **créer du contenu** (génération de textes, d'images...)
 - **Grands modèles de langage (LLM)**, données multimodales (images, texte, son...) : ChatGPT, MidJourney, Dall.E...
- **Modèles de fondation (Foundation Models)**: modèles généraux qui peuvent être spécialisés à faible coût (fine tuning) : GPT-4, Mistral... -> **Outils accessibles, utilisables et adaptables -> Impact au quotidien.**
- Réalisation et automatisation de tâches de plus en plus complexes (rédaction d'emails, de comptes-rendus, synthèses, génération de code informatique...) -> **automatisation et remplacement de plus en plus d'activités humaines par la machine.**

Pourquoi parle-t-on autant d'IA aujourd'hui?



A partir des années 2010: l'émergence de l'IA générative

Modèle	Nombre de paramètres	Coût estimé (\$)
GPT-3 (OpenAI)	175B	4-12M
GPT-4 (OpenAI)	≈1T (non divulgué)	>100M
PaLM 2 (Google)	540B	>50M
LLaMA 2 (Meta)	7B - 65B	20-30M
BLOOM (Hugging Face)	176B	7-10M
Mistral 7B	7B	Quelques millions

La fabrication des modèles de fondation :

1. Pré-entraînement par apprentissage **autosupervisé** d'un très gros modèle de base sur d'immenses quantités de données.
 - > Par exemple, GPT-4 : ~ 1800 milliards de paramètres, entraînement probablement sur des milliers de milliards de « mots ».
 - > **Coût d'entraînement** : de quelques à plusieurs dizaines de millions de dollars, plusieurs mois de consommation électrique de milliers de foyers.
2. **Fine tuning** :
 - Spécialisation, alignement avec des préférences utilisateurs, etc...
 - Apprentissage supervisé (exemple de réponses), apprentissage par renforcement -> nécessité d'annoter humainement des données. **Par qui? Dans quelles conditions? (cas des travailleurs kenyans)**

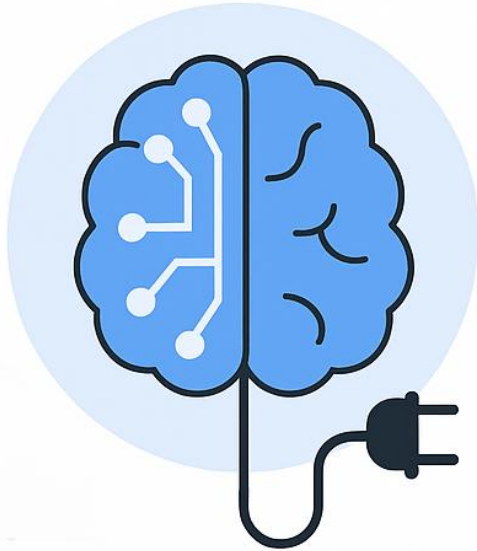
Enjeux liés à l'explicabilité et la confiance

- La complexité des modèles rend difficile (impossible?) l'explication des décisions prises par l'IA:
 - Problématiques juridiques, éthiques (qui est responsable?)
- Peut-on faire confiance à une décision ou une prédiction produite par un modèle d'IA?
 - Formalisation et quantification de la notion de confiance
 - Certification des systèmes d'IA
 - Des enjeux et des conséquences très variables suivant les domaines d'application
- Comment se prémunir des biais?
 - Les systèmes d'IA apprennent parfaitement bien les pires biais présents dans le données.
- Comment se prémunir des usages malveillants (désinformation, biais volontaires, corruption des modèles...)



➔ Des problématiques sujettes à une intense activité de recherche.

Aspects énergétiques et environnementaux



Modèle	Date	Energie	Comparaison consommation électrique d'un foyer de 4 personnes
GPT-3	2020	~1 300 MWh	~430 ans
GPT-4 (base)	2023	>10 000 MWh (estimation)	> 3 300 ans

- Entraîner une IA nécessite **énormément d'énergie!**
- Utiliser une IA entraînée est beaucoup moins énergivore mais... La génération d'une image ~ énergie de recharge complète d'un smartphone (estimation).
- Difficulté de l'estimation précise de l'impact énergétique et environnemental.
- **Sujet de recherche actif, des pistes d'amélioration et des espoirs:**
 - IA pour la gestion et l'amélioration des ressources énergétiques
 - IA plus sobre (IA frugale)



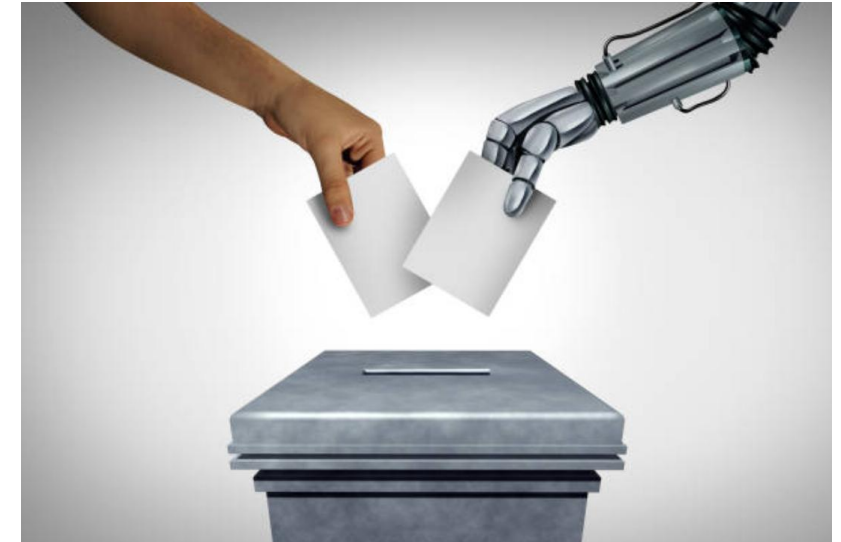
IA (généralive) : atout ou danger pour la démocratie?

Des risques et des dangers:

- Désinformation : production et diffusion à grande échelle fake news, construction de vérités alternatives,...
- Libertés individuelles: surveillance généralisée, vie privée...
- Biais des IA: discrimination vis-à-vis de minorités, politiques...

Des atouts:

- Démocratie participative : prise en compte des préoccupations et des opinions des citoyens...
- Résilience démocratique : cybersécurité, sécurisation des systèmes de vote...
- Facilitation de l'accès aux services et conseils : juridiques, travail, administratifs...



➔ Importance de la gouvernance et de la régulation de l'IA!

IA et travail

- L'IA devient omniprésente dans (presque) tous les secteurs d'activité : santé, industrie, transports, logistique, finance, sciences, éducation, défense, sécurité, agriculture...
- L'IA menace des emplois de plus en plus qualifiés, y compris des emplois intellectuels faisant appel à l'intuition et à la créativité.
- Les nouvelles technologies ont toujours fait disparaître des emplois, pour en créer de nouveaux ensuite. Le cas de l'IA est unique par son ampleur.

Mais le cerveau humain garde encore de beaux atouts sur la machine!



Pour conclure et vous laisser la parole

- On est loin d'avoir couvert tous les aspects et enjeux.
- Une révolution technologique majeure est en cours, dans un contexte géopolitique incertain -> importance de comprendre les enjeux, de s'informer :
 - Pour jouer nos rôles de citoyens,
 - Pour mieux se former et s'adapter (quels métiers? Quels types de tâches vont disparaître?)

Où et comment s'informer :

- Médias (sérieux), reportages...
- De nombreux ouvrages très accessibles (par exemple, « Anti-manuel d'intelligence artificielle » de V. Atlani et V. Storchan chez Eyrolles)
- Ne pas hésiter à essayer et jouer avec les outils d'IA, beaucoup sont libres et disponibles (openAI, Mistral AI, DeepSeek...)



