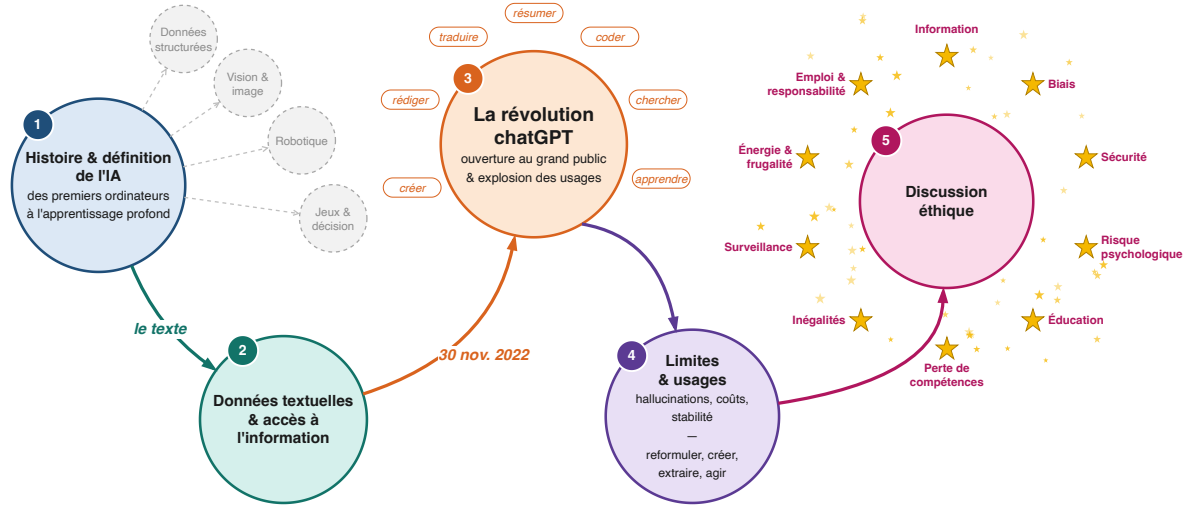
PARTIE 1 :
DES MODÈLES DE LANGUE À LEURS USAGES

Les enjeux de l'intelligence artificielle dans le monde du travail
16 novembre 2026, Institut des sciences sociales du travail

Vincent Guigue
vincent.guigue@agroparistech.fr
https://vguigue.github.io

Plan de la présentation



Partie 1 — Modèles & usages

Partie 2 — Risques & enjeux

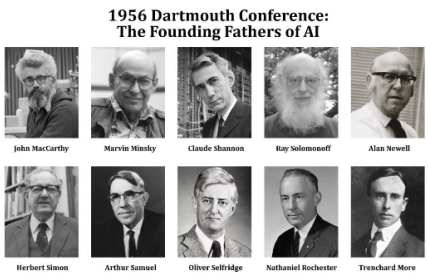
Introduction : une histoire de l'Intelligence Artificielle

Un rapide tour historique de l'Intelligence Artificielle

Naissance de l'informatique... Et de l'Intelligence Artificielle



A. Turing



Dartmouth conference creation of IA



Y. Lecun
Automated cheque reading



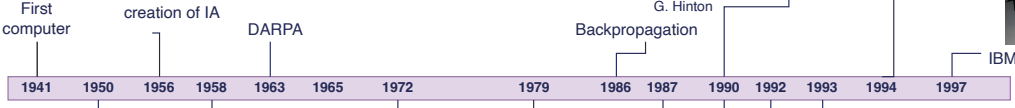
E. Dickmanns : 1000km in autonomous vehicle



G. Hinton

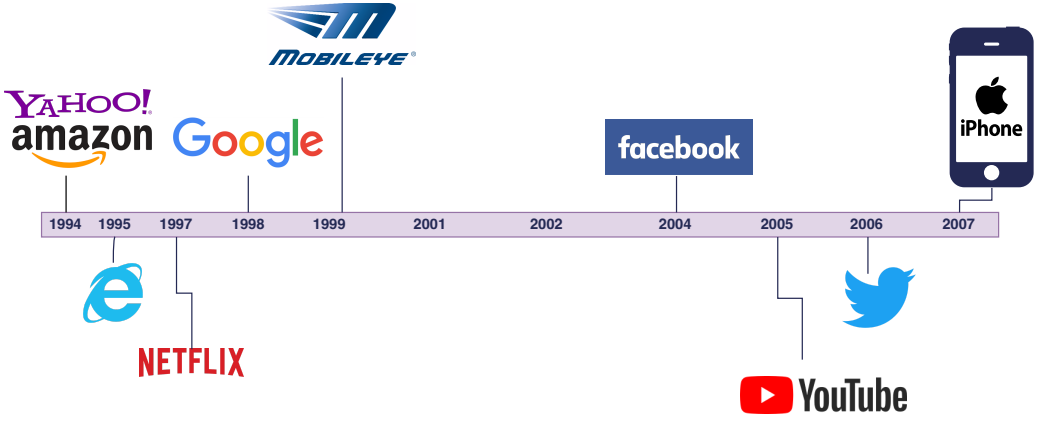


IBM Deeper Blue



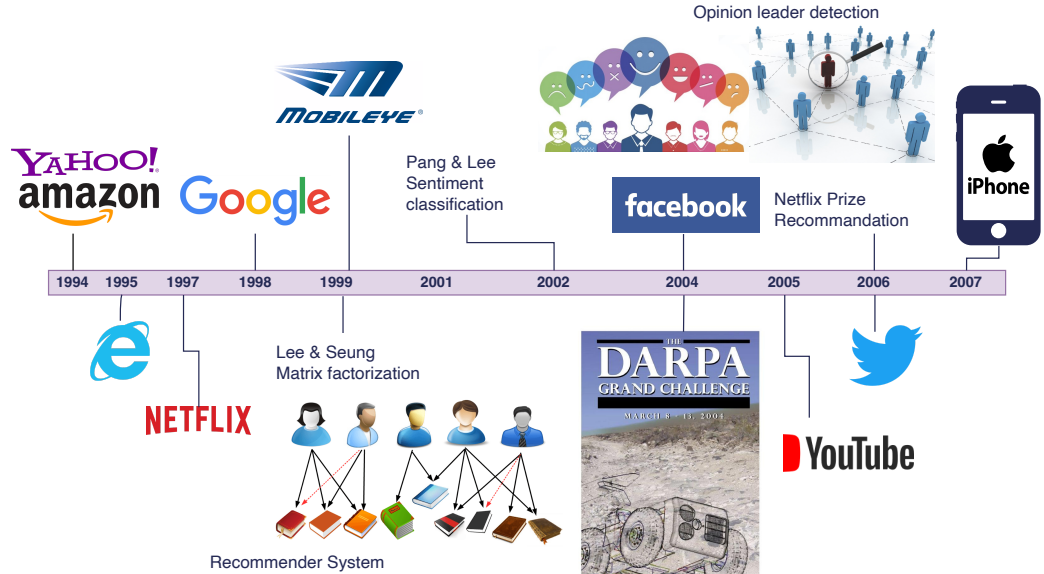
Un rapide tour historique de l'Intelligence Artificielle

Emergence (ou refondation) des GAFAM/GAMMA



Un rapide tour historique de l'Intelligence Artificielle

Emergence (ou refondation) des GAFAM/GAMMA



Un rapide tour historique de l'Intelligence Artificielle

Formation d'une vague de l'Intelligence Artificielle



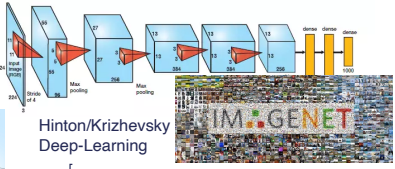
Thrun:
DARPA Gd Challenge
victory





Google
car





Hinton/Krizhevsky
Deep-Learning

K. Cho
Traduction auto.

Translate (v2)

AlphaGo

ChatGPT

2005 2007 2008 2009 2010 2011 2012 2014 2015 2016 2017 2020 2023

Y. Koren
Challenge
Netflix



kaggle

SIRI

amazon alexa

Google
DeepMind
Acquisition : $400M



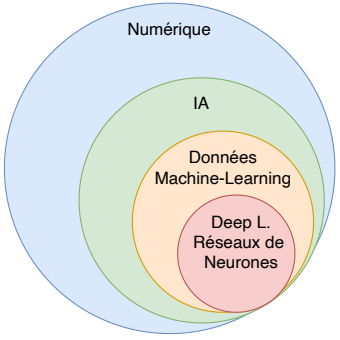
IBM Jeopardy win

MOBILEYE
An intel company
Acquisition :
$15B

AlphaFold

OpenAI
DALL·E 2

Artificial Intelligence & Machine Learning



Input (**X**)	Output (**Y**)	Application
email →	spam? (0/1)	spam filtering
audio →	text transcript	speech recognition
English →	Chinese	machine translation
ad, user info →	click? (0/1)	online advertising
image, radar info →	position of other cars	self-driving car
image of phone →	defect? (0/1)	visual inspection

IA : programmes informatiques qui s'adonnent à des tâches qui sont, pour l'instant, accomplies de façon plus satisfaisante par des êtres humains car elles demandent des processus mentaux de haut niveau.

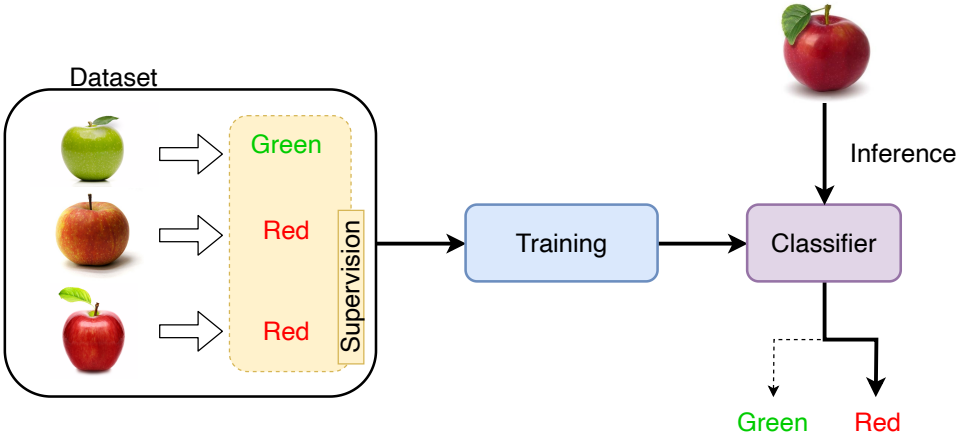
Marvin Lee Minsky, 1956

N-AI (Narrow Artificial Intelligence), dédiée à une tâche unique

$\neq$ **IA-G (IA Générale)**, qui remplace l'humain dans les systèmes complexes.

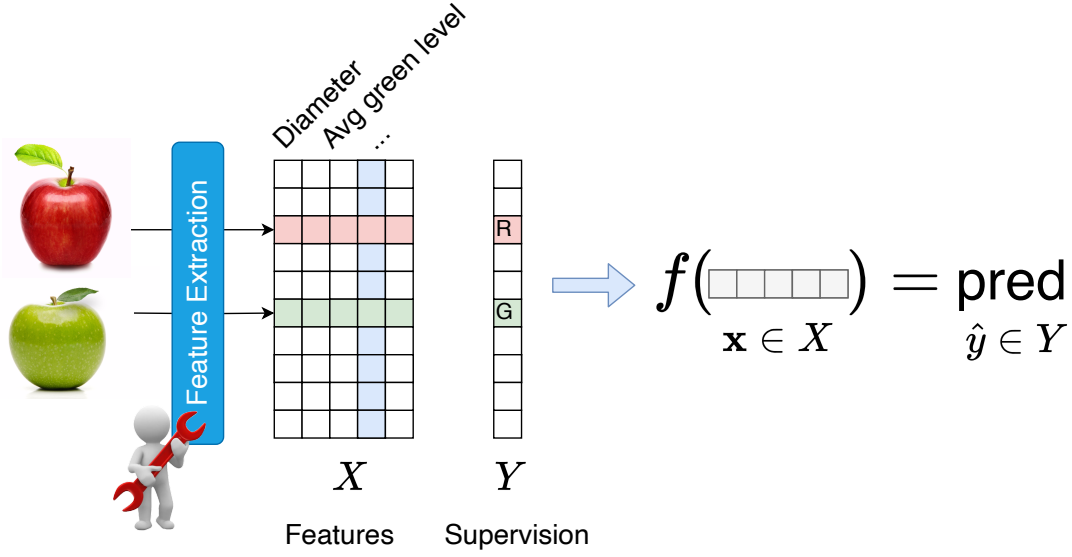
Andrew Ng, 2015

Chaîne de Traitement Supervisé & Modèles

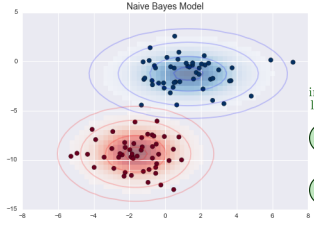


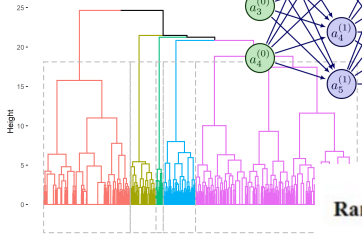
- Promesse = construire un modèle *uniquement* à partir d'observations

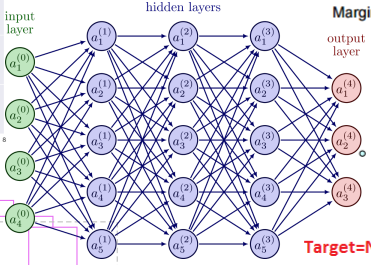
Chaîne de Traitement Supervisé & Modèles

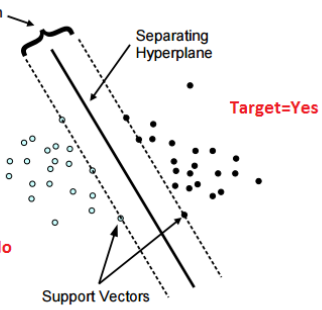


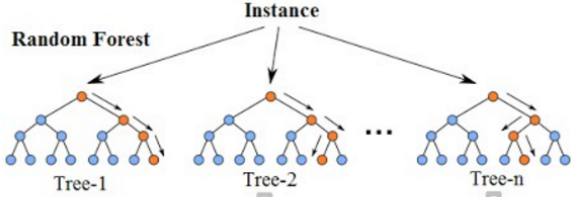
Chaîne de Traitement Supervisé & Modèles



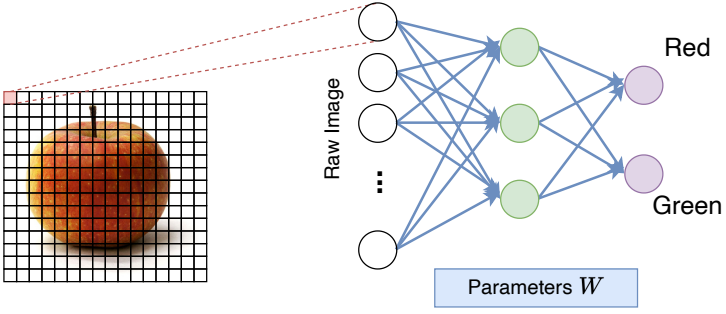






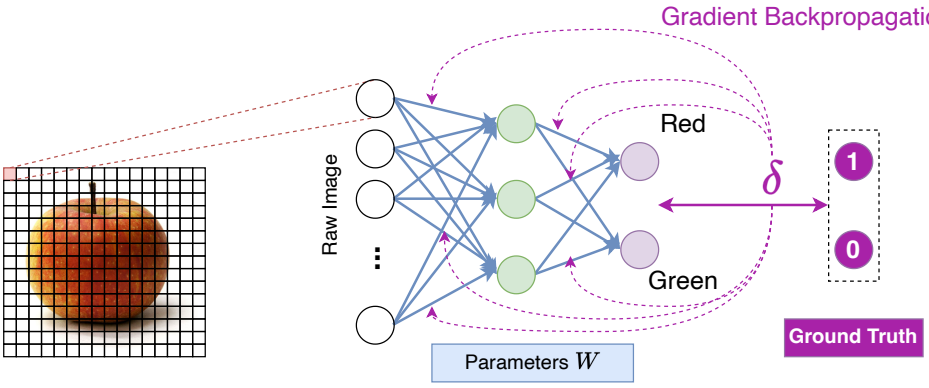


Chaîne de Traitement Supervisé & Modèles



- Initialisation aléatoire... Et prise de décision aléatoire (au début !)

Chaîne de Traitement Supervisé & Modèles

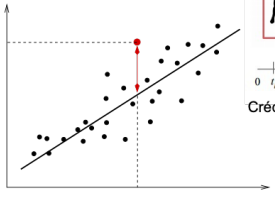


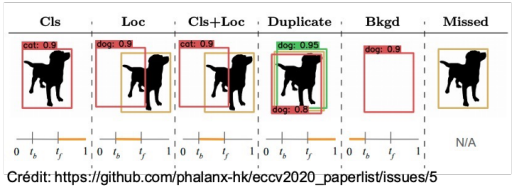
- Mise à jour des poids
- Pas de taille epsilon, nombreuses itérations sur les données

Mesurer les performances des modèles

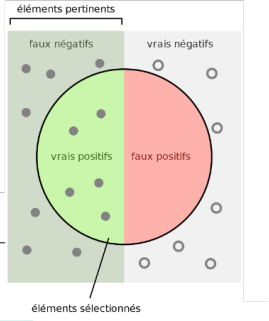
Estimer la performance en généralisation...

Est aussi important que de construire le modèle



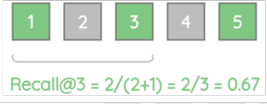


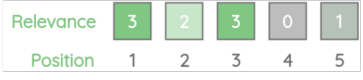
Crédit: https://github.com/phalanx-hk/eccv2020_paperlist/issues/5

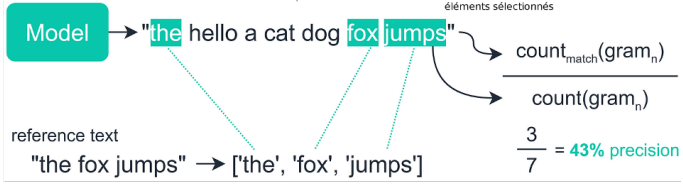


Précision = 

Rappel = 

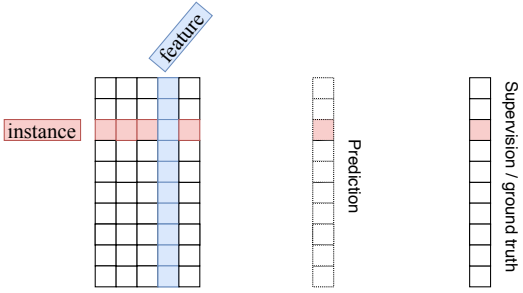






Mesurer les performances des modèles

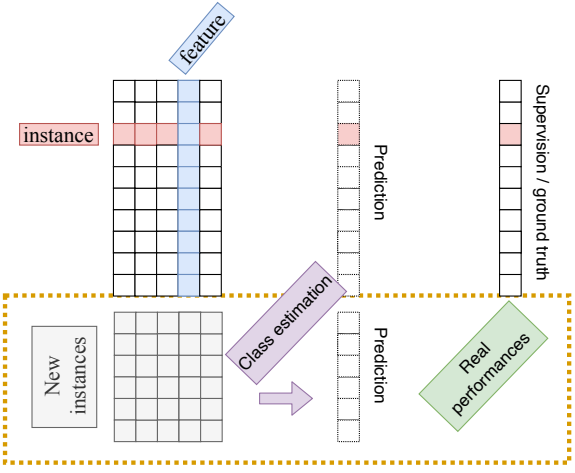
Estimer la performance en généralisation...
Est aussi important que de construire le modèle



Mesurer les performances des modèles

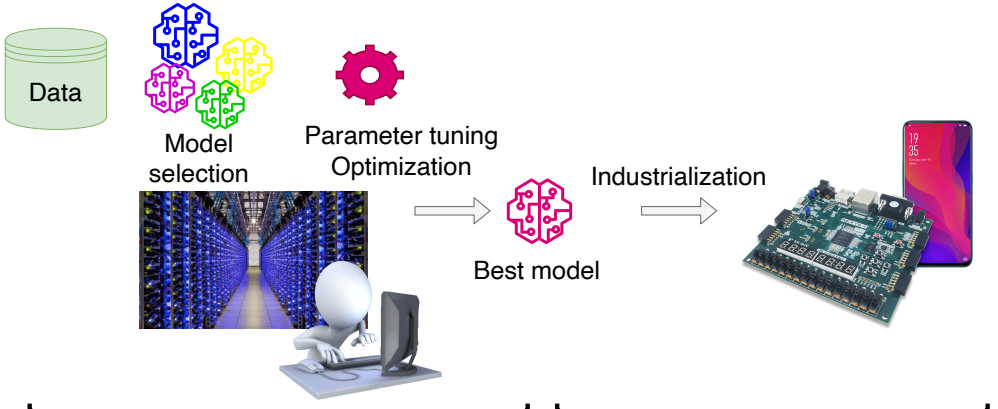
Estimer la performance en généralisation...

Est aussi important que de construire le modèle



Apprentissage/Construction *vs* Déployement/Exploitation

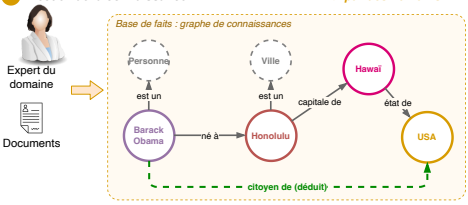
Différentes étapes en apprentissage automatique



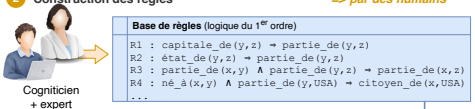
La place de l'humain dans une décision automatique

1. Les systèmes experts / IA Symbolique

1 Recueil de la connaissance *=> par des humains*



2 Construction des règles *=> par des humains*



3 Décision automatisée par la logique

faits + règles

Nouveau fait : citoyen_de(Obama, USA)
Chaînage avant : né_à + R1, R2 → R3 → R4

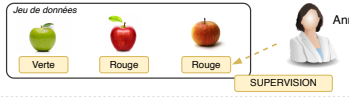
Moteur d'inférence

✔ *décision explicable & traçable*
✘ *connaissances coûteuses à recueillir, fragiles à maintenir*

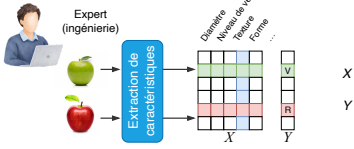
Humain : connaissances + règles | **Machine** : déduction

2. L'apprentissage automatique / machine-learning

1 Collecte et annotation des données *=> par des humains*

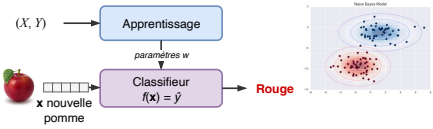


2 Conception des caractéristiques *=> par l'expert*



X : **Description** des données
Y : **Etiquettes** (supervision)

3 Apprentissage automatique de la décision



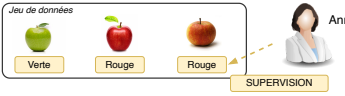
✔ *moins cher à développer et mettre à jour*
✘ *moins de transparence et de garanties*

Humain : données, annotations, caractéristiques | **Machine** : paramètres

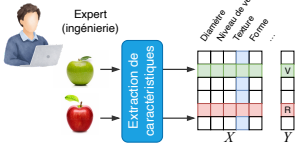
La place de l'humain dans une décision automatique

2. L'apprentissage automatique / machine-learning

1 Collecte et annotation des données => *par des humains*

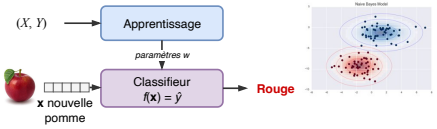


2 Conception des caractéristiques => par l'expert



X : **Description** des données
Y : **Etiquettes** (supervision)

3 Apprentissage automatique de la décision



✔ *moins cher à développer et mettre à jour*
✘ *moins de transparence et de garanties*

Humain : données, annotations, caractéristiques | **Machine** : paramètres

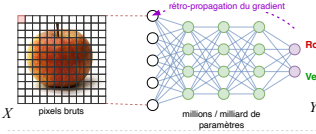
3. L'apprentissage de représentation / deep-learning

1 Beaucoup plus de données... et de calcul



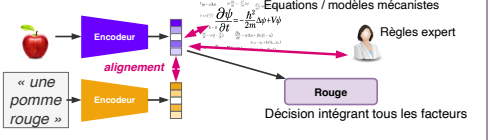
2 Des opérateurs complexes pour des données brutes

Convolutions | Récurrences | Attention (Transformers)



$f(x)$: fonction très complexe
Extraction automatique des caractéristiques

3 Systèmes flexibles / expertise / multimodalité



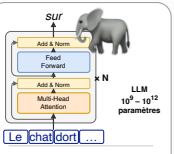
✔ *décision complexe multifactorielle très robuste*
✘ *(beaucoup) moins de transparence*

Humain : annotation | **Machine** : représentations + décision

Panorama (non exhaustif) de l'IA

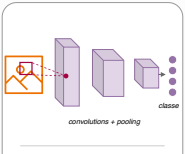
① Modalités & architectures

Texte



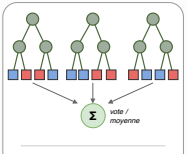
Transformer

Images



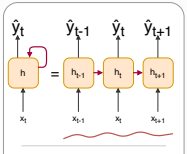
Réseau convolutif (CNN)

Données tabulaires



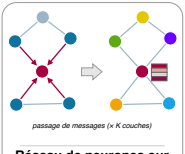
Forêts d'arbres

Séries temporelles



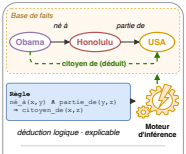
Réseau récurrent (RNN)

Graphes



Réseau de neurones sur graphes

Expertise



Systèmes experts

② Applications

Santé
Imagerie médicale, aide au diagnostic, dossiers patients

Agriculture de précision
Drones & satellites, capteurs, élevage

IoT
Flux de capteurs, détection d'anomalies, IA embarquée

Maintenance prédictive
Prévision de pannes, durée de vie résiduelle

Recommandation
E-commerce, streaming, filtrage collaboratif

Accès à l'information
Recherche, réseaux sociaux, veille

Codage & rédaction
Assistants de code, rédaction de documents : tous les métiers

Question-réponse & assistants
Chatbots, RAG, résumé, traduction

Détection de fraude
Transactions bancaires, cybersécurité

Véhicule autonome & robotique
Perception, détection d'objets, contrôle

Chimie & médicament
Molécules = graphes, protéines, matériaux

Énergie & finance
Prévision de consommation, marchés, risque

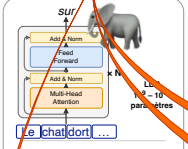
Contrôle qualité industriel
Détection de défauts par vision

Science
Des maths aux plans d'expériences : preuves, simulation, découverte

Panorama (non exhaustif) de l'IA

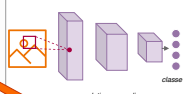
① Modalités & architectures

Texte



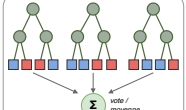
Transformer

Images



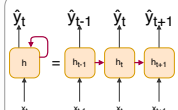
Réseau convolutif (CNN)

Données tabulaires



Forêts d'arbres

Séries temporelles



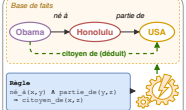
Réseau récurrent (RNN)

Graphes



Réseau de neurones sur graphes

Expertise



Systèmes experts

② Applications

Santé
Imagerie médicale, aide au diagnostic, dossiers patients

Agriculture de précision
Drones & satellites, capteurs, élevage

IoT
Flux de capteurs, détection d'anomalies, IA embarquée

Maintenance prédictive
Prévision de pannes, durée de vie résiduelle

E-commerce, streaming, filtrage collaboratif

Accès à l'information
Recherche, réseaux sociaux, veille

Codage & rédaction
Assistants de code, rédaction de documents : tous les métiers

Question-réponse & assistants
Chatbots, RAG, résumé, traduction

Détection de fraude
Transactions bancaires, cybersécurité

Véhicule autonome & robotique
Perception, détection d'objets, contrôle

Chimie & médicament
Molécules = graphes, protéines, matériaux

Énergie & finance
Prévision de consommation, marchés, risque

Contrôle qualité industriel
Détection de défauts par vision

Science
Des maths aux plans d'expériences : preuves, simulation, découverte

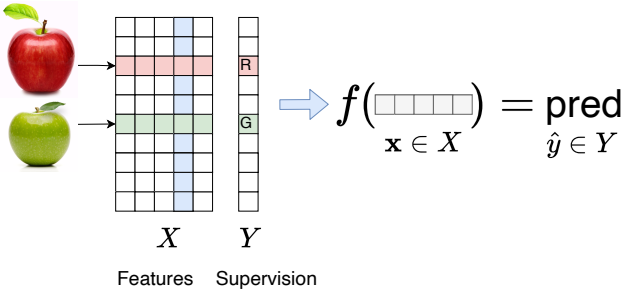
Deep learning & Apprentissage de représentations

[Application aux données textuelles]

Des données tabulaires au texte

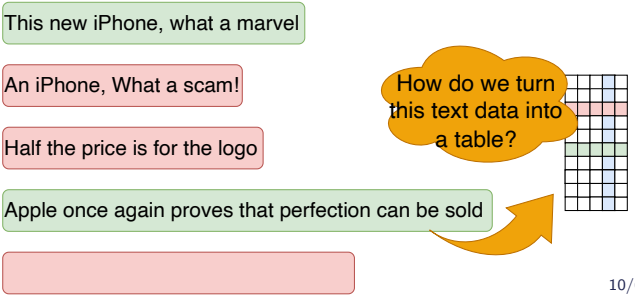
- **Données tabulaires**
  - Dimension fixe
  - Valeurs continues

  ⇒ Un terrain de jeu idéal pour l'apprentissage automatique



- **Données textuelles**
  - Longueurs variables
  - Valeurs discrètes

  ⇒ Complexes pour l'apprentissage automatique

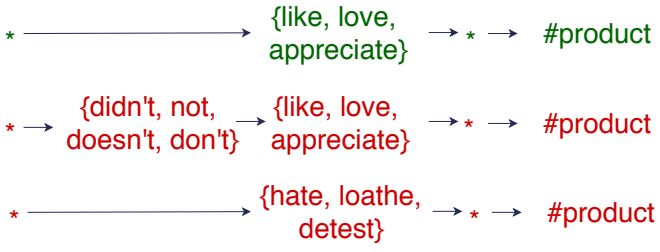


IA + Texte : Trait. Auto. du Langage Naturel (TALN)

TALN = plus grande communauté scientifique en IA

Linguistique [1960-2010]

Systèmes à base de règles :

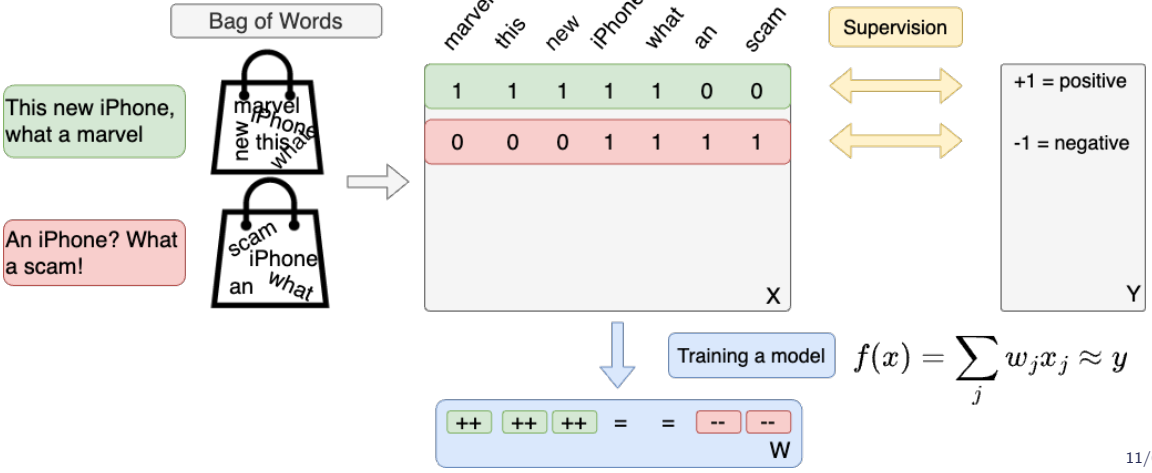


- Nécessite une expertise humaine
- Extraction de règles ⇔ données très propres
- Très grande précision
- Faible rappel
- Système interprétable

IA + Texte : Trait. Auto. du Langage Naturel (TALN)

TALN = plus grande communauté scientifique en IA

Apprentissage auto. [1990-2015]



IA + Texte : Trait. Auto. du Langage Naturel (TALN)

TALN = plus grande communauté scientifique en IA

Linguistique [1960-2010]

- Nécessite une expertise humaine
- Extraction de règles ⇔ données très propres

\+ Système interprétable
\+ Très grande précision
− Faible rappel

Apprentissage auto. [1990-2015]

- Peu d'expertise nécessaire
- Extraction statistique ⇔ robuste aux données bruitées

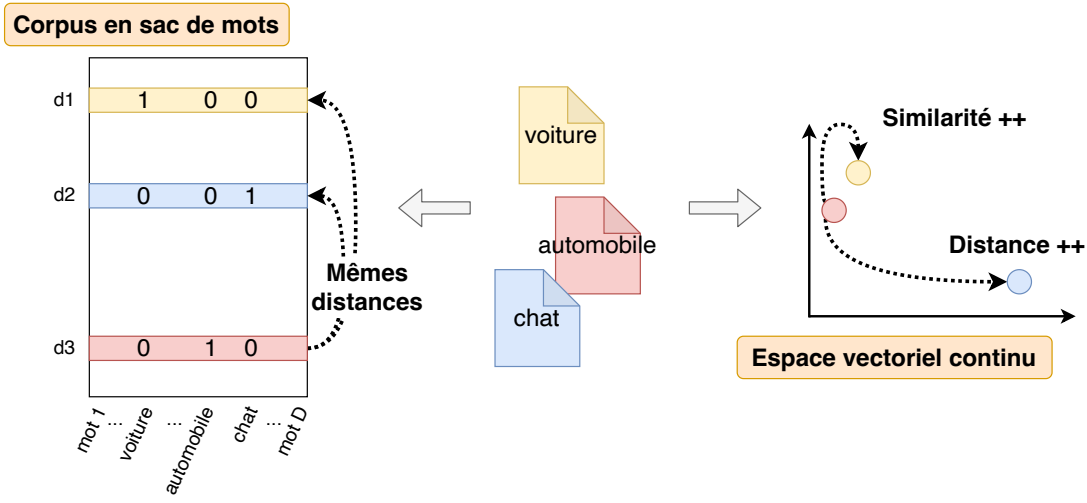
≈ Système moins interprétable
− Moindre précision
\+ Meilleur rappel

Précision = critère d'acceptation par l'industrie

→ Lien vers les métriques

Apprentissage de représentations pour les données textuelles

Du sac de mots aux représentations vectorielles [2008, 2013, 2016]

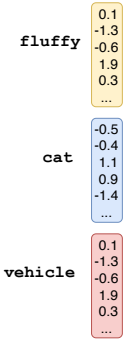


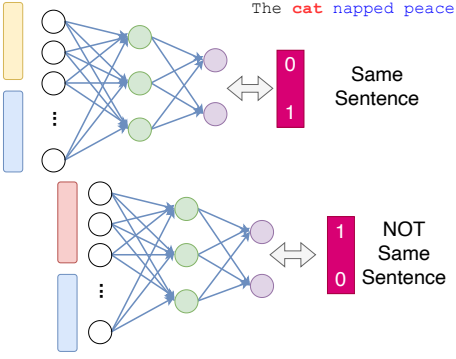
LeCun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

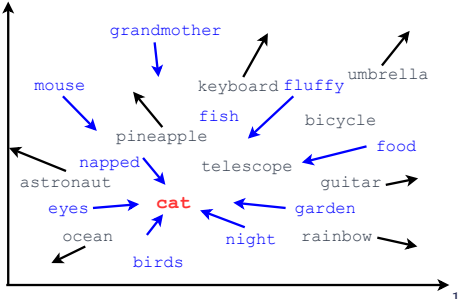
Apprentissage de représentations pour les données textuelles

Du sac de mots aux représentations vectorielles [2008, 2013, 2016]

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.
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.
The cat napped peacefully in the warm sunlight.

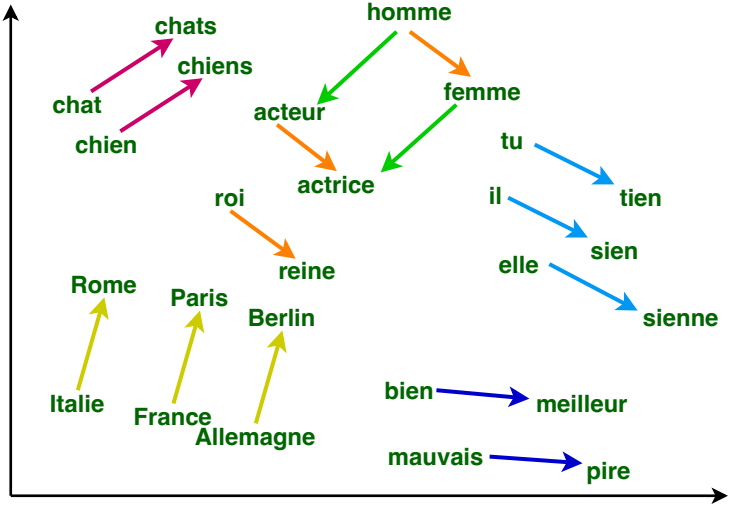






Apprentissage de représentations pour les données textuelles

Du sac de mots aux représentations vectorielles [2008, 2013, 2016]



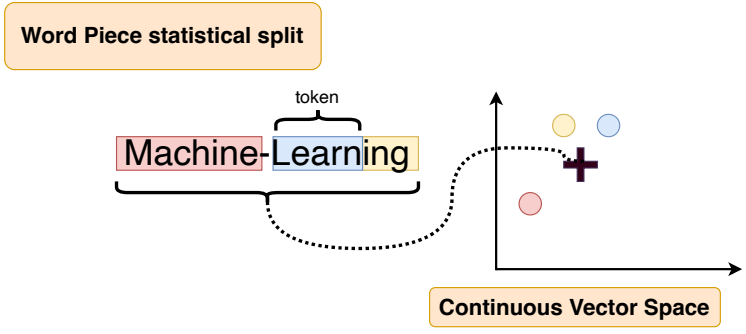
- Espace sémantique :
  significations similaires
  $\Leftrightarrow$
  positions proches
- Espace structuré :
  régularités grammaticales, connaissances de base, ...

Distributed representations of words and phrases and their compositionality, Mikolov et al. NeurIPS 2013

Apprentissage de représentations pour les données textuelles

Du sac de mots aux représentations vectorielles [2008, 2013, 2016]

Des mots aux tokens

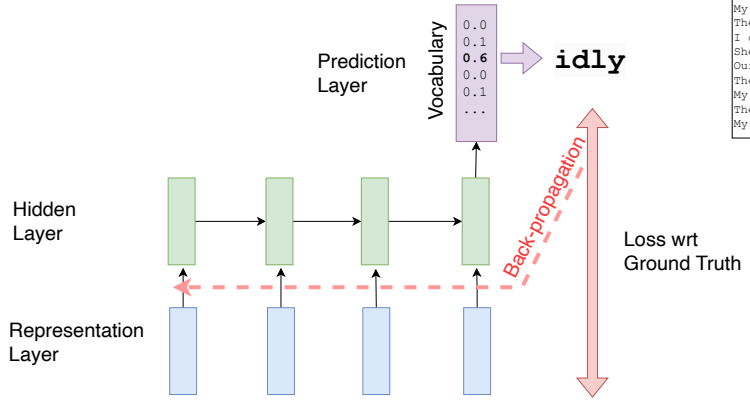


- Représentation des mots inconnus
- Adaptation aux domaines techniques
- Résistance aux fautes d'orthographe

Enriching word vectors with subword information. Bojanowski et al. TACL 2017.

Agrégation des représentations de mots : vers l'IA générative

- Génération et représentation
- Nouvelle manière d'apprendre les positions des mots

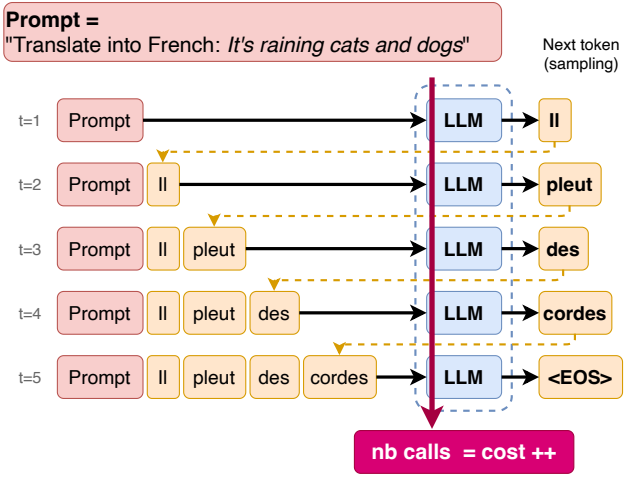


The fluffy cat napped lazily in the sunbeam.
I adopted a stray cat from the shelter last week.
My cat loves to chase after toy mice.
The black cat stealthily crept through the dark alley.
I often find my cat perched on the windowsill, watching birds.
She gently stroked her cat's fur as it purred contentedly.
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.

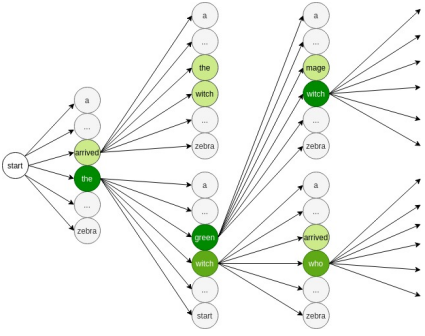
Corpus

Sequence to Sequence Learning with Neural Networks, Sutskever et al. NeurIPS 2014

Inférence & Beam Search



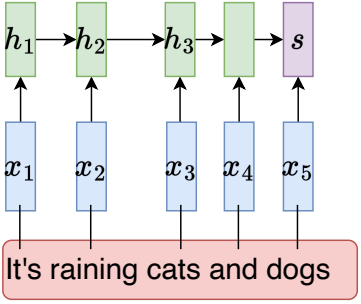
- Coût élevé $\approx$ 1 appel / token
- Principe du max de vraisemblance
- Tâches historique du TALN =
  - Archi. spécifique / data spécifique
  - Ajouter des contraintes a posteriori dans le décodage



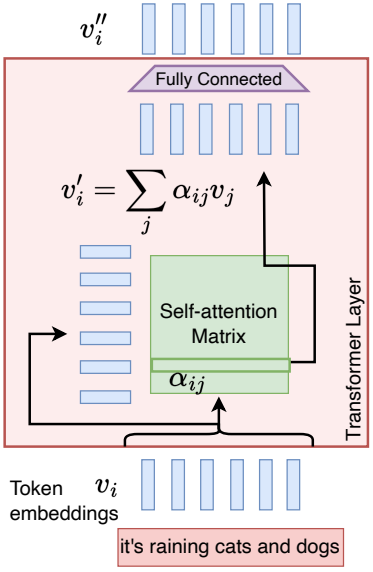
Architecture Transformer : agrégation à l'état de l'art

Réseau de neurones récurrents :

$$h_{t+1} = h_t W_1 + x_{t+1} W_2$$



Transformer :

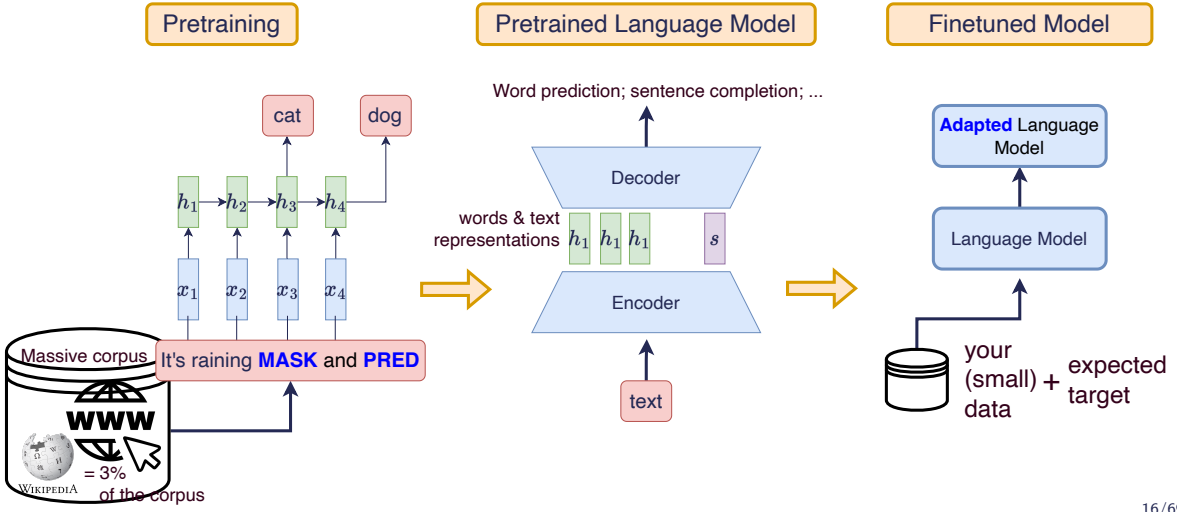


Attention is all you need, Vaswani et al. NeurIPS 2017

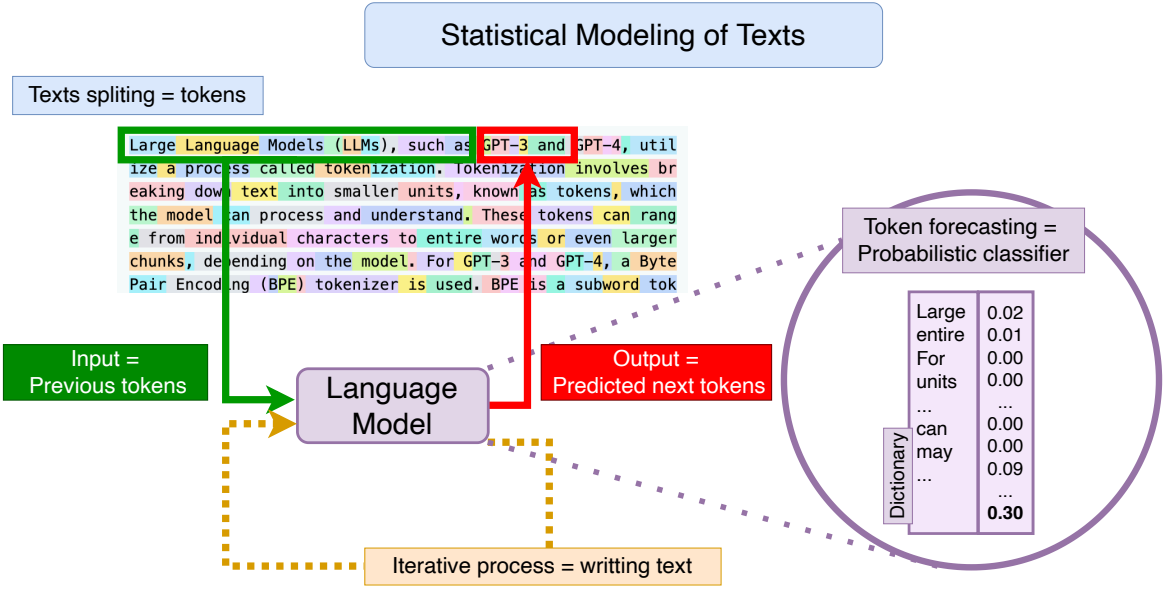
Sequence to Sequence Learning with Neural Networks, Sutskever et al. NeurIPS 2014

Un nouveau paradigme de développement depuis 2015

- Jeu de données massif + architecture massive ⇒ coût d'entraînement +++
- Architecture pré-entraînée + zéro-shot / affinage



Au bout du compte: un perroquet stochastique :)

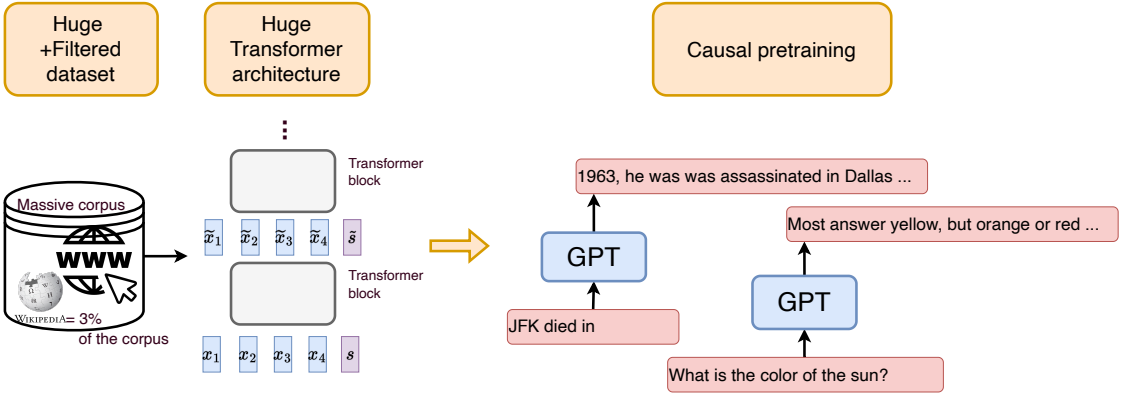


chatGPT

30 novembre 2022
1 million d'utilisateurs en 5 jours
100 millions à la fin janvier 2023
1.16 milliard en mars 2023

Les ingrédients de chatGPT

0. Transformer + données massives (GPT)



- Grammaire : accord singulier/pluriel, concordance des temps
- Connaissances : entités, nom, lieux, dates, ...

Language Models are Few-Shot Learners, Brown et al. 2020

Les ingrédients de chatGPT

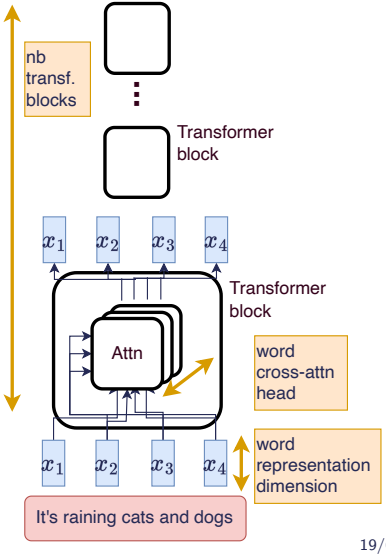
1. Plus, c'est mieux ! (GPT)

+ plus de mots en entrée [500 ⇒ 2k, 32k, 100k]
+ plus de dim. dans l'espace des mots [500-2k ⇒ 12k]
+ plus de têtes d'attention [12 ⇒ 96]
+ plus de blocs/couches [5-12 ⇒ 96]

175 milliards de paramètres...

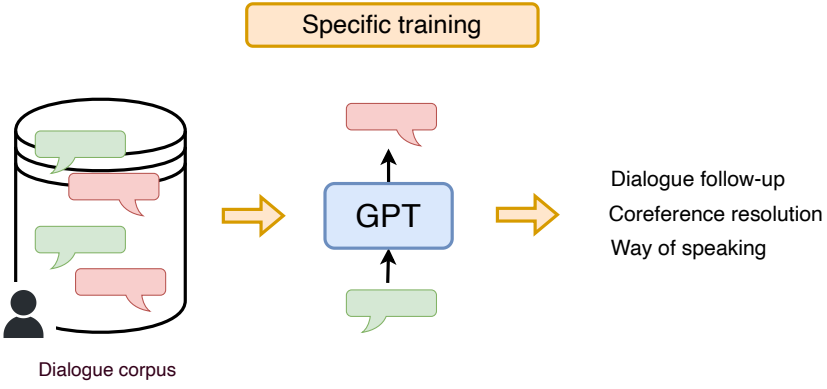
Qu'est-ce que cela signifie ?

- $1{,}75 \cdot 10^{11} \Rightarrow$ 300 Go + 100 Go (stockage pour l'inférence) $\approx$ 400 Go
- GPU NVidia A100 = 80 Go de mémoire (=20k€)
- Coût de (1) entraînement : 4,6 millions €



Les ingrédients de chatGPT

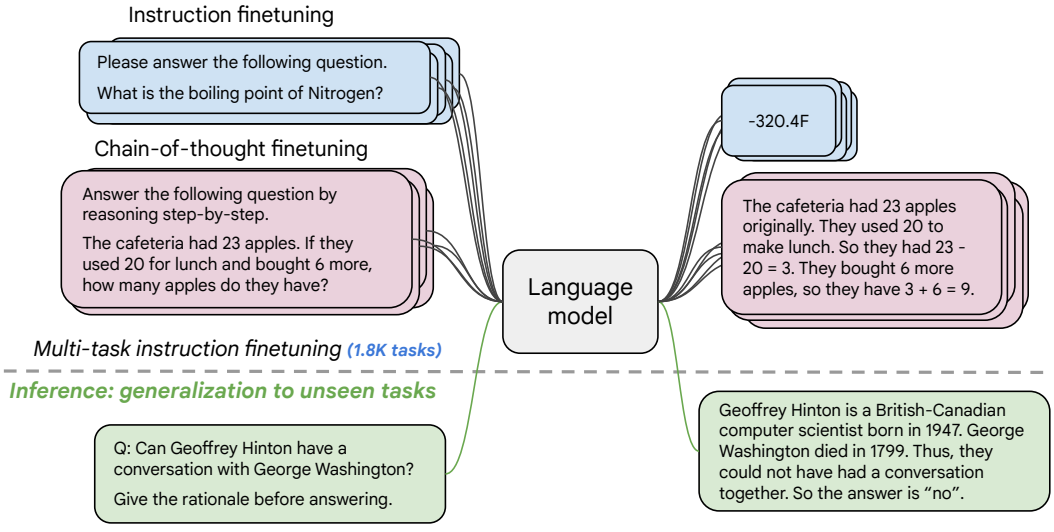
2. Suivi du dialogue



- **Données très propres** Données générées/validées/classées par des humains

Les ingrédients de chatGPT

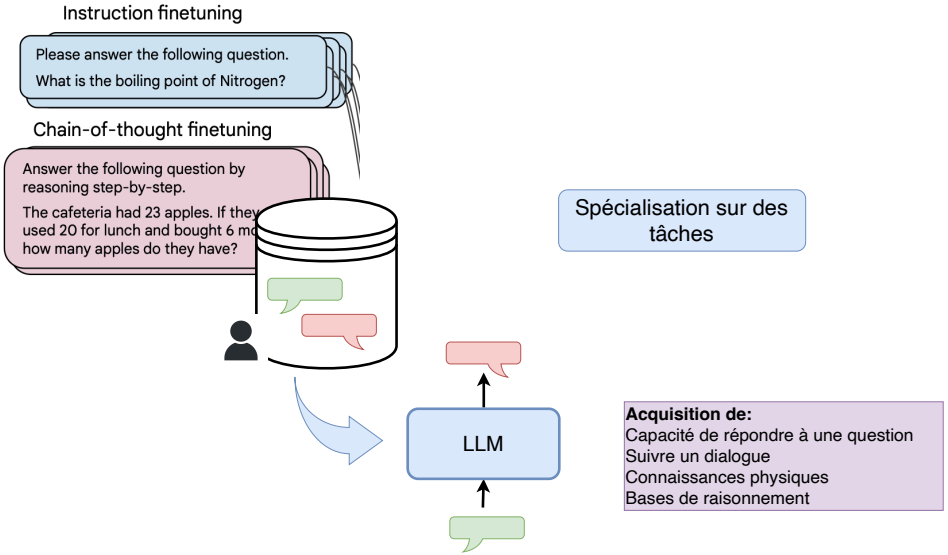
3. Ajustement fin sur des tâches de raisonnement (±) complexes



Scaling Instruction-Finetuned Language Models, Chung et al., JMLR 2024

Les ingrédients de chatGPT

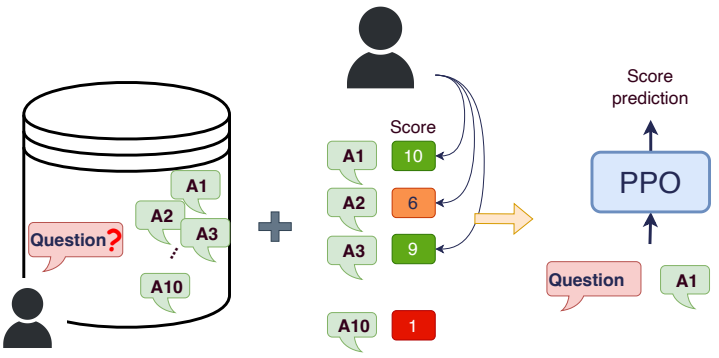
3. Ajustement fin sur des tâches de raisonnement (±) complexes

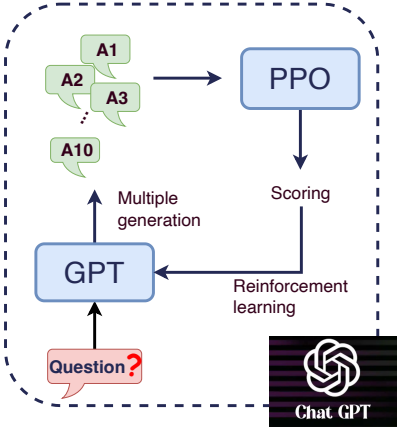


Scaling Instruction-Finetuned Language Models, Chung et al., JMLR 2024

Les ingrédients de chatGPT

4. Instructions + classement des réponses



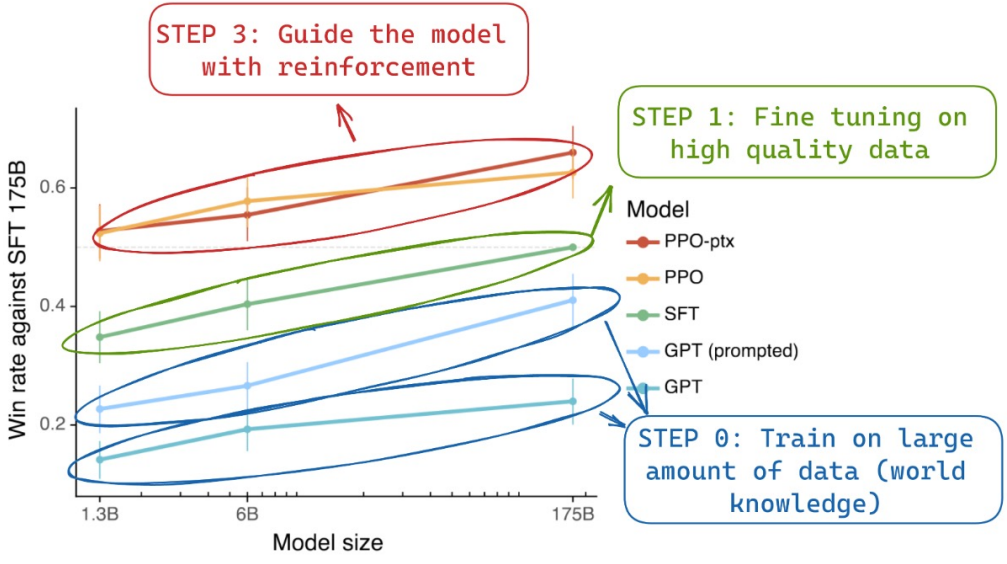


- Base de données créée par des humains
- Amélioration des réponses
- ... Aussi un moyen d'éviter les sujets sensibles = censure

Training language models to follow instructions with human feedback, Ouyang et al., 2022

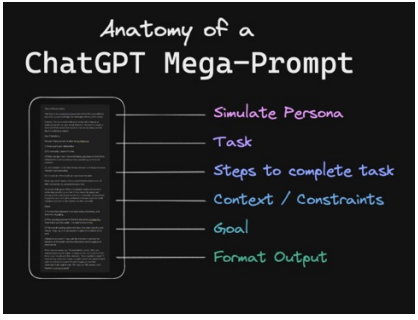
Steps & Performance

Données Massives ⇒ Données Haute Qualité (dialogue) ⇒ Taches ⇒ RLHF



Utilisation de chatGPT & Prompting

- Interroger chatGPT = compétence ⇒ *prompting*
  - Bonne question : *... en détail*, *... étape par étape*
  - Nombre d'élts, ex. : *3 qualités pour ...*
  - Contexte : *cellule* pour un biologiste / assistant juridique
- Ne pas s'arrêter à la première question
  - Détailler certains points
  - Réorienter la recherche
  - Dialoguer
- Reformulation
  - *Explain like I'm 5*, article scientifique, mode pote, ...
  - Résumer, développer
  - Ajouter des erreurs (!)



https://chatgptprompts.guru/what-makes-a-good-chatgpt-prompt/

⇒ Besoin de **pratique** [1 à 2 heures], échanges avec les collègues

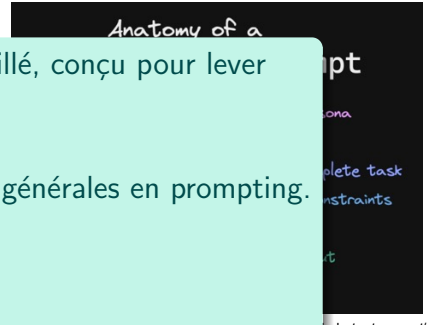
Utilisation de chatGPT & Prompting

- Interroger chatGPT = compétence ⇒ *prompting*
  - Bonne question : *... en détail*, *... étape par étape*
  - Nombre d'élts, ex. : *3 qualités pour ...*
  - Contexte : *cellule* pour un biologiste / assistant
- Ne p
  -
  -
  -
- Refo
  -
  -
  - Ajouter des erreurs (!)

⇒ Besoin de **pratique** [1 à 2 heures], échanges avec les collègues

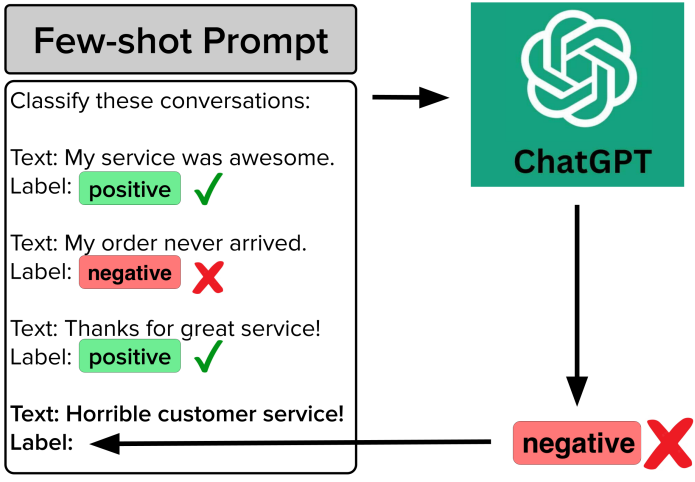
- **Prompt = cahier des charges** : long, détaillé, conçu pour lever toute ambiguïté sur les directions attendues.
- Le prompting requiert
  une **expertise métier** >> connaissances générales en prompting.





Vers le *few-shot learning*

- Apprendre sans modifier le modèle = des exemples dans le prompt

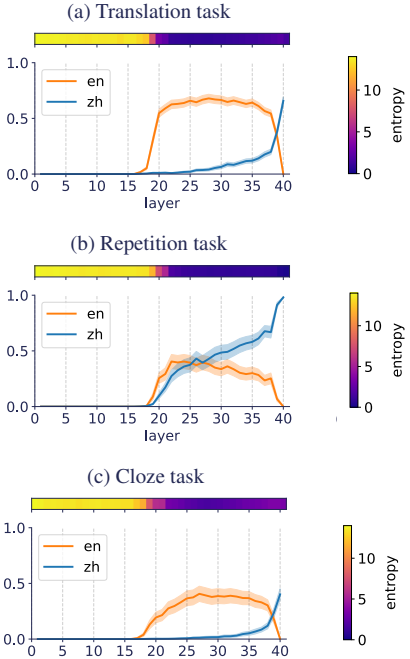


Gestion des langues

- Les modèles de langue sont aujourd'hui multilingues :

⇒ Restez dans la langue où vous êtes à l'aise
⇒ Demandez des réponses dans n'importe quelle langue

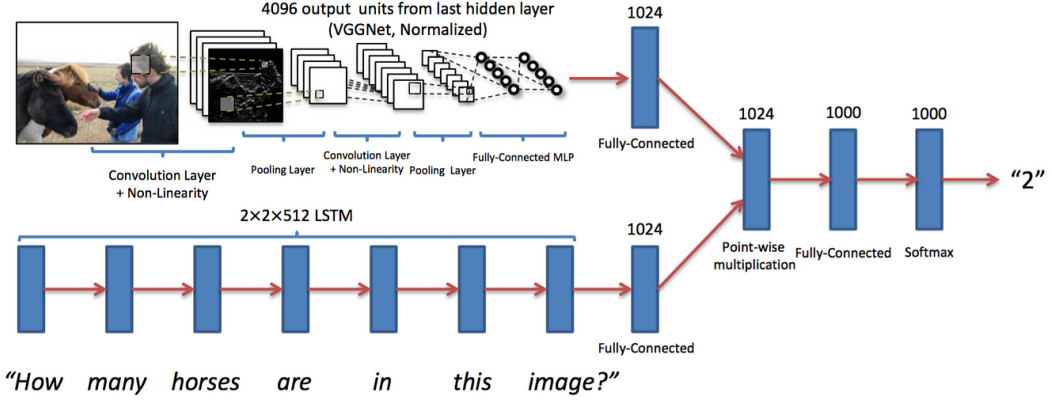
[Wendler et al. 2024] Do Llamas Work in English? On the Latent Language of Multilingual Transformers



GPT-4 & Multimodalité

Fusionner info. texte + image. **Apprendre** l'information conjointement

Exemple du VQA : Visual Question Answering (questions visuelles)



⇒ Rétropropager l'erreur ⇒ modifier les repr. des mots + l'analyse d'image

VQA : Visual Question Answering, arXiv, 2016 , A. Agrawal et al.

GPT-4 & Multimodalité



VLA : Vision Language Action Model

Vision Encoder

State Encoder

Text Encoder

Put the apple in the bowl

Intermediate representation

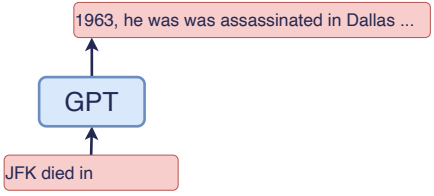
Decoder

Series of instructions

Les limites techniques

chatGPT et la relation à la vérité

1. **Vraisemblance** = grammaire, accords, concordance des temps, enchaînements logiques...
   ⇒ Connaissances répétées
2. Prédit le mot le plus **plausible**...
   ⇒ produit des **hallucinations**
3. chatGPT ≠ **graphes de connaissances**
4. Réponses brillantes...
   Et erreurs absurdes !
   + on ne peut pas prévoir les erreurs
5. Quantification de l'incertitude peu fiable



Début de phrase

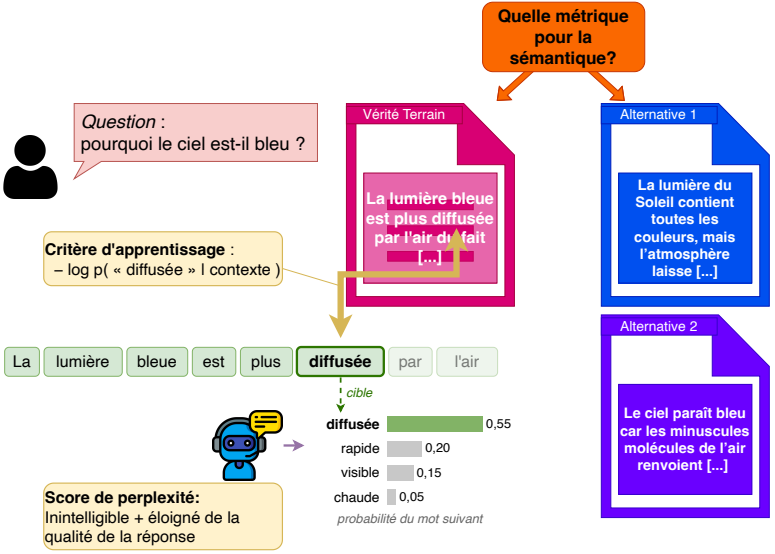
Dans sa bosse, le dromadaire stocke...

Suites possibles : $P(\text{suite} \mid \text{contexte})$

✗	de l'eau	.58← **choisi**
✓	de la graisse	.24
	des réserves	.11
	de la nourriture	.05

IA générative : comment évaluer la performance ?

- Comment évaluer par rapport à la vérité terrain ?
- Comment évaluer la confiance du système / la plausibilité de la génération ?

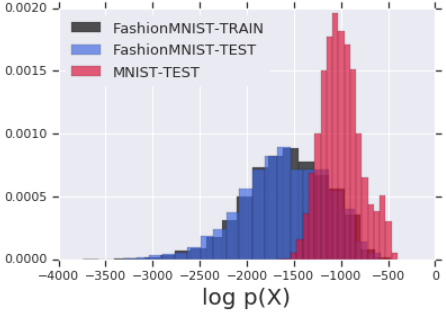


LLM as a judge

Evolution en 2024-25
⇒ hausse qualité des résultats

IA générative : comment évaluer la performance ?

- Comment évaluer par rapport à la vérité terrain ?
- Comment évaluer la confiance du système / la plausibilité de la génération ?



Plausibilité



Entraînement



Test

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

Stabilité / Prévisibilité

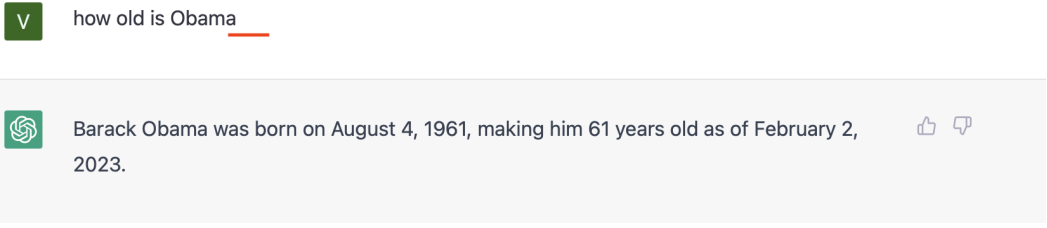
- Difficile d'encadrer un comportement
- Impossible de prédire les bonnes / mauvaises réponses



⇒ Des modèles qui se discréditent régulièrement

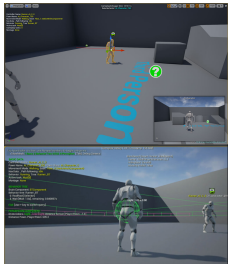
Peu / pas utilisés dans les jeux vidéo

⇒ Impossible de certifier ces modèles pour des applications critiques



Stabilité / Prévisibilité

- Difficile d'encadrer un comportement
- Impossible de prédire les bonnes / mauvaises réponses



⇒ Des modèles qui se discréditent régulièrement

Peu / pas utilisés dans les jeux vidéo

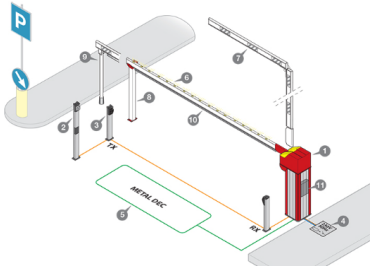
⇒ Impossible de certifier ces modèles pour des applications critiques

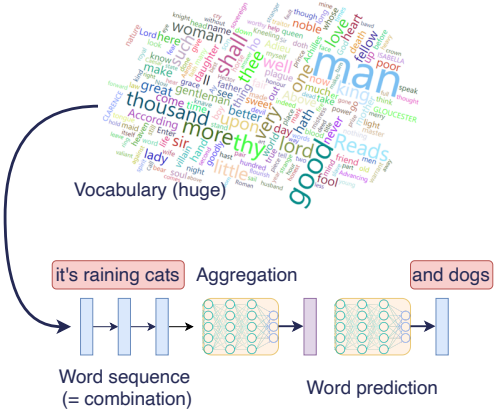
V how old is obama?

As of 2021, Barack Obama was born on August 4, 1961, so he is 60 years old.

V and today?

Stabilité, explicabilité... Et complexité





- Système simple
- Tests exhaustifs des entrées/sorties
- **Prévisible** & **explicable**

- Grande dimension
- Combinaisons non linéaires complexes
- **Non prévisible** & **non explicable**

Stabilité, explicabilité... Et complexité

Interprétabilité *vs* explication a posteriori

Réseaux de neurones = **non interprétables** (presque toujours)

trop de combinaisons à anticiper

Réseaux de neurones = **explicables a posteriori** (presque toujours)



[Accident Uber, 2018]

- Système simple
- Tests exhaustifs des entrées/sorties
- **Prévisible** & **explicable**

- Grande dimension
- Combinaisons non linéaires complexes
- **Non prévisible** & **non explicable**

Transparence : open source / poids ouverts

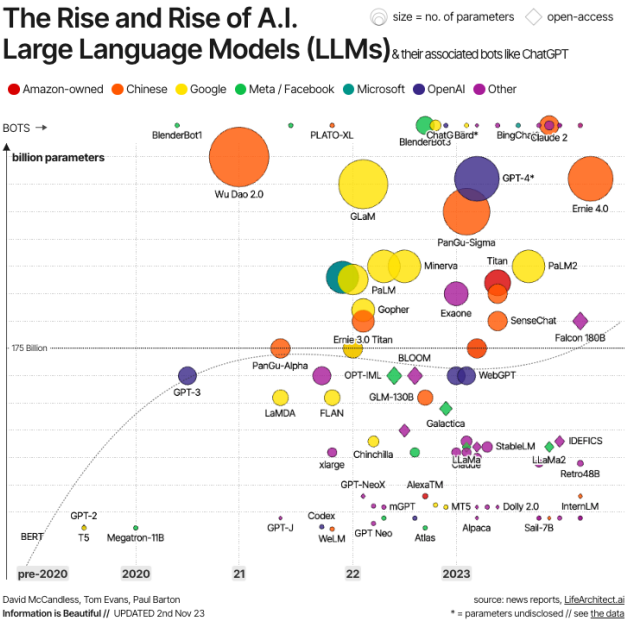
- Puis-je le modifier ? — Adaptation
- Quelles données d'entraînement ? — Contamination des données / compétences
- Quelle ligne éditoriale / censure ? — Accès à l'information
- Pourquoi cette réponse ? — Explicabilité / interprétabilité

Foundation Model Transparency Index Scores by Major Dimensions of Transparency, 2023

Source: 2023 Foundation Model Transparency Index

Major Dimensions of Transparency	Meta Llama 2	BigScience BLOOMZ	OpenAI GPT-4	stability.ai Stable Diffusion 2	Google PaLM 2	ANTHROPIC Claude 2	cohere Command	AI21labs Jurassic-2	Inflection Inflection-1	amazon Titan Text	Average
Data	40%	60%	20%	40%	20%	0%	20%	0%	0%	0%	20%
Labor	29%	86%	14%	14%	0%	29%	0%	0%	0%	0%	17%
Compute	57%	14%	14%	57%	14%	0%	14%	0%	0%	0%	17%
Methods	75%	100%	50%	100%	75%	75%	0%	0%	0%	0%	48%
Model Basics	100%	100%	50%	83%	67%	67%	50%	33%	50%	33%	63%
Model Access	100%	100%	67%	100%	33%	33%	67%	33%	0%	33%	57%
Capabilities	60%	80%	100%	40%	80%	80%	60%	60%	40%	20%	62%
Risks	57%	0%	57%	14%	29%	29%	29%	29%	0%	0%	24%
Mitigations	60%	0%	60%	0%	40%	40%	20%	0%	20%	20%	26%
Distribution	71%	71%	57%	71%	71%	57%	57%	43%	43%	43%	59%
Usage Policy	40%	20%	80%	40%	60%	60%	40%	20%	60%	20%	44%
Feedback	33%	33%	33%	33%	33%	33%	33%	33%	33%	0%	30%
Impact	14%	14%	14%	14%	14%	0%	14%	14%	14%	0%	11%
Average	**57%**	**52%**	**47%**	**47%**	**41%**	**39%**	**31%**	**20%**	**20%**	**13%**	

Coûts / Frugalité



Paramètres

1998	LeNet-5	= 0,06M
2011	Senna	= 7,3M
2012	AlexNet	= 60M
2017	Transformer	= 65M / 210M
2018	BERT	= 110M / 340M
2019	GPT2	= 1 500M
2020	GPT3	= 175 000M

Comportements différents, coûts différents

Les IA sont démasquées !

Mistral/**Ministral**

SEMI-OUVERT | 8 MDS DE PARAMÈTRES | SORTIE 10/2024

Optimisé pour un temps de réaction rapide, ce modèle est idéal pour des applications nécessitant des réponses immédiates et peut supporter plus de 100 langues. Sorti en octobre 2024.

Impact énergétique de la discussion

8 milliards param. taille du modèle	×	**128** tokens taille du texte	=	**0.30** Wh énergie conso.

Ce qui correspond à :

0.30g CO_2 émis	**5**min ampoule LED	**33**s vidéos en ligne

Voir plus

DeepSeek/**DeepSeek v3**

SEMI-OUVERT | 671 MDS DE PARAMÈTRES | SORTIE 12/2024

Sorti en décembre 2024, le modèle DeepSeek V3 possède une architecture Mixture-of-Experts qui lui permet d'être d'une très grande taille en diminuant les coûts d'inférence.

Impact énergétique de la discussion

671 milliards param. taille du modèle	×	**225** tokens taille du texte	=	**6**Wh énergie conso.

Ce qui correspond à :

6g CO_2 émis	**2**h ampoule LED	**12**min vidéos en ligne

Voir plus

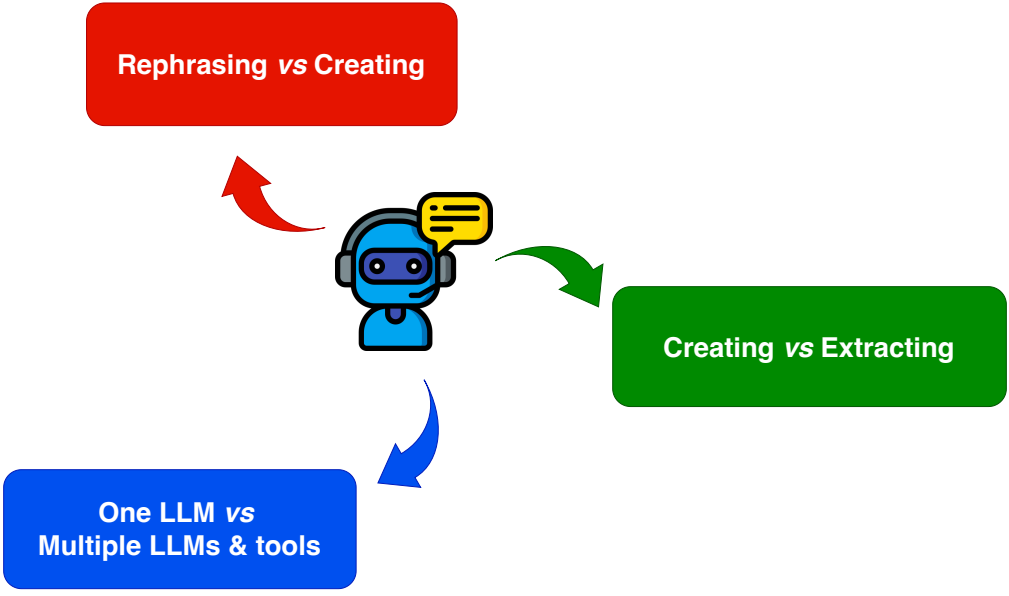
Tout ce qui dépasse les capacités / l'entraînement du LLM

- Calculs simples
  (multiplication, division)
- Générer des noms d'animaux en n syllabes (en cours)
- Jouer aux échecs
- Suivre un raisonnement causal (complexe)
- ...



Typologie des usages des LLMs

Trois questions pour cadrer les usages



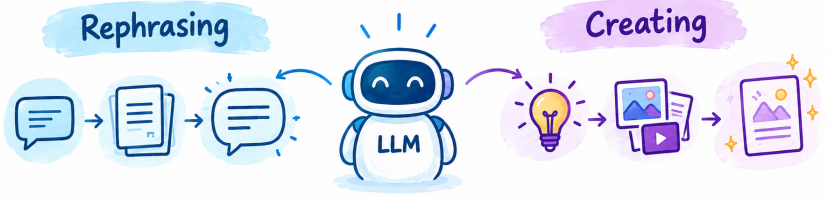
(1) Reformuler *vs* créer du contenu

Reformuler *j'apporte le contenu*

- **Traduire**/**relire** un paragraphe
- **Reformuler** : style, ton, -10%
- Paragraphe à partir de **mes** puces
- Compte-rendu depuis **mes** notes
- Lettre de motivation : CV + offre
- Tableaux, OCR, LaTeX/Excel
- Le résumé de **mon propre** article

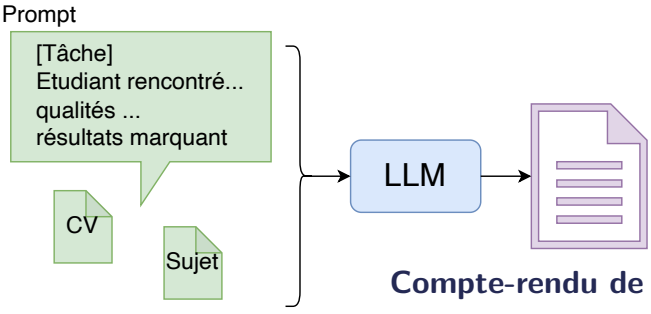
Créer *le modèle apporte le contenu*

- **Trouver** l'inspiration [page blanche]
- **Proposer** un plan, un acronyme
- **Répondre** là où je ne sais pas
- **Tester** une idée, ses faiblesses
- Rédiger sur un sujet donné
  mais non détaillé
- Rédiger une relecture, un état de l'art

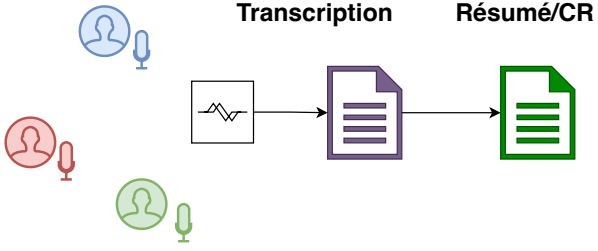


(1) Reformuler : exemples

Construire une lettre de recommandation



Compte-rendu de réunion



(1) Questions éthiques

Reformuler

Le contenu est **le mien**. Le modèle touche à la *forme*.

→ Suis-je toujours l'auteur ? **oui**

→ Dois-je le déclarer ?
comme un correcteur ? traducteur ?

→ Risque = **distorsion silencieuse**
nuance perdue, précaution effacée

→ Risque = **confidentialité**
texte inédit envoyé sur un serveur

Créer

Le contenu vient **du modèle**, *i.e.* d'un corpus que je n'ai pas choisi.

→ Suis-je toujours ~~l'auteur~~ responsable ? **oui**

→ Qui dois-je citer ?
le modèle n'a pas de source

→ Risque = **hallucination**
plausible, bien écrit, faux

→ Risque = **uniformisation de la pensée**
l'idée la plus probable, pas la meilleure

⇒ Même outil, même zone de saisie, **deux débats totalement différents**.

(2) Créer *ex-nihilo* *vs* extraire de l'information

Ex-nihilo *mémoire paramétrique*

- La réponse est **dans les poids**
- *JFK est mort en...* ⇒ réponse probabiliste
- Pas de source, pas de traçabilité
- Figé au moment de l'entraînement

⇒ hallucination (mesure de *factualité*)

Extraction *le document est fourni*

- La réponse est **dans l'entrée**
- Le modèle *localise*, il n'invente pas
- La source est connue *par construction*
- À jour *= aussi frais que le document*

⇒ vérifiable (mesure de *fidélité*)

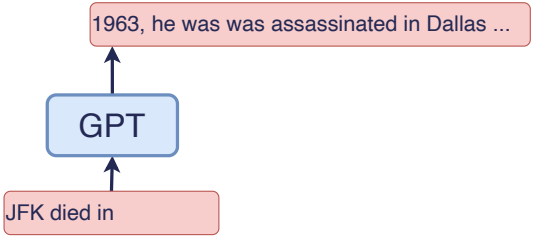


(2) Créer *ex-nihilo* *vs* extraire de l'information

Ex-nihilo *mémoire paramétrique*

- La réponse est **dans les poids**
- *JFK est mort en...* ⇒ réponse probabiliste
- Pas de source, pas de traçabilité
- Figé au moment de l'entraînement

⇒ hallucination (mesure de *factualité*)



Extraction *le document est fourni*

- La réponse est **dans l'entrée**
- Le modèle *localise*, il n'invente pas
- La source est connue *par construction*
- À jour = *aussi frais que le document*

⇒ vérifiable (mesure de *fidélité*)

Passage Sentence

In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under gravity.

Question

What causes precipitation to fall?

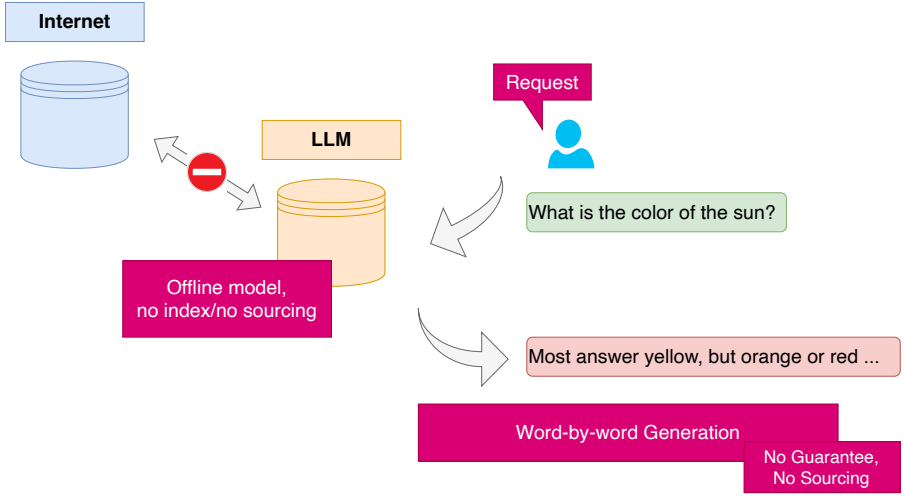
Answer Candidate

gravity

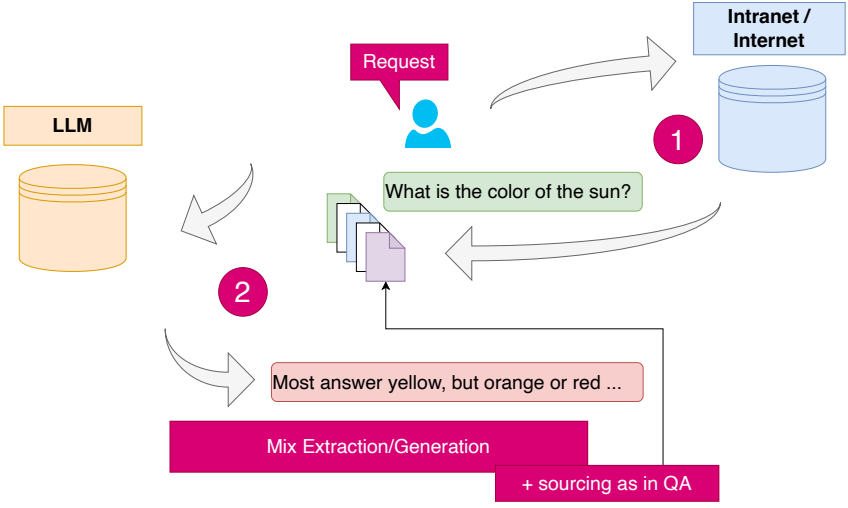
⇒ Même question, même interface... mais une seule des deux est vérifiable.

(2) LLMs ⇒ RAG : mémoire *vs* extraction d'information

- Poser des questions à ChatGPT... Un usage surprenant !
- Mais est-ce raisonnable ? [Une vraie question ouverte (!)]



(2) LLMs ⇒ RAG : mémoire *vs* extraction d'information



- RAG : *Retrieval-Augmented Generation*
- Taille d'entrée limitée (2k, 32k, 200k tokens)

(3) Des LLMs aux agents

LLM *un outil / une tâche*

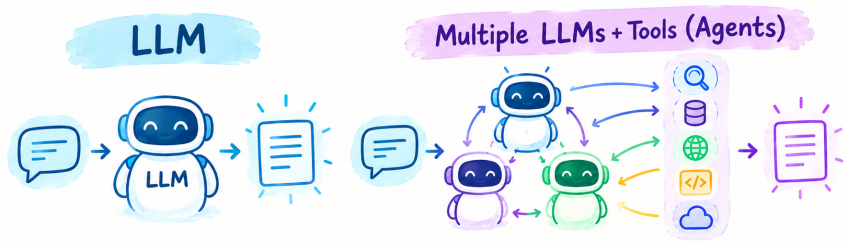
- Une question → une réponse
- **Je** lis, **je** juge, **je** garde ou je jette
- Le modèle produit du **texte**
- Aucun effet sur le monde
- Efficace sur de petites tâches

⇒ validation humaine **à chaque étape**

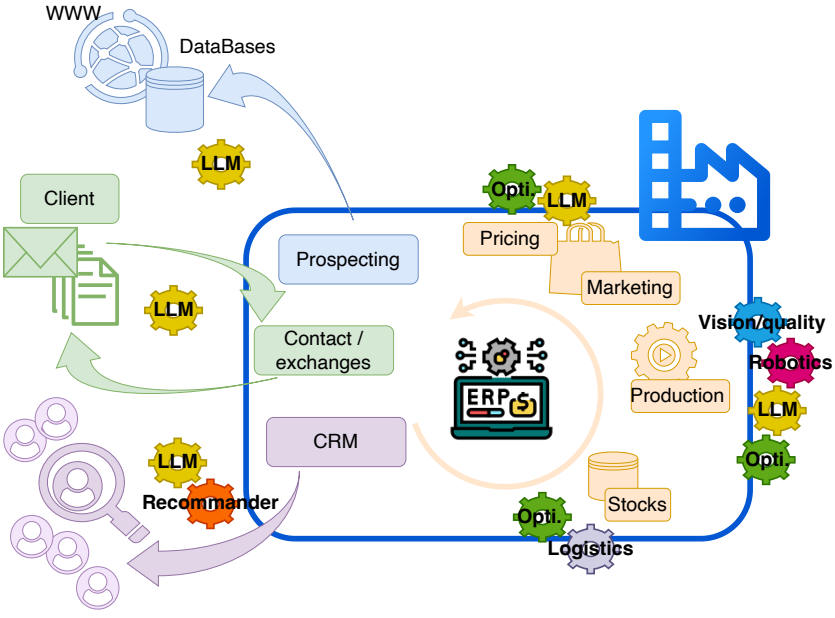
Agent *un système / des tâches*

- Un objectif → une **chaîne** d'actions
- Appels d'**outils**, boucles, essais
- Le modèle produit des **actions**
  requêtes, fichiers, emails, code
- Effets sur le monde

⇒ validation humaine **à la fin**



(3) Intégration (multiple) dans un système complexe



- Liste de prospects
- Traitement des emails, identification des besoins
- Interaction avec l'ERP $\Rightarrow$ bases de propositions
- Tarification
- Rédaction d'une proposition commerciale

(3) Toolformer : quand le LLM fait appel à des outils

Le LLM [2023] :

1. Identifie ses propres faiblesses
2. Appelle des outils/API pour mieux répondre

⇒ Sources de données contrôlées (SQL, Wikipédia) = RAG++ ; calculatrice ; traducteur ; moteur de calcul spécialisé

Le LLM au cœur du système

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.

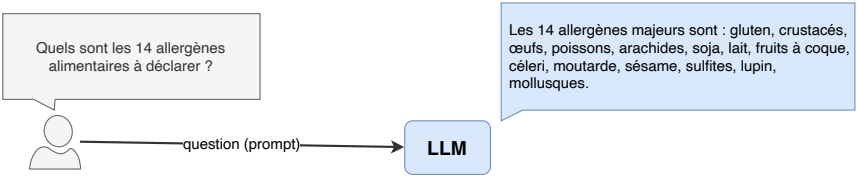
Out of 1400 participants, 400 (or **[Calculator(400 / 1400) → 0.29]** 29%) passed the test.

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

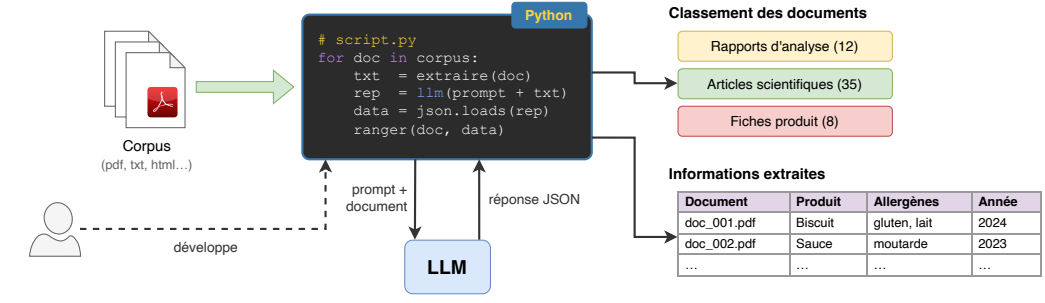
The Brown Act is California's law **[WikiSearch("Brown Act") → The Ralph M. Brown Act is an act of the California State Legislature that guarantees the public's right to attend and participate in meetings of local legislative bodies.]** that requires legislative bodies, like city councils, to hold their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

(3) Positionnement par rapport aux outils

1. Usage direct — interaction manuelle, une question à la fois

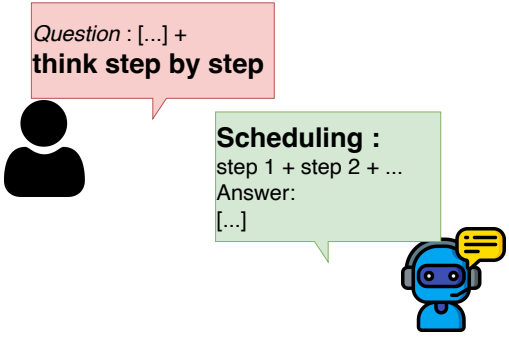


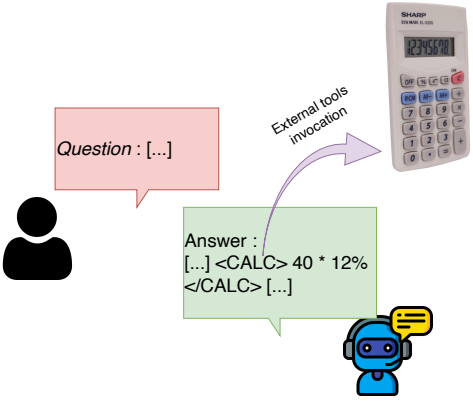
2. LLM intégré dans une chaîne de traitement — traitement automatique de N documents



(3) De la *Chain-of-Thought* aux actions

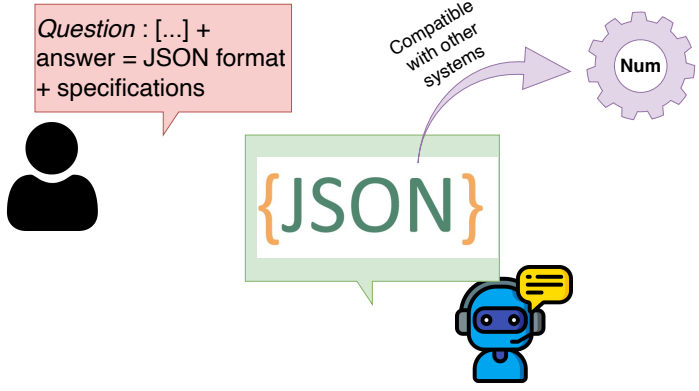
1 *Chain-of-Thought* / Toolformer





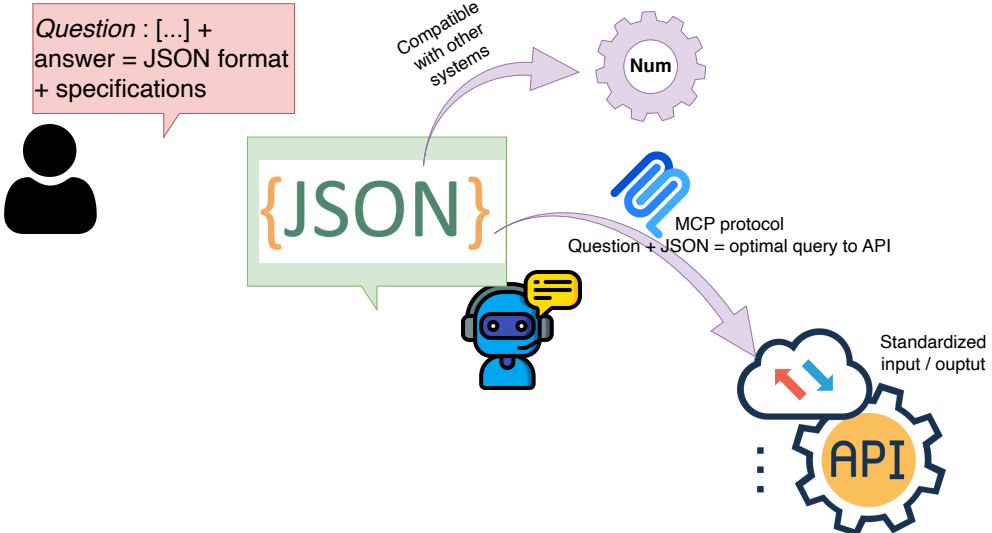
(3) De la *Chain-of-Thought* aux actions

1. *Chain-of-Thought* / Toolformer
2. Des réponses textuelles à l'information structurée



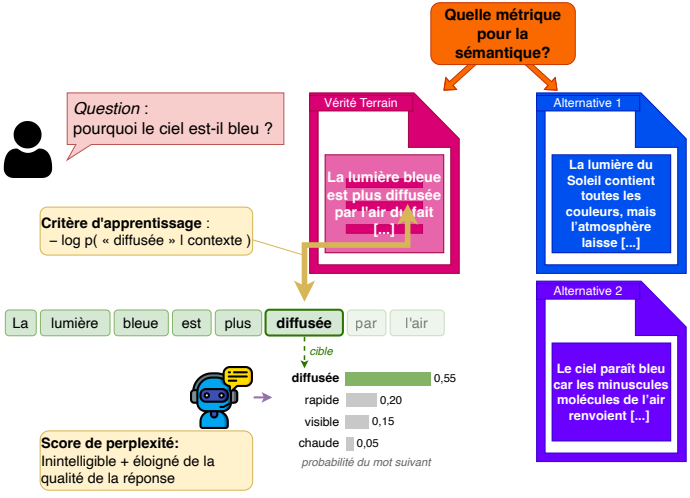
(3) De la *Chain-of-Thought* aux actions

1. *Chain-of-Thought* / Toolformer
2. Des réponses textuelles à l'information structurée



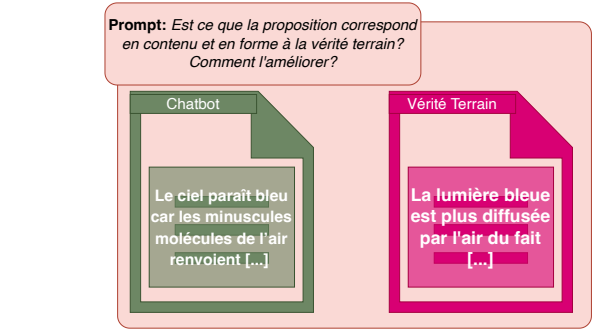
(3) De la *Chain-of-Thought* aux actions

1. *Chain-of-Thought* / Toolformer
2. Des réponses textuelles à l'information structurée
3. LLM juge (*LLM as a judge*) / réflexivité + boucle d'amélioration



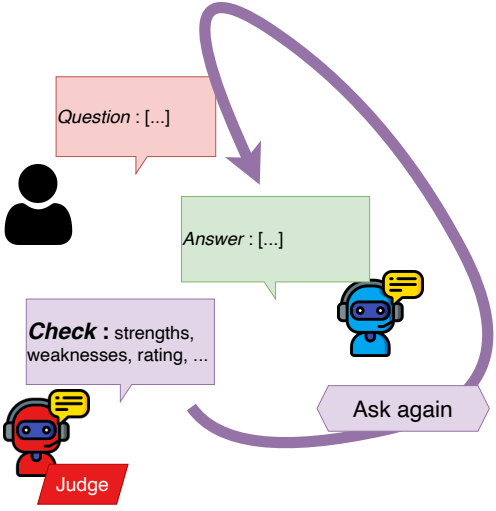
(3) De la *Chain-of-Thought* aux actions

1. *Chain-of-Thought* / Toolformer
2. Des réponses textuelles à l'information structurée
3. LLM juge (*LLM as a judge*) / réflexivité + boucle d'amélioration
4. Parallélisme



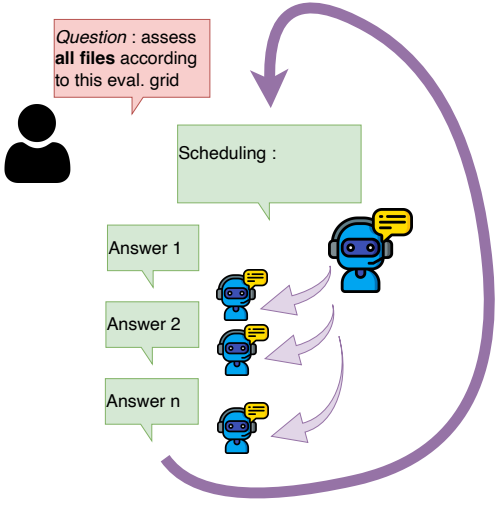
(3) De la *Chain-of-Thought* aux actions

1. *Chain-of-Thought* / Toolformer
2. Des réponses textuelles à l'information structurée
3. LLM juge (*LLM as a judge*) / réflexivité + boucle d'amélioration
4. Parallélisme
5. MCP / actions



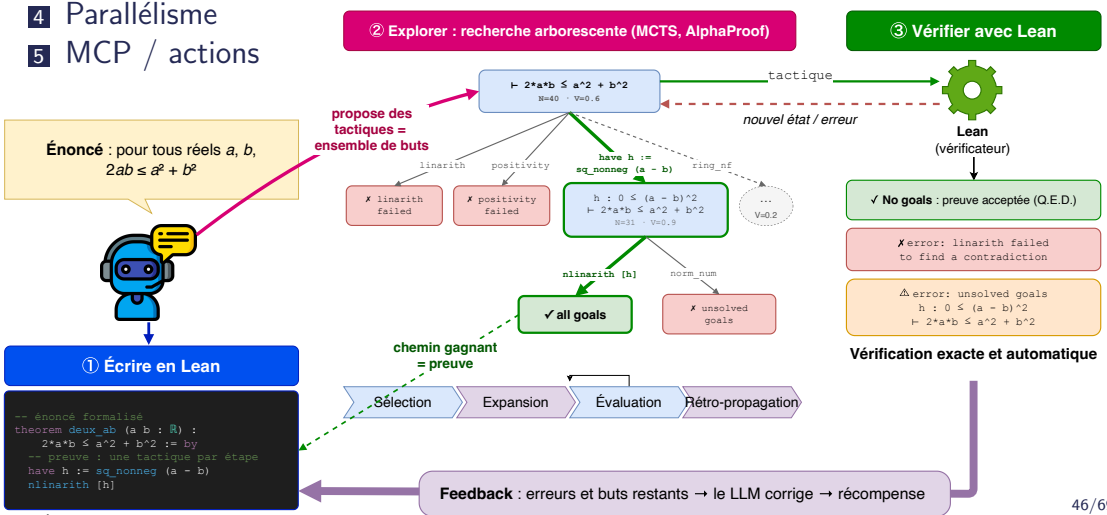
(3) De la *Chain-of-Thought* aux actions

1. *Chain-of-Thought* / Toolformer
2. Des réponses textuelles à l'information structurée
3. LLM juge (*LLM as a judge*) / réflexivité + boucle d'amélioration
4. Parallélisme
5. MCP / actions



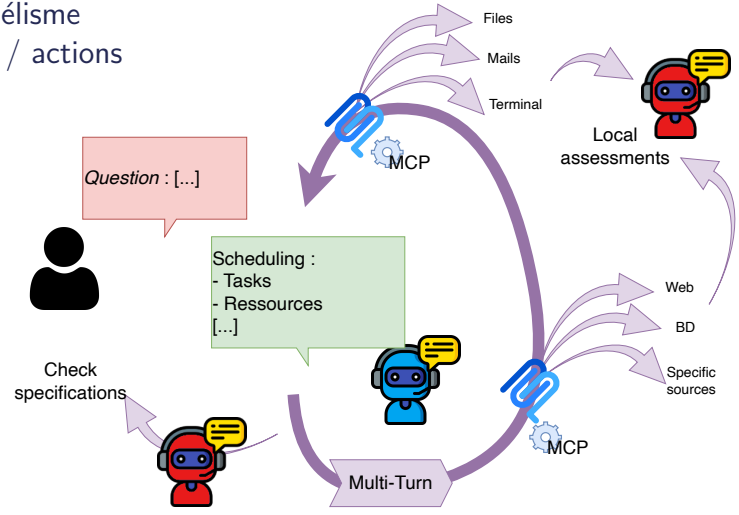
(3) De la *Chain-of-Thought* aux actions

1. *Chain-of-Thought* / Toolformer
2. Des réponses textuelles à l'information structurée
3. LLM juge (*LLM as a judge*) / réflexivité + boucle d'amélioration
4. Parallélisme
5. MCP / actions



(3) De la *Chain-of-Thought* aux actions

1. *Chain-of-Thought* / Toolformer
2. Des réponses textuelles à l'information structurée
3. LLM juge (*LLM as a judge*) / réflexivité + boucle d'amélioration
4. Parallélisme
5. MCP / actions



(3) De l'importance du Harnais

LLM alone

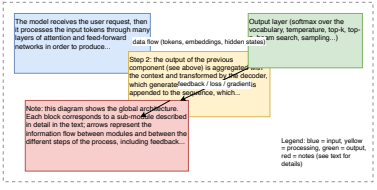
it knows the XML syntax... but never sees the result

Question : draw a figure explaining [...]

Answer :

```
<mxfile>
<mxCell value="..."/>
[...]
</mxfile>
```

one shot, no feedback

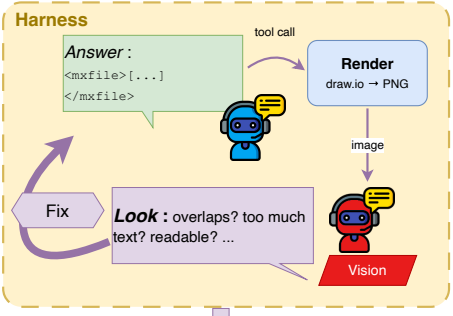


text-heavy, overlaps, unreadable

LLM + harness

render, look, fix: generation becomes a loop

Question : draw a figure explaining [...]



clear, checked, polished

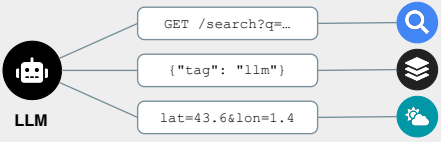
(3) Exemple d'usage du Harnais

[Remi Decoupes]

Serveur MCP : *Model Context Protocol*

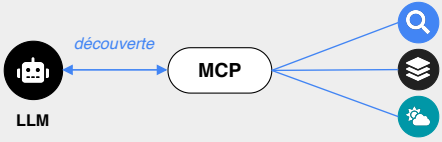
Avant

Le LLM doit écrire la requête propre à chaque API



Avec MCP

Le LLM découvre outils, ressources et paramètres



Un serveur MCP expose

Des outils

Des actions, avec leurs paramètres décrits

Des ressources

Des données à lire : fichiers, documents

Des prompts

Des modèles de requêtes prêts à l'emploi

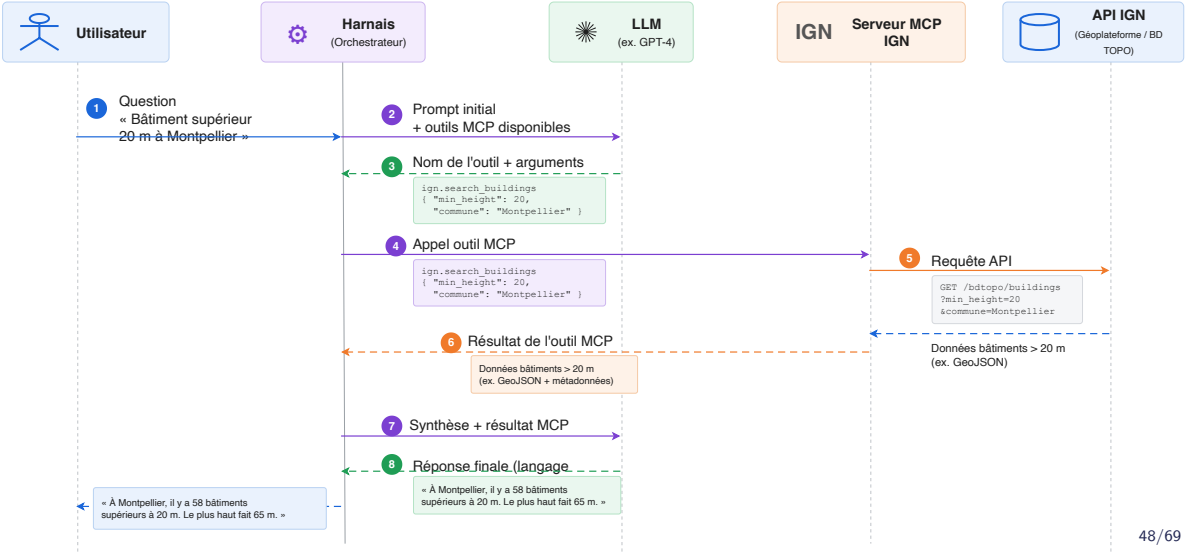
- Web
- Mail
- Word, Excel, Powerpoint
- Drawio, ...

(3) Exemple d'usage du Harnais

[Remi Decoupes]

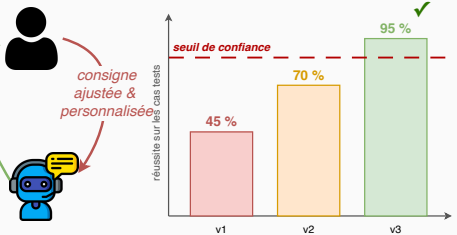
Requête complexe:

Combien y a-t-il de batiments de plus de 20m de haut à Montpellier?



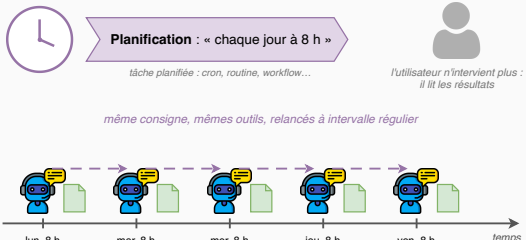
(3) Les agents récurrents

1 Mise au point : l'agent devient fiable



automatiser

2 Lancement récurrent : l'agent tourne seul



Mails pendant les vacances

toutes les heures

« Préviens-moi si un message urgent arrive ; sinon, résume-moi la journée. »

Boîte mail accès en lecture seule → **Alerte** immédiate (SMS, notification) ; **Résumé** du jour, chaque soir

Veille technologique

chaque lundi à 8 h

« Quoi de neuf cette semaine sur les agents IA ? Garde le meilleur. »

Web, arXiv, blogs, brevets, dépôts de code → **Note de veille** : 5 nouveautés + liens sources

Revue de presse

chaque matin à 7 h

« Synthèse de l'actualité de mon secteur, classée par thème. »

Flux RSS, sites de presse, dépêches → **Revue de presse** par thème, avec citations

IA générative & travail :

ce que mesurent les premières études

Un métier = un panier de tâches

Approche par les tâches

On ne remplace pas un **métier**, on transforme des **tâches** :

- certaines sont automatisées
- d'autres sont augmentées
- de nouvelles apparaissent (vérifier, cadrer, intégrer)

Autor, Levy & Murnane (*QJE*, 2003) ; Acemoglu & Autor (2011) ; nouvelles tâches : Acemoglu & Restrepo (2019), Autor et al. (*QJE*, 2024)

Ce qui change avec l'IA générative :

- Vagues précédentes : tâches **routinières**
- LLMs : **écrit**, analyse, code, conseil
  $\Rightarrow$ des métiers qualifiés

Exposition des emplois

- Eloundou et al. (*Science*, 2024) :
  $\approx 80\%$ des actifs américains ont
  $\geq 10\%$ de leurs tâches exposées ;
  $\approx 19\%$ en ont au moins la moitié
- OIT (2023, 2025) : $\approx$ 1 emploi sur 4 exposé dans le monde ;
  l'**augmentation** domine l'**automatisation** ; emplois administratifs les plus exposés (plus souvent occupés par des femmes)

Exposé $\neq$ remplacé : un potentiel technique, pas un effet mesuré.

Gains de productivité : expériences contrôlées

Étude	**Population / tâche**	**Résultat**
Noy & Zhang *Science* 2023	453 professionnels rédaction	temps -40%, qualité $+18\%$; écarts entre participants réduits
Brynjolfsson et al. *QJE* 2025	5 000 téléconseillers support client	$+14\%$ de demandes résolues/heure ; $+34\%$ pour les novices, ≈ 0 pour les plus expérimentés
Dell'Acqua et al. (HBS / BCG) 2023	758 consultants tâches de conseil	dans le périmètre : $+12\%$ de tâches, -25% de temps, qualité $+40\%$; hors périmètre : -19 points de réponses correctes
Peng et al. (GitHub) 2023	développeurs tâche de code isolée	tâche réalisée 56% plus vite

Des gains **réels**, **concentrés** sur certaines tâches
qui **compressent les écarts** de compétence.

Le périmètre irrégulier de l'IA (*jagged frontier*)

- Capacités du modèle $\neq$ intuition de la difficulté
  - ↗↗ synthèse complexe
  - ↘↘ calcul simple ou un détail factuel
- Danger = **hors du périmètre** : la réponse reste fluide et convaincante
- Trois postures observées :
  - Centaures (15%) : partage net des tâches humain / machine *upskilling*
  - Cyborgs (60%) : imbrication continue + vérification *new skilling*
  - Self-automators (25%) : délégation brute *no skilling*

Usage raisonnable

Savoir **où s'arrête le périmètre** pour **sa** tâche = compétence professionnelle.

Lien avec les limites

Relation à la vérité, instabilité, absence d'explication : le périmètre bouge d'une version de modèle à l'autre.

Dell'Acqua et al., *Navigating the Jagged Technological Frontier*, HBS Working Paper 24-013, 2023

Randazzo al., *Cyborgs, Centaurs and Self-Automators*, HBS Working Paper, 2024

Perception, mesure & économie réelle

Perception ≠ mesure

METR (2025) : 16 développeurs open source expérimentés, sur leurs propres projets

- avec assistants IA : 19% plus lents
- Estimation perso. ≈ 20% plus rapides (!)

Laboratoire ≠ économie

Humlum & Vestergaard (2025), Danemark, 25 000 salariés : adoption massive, effets sur salaires et heures travaillées **non détectables**; gain de temps déclaré ≈ 3%

Ce qui fait la différence :

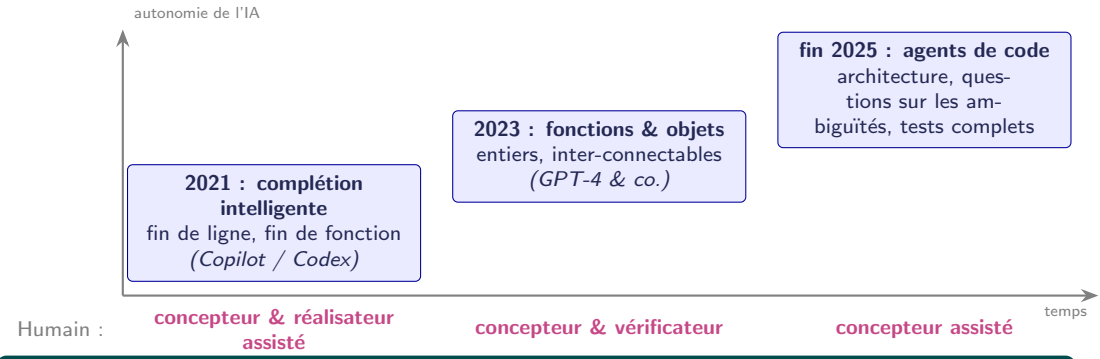
- tâche isolée *vs* travail réel (contexte, qualité attendue, relecture)
- novice *vs* expert
- coût de **vérification** sous estimé
- réorganisation du travail : gain = temps, volume, qualité ?

Gains individuels = productivité collective

⇔

transformation de l'organisation

Seuils de performance : l'exemple du code

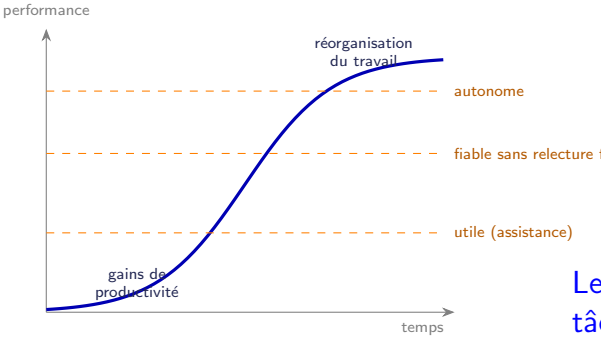


Impacts sur l'emploi = **seuils** de performance

Une tâche bascule quand l'IA la réalise **assez bien pour se passer d'une supervision fine**. Le métier ne disparaît pas : **la place de l'humain dans la chaîne change**.

SWE-bench Verified : quelques % (2023) → >75 % (fin 2025) durée des tâches réalisables par un agent : ×2 tous les ∼7 mois (METR, Kwa *et al.* 2025)

Raisonner en seuils, métier par métier



- Sous le seuil : un outil d'assistance (+x % de productivité)
- Au-dessus : la tâche change de main (réorganisation, effectifs)
- Métier = panier de tâches
  $\Rightarrow$ seuils franchis **différentes dates**

Le code informatique est en avance : tâches **vérifiables** (tests), données massives

Exercice

Pour votre métier : quelles tâches ont déjà franchi le seuil ?
Lesquelles le franchiront en 2027 ?

La responsabilité reste humaine

- L'opérateur vérifie moins... mais reste responsable
  - biais d'automatisation, baisse de vigilance
  - relire 2 000 lignes générées $\neq$ les avoir écrites
- Par défaut : l'opérateur humain répond de l'erreur
  - celui qui valide le code, signe le rapport, envoie le mail
  - AI Act (art. 14) : *supervision humaine* des systèmes à haut risque
- Le contre-exemple BYD
  - juil. 2025 : le constructeur couvre les dommages de son stationnement automatique
  - mai 2026 : étendu à la conduite assistée en ville (1 an)
  - usage du stationnement auto : 21 % $\rightarrow$ 93 % (chiffres BYD)



Lien entre confiance & responsabilité : adoption $\nearrow$ quand quelqu'un d'autre que l'utilisateur assume le risque ET/OU adaption $\nearrow$ en cas de confiance au-dessus d'un seuil

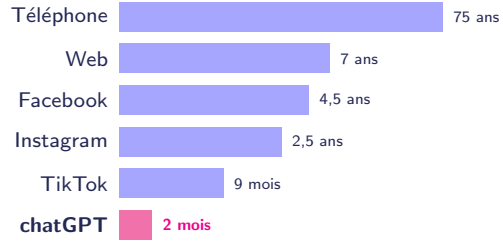
$\Rightarrow$ quid des décisions imbriquées? $\Rightarrow$ quid des décisions militaires?

Une transformation à grande vitesse

Une adoption sans précédent

Temps pour atteindre 100 millions d'utilisateurs

(échelle logarithmique)



Ordres de grandeur couramment cités (UBS 2023)

nov. 2022 1 M d'utilisateurs en 5 jours

janv. 2023 100 M d'utilisateurs

fév. 2026 900 M d'utilisateurs / semaine

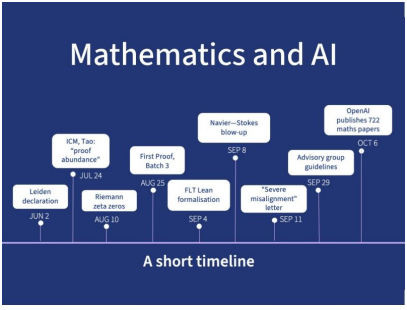
fin 2026 $\geq$ 1 milliard / semaine

Aux USA, 39 % des adultes ont utilisé l'IA générative **2 ans** après son lancement
PC : 20 % après 3 ans

Bick, Blandin & Deming, *The rapid adoption of generative AI*, 2024

LLMs & mathématiques : une accélération inouie

- 2021 GSM8K : problèmes niveau collège... souvent ratés
- 2022 Minerva : problèmes de compétition niveau lycée ($\sim$50 %)
- 2024 AlphaProof + AlphaGeometry 2 : **argent** aux Olympiades (IMO)
- 2025 **Or** à l'IMO, en langage naturel (DeepMind, OpenAI)
- 2025 FrontierMath (niveau recherche) : de $\sim$2 % à $>$50 % des problèmes résolus au moins une fois
- 2026 Problèmes d'Erdős résolus de manière autonome (Harmonic, DeepMind)
- 2026 Annonce d'OpenAI sur Navier–Stokes
  non vérifiée indépendamment
- 2026 722 articles de haut niveau [OpenAI]



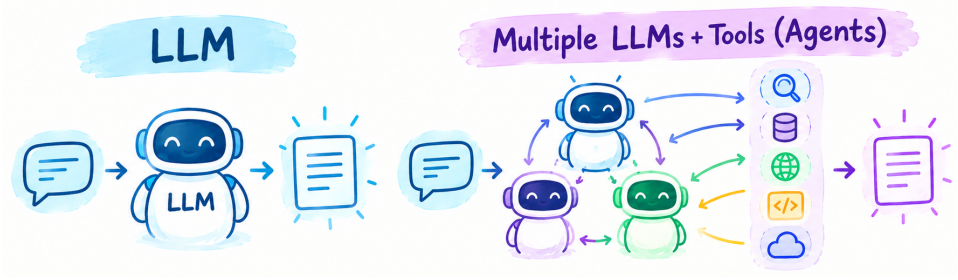
Academy for the Mathematical Sciences

De l'exercice de collège à la recherche ouverte en moins de 5 ans.

Les systèmes agentiques : une histoire très récente

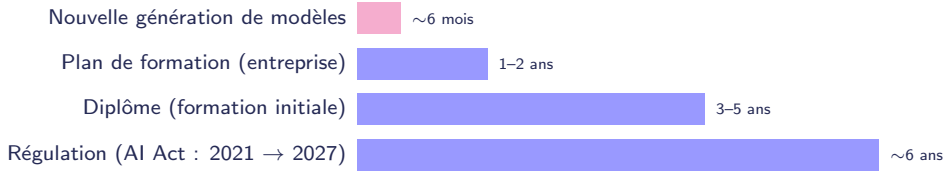
2023 Plugins, AutoGPT : démonstrations spectaculaires... peu fiables / pb sécurité
2024 Appels d'outils, protocole MCP (nov.)
2025 Agents de code, navigateurs agentiques
Q4 2025 **Agents fiables sur des tâches longues** ⇒ bascule des usages pro

Les mesures d'impact sur le travail (2023–2024, cf. *jagged frontier*) portent sur des **chatbots**.
⇒ Sur les **agents**, nous avons **moins d'un an de recul**.



Comment faire face à cette vitesse ?

Des horloges désynchronisées



- Former aux **principes**, pas aux outils
- Veille organisée (qui ? combien de temps ?)
- Expérimentations courtes & réversibles
- Chartes **datées** & révisées
- Retours d'expérience partagés
- Accepter de décider dans l'incertitude

La nouvelle vague robotique

Robots d'hier, robots d'aujourd'hui

Robotique industrielle (1980–2020)

- Chère : achat + intégration + cellule de sécurité
- Programmée point par point
- Peu adaptable : une tâche, un environnement

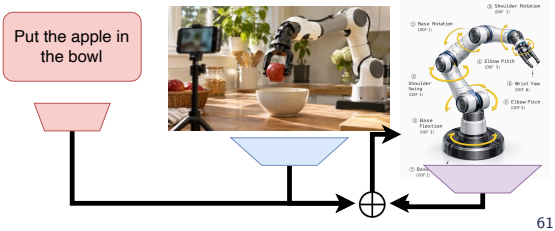
⇒ Gros volumes, chaînes à forte valeur ajoutée (automobile, électronique)



Nouvelle vague (2024–...)

- **Produite en masse** : bras, humanoïdes & quadrupèdes à quelques milliers d'euros [unitree]
- **Algorithmes adaptables** : instruction en langue naturelle, généralisation
- Apprentissage par démonstration [VLA]

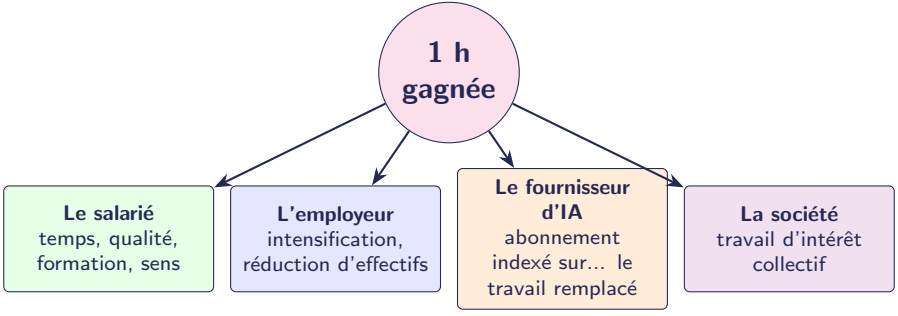
⇒ Petites séries, PME, logistique, agriculture, services



Les travailleurs face à l'IA

valeur, choix, psychologie, compétences

Du temps libéré... mais qui capte la valeur ?



- Le progrès technique ne partage pas ses gains **automatiquement** : c'est un rapport de force
  D. Acemoglu & S. Johnson, *Power and Progress*, 2023
- Quid d'une **réflexion sociale** sur la suppression des emplois ?
  dialogue social, CSE (consultation sur les nouvelles technologies), fiscalité...

Le marché visé : la masse salariale

Arthur Mensch (Mistral AI) – Assemblée nationale, mai 2026

Chez Mistral, le marché est estimé à **10 % de la masse salariale** *des entreprises. À l'échelle européenne, d'ici 3 à 6 ans, ce ratio représenterait de l'ordre de* **1 000 milliards d'euros**.

(propos rapportés – recopier le verbatim du compte rendu officiel de la commission d'enquête)

OpenAI / Anthropic : Pertes courantes = 2 à 3 fois le CA
⇒ Le prix des outils est *pour l'instant* décorrélé de leur coût

Ordre de grandeur

- Coût salarial moyen ≈ 50 k€/an
- 10–15 % ⇒ 5–7,5 k€/an/salarié
- ≈ 400–600 €/mois
  20 à 30× un abonnement grand public

- Valeur transférée **hors de l'entreprise**... et souvent hors d'Europe
- Qui négocie ce transfert ?
- Que devient le salarié dont 10 % du coût passe à l'IA ?
- Quelle fiscalité?

Utiliser l'IA : un choix personnel ?

L'exemple des enseignants

Si (presque) tous les étudiants utilisent l'IA, l'enseignant a-t-il **réellement le choix** de s'y mettre ou non ?

Évaluations, détection, conception des exercices : tout est déjà transformé.

Niveau	**Ce qui s'y décide**	**Exemples**
Individu	style, part de délégation, vérification	rédiger seul ou non, relire tout ou non
Équipe	normes & conventions	"tout code généré est relu par un pair"
Entreprise	outils autorisés, outils *mis à jour*, données, licences, chartes	Copilot, IA locale, *shadow AI*
Société	droit, régulation, protection sociale	AI Act, droit du travail, fiscalité

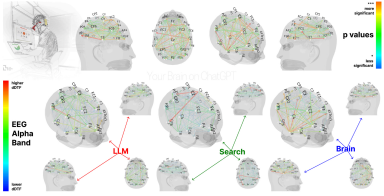
Frontière floue entre les choix individuels et ceux contraints par l'organisation.
Quelle possibilité de refuser l'IA? Quelle possibilité de l'utiliser en cas d'interdiction (shadow IA)

Impact psychologique du travail avec l'IA [NOUVEAU]

- Perte de sens : relire plutôt que produire, valider plutôt que décider
- Motivation : comment rester motivé quand une "calculatrice" fait la tâche ?
- Illusion de compétence : comprendre une solution $\neq$ savoir la produire
- Dette cognitive : moins d'engagement mesuré à l'EEG

  N. Kosmyna *et al.*, *Your Brain on ChatGPT*, 2025
- Anxiété & identité professionnelle : "à quoi je sers ?"
- Complaisance des modèles (*sycophancy*) : toujours d'accord avec vous

Ces axes marchent aussi pour les élèves, les étudiants, ...

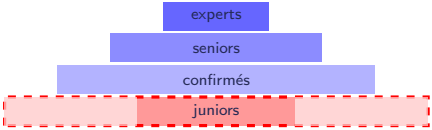




Compétences : les juniors en première ligne

- Les tâches confiées aux débutants sont les **plus automatisables**
- Les seniors transmettent **directives & spécifications**... aux LLMs plutôt qu'aux juniors
- Emploi des 22–25 ans dans les métiers les plus exposés (dév., service client) : **−13 %** (relatif) depuis fin 2022

Brynjolfsson, Chandar & Chen, *Canaries in the Coal Mine?*, Stanford 2025



Un bouleversement RH

Qui seront les seniors de demain ? Combien seront-ils ?
Quelles compétences auront-ils... ou n'auront-ils plus ?

On devient expert **en faisant** les tâches que l'on délègue aujourd'hui.

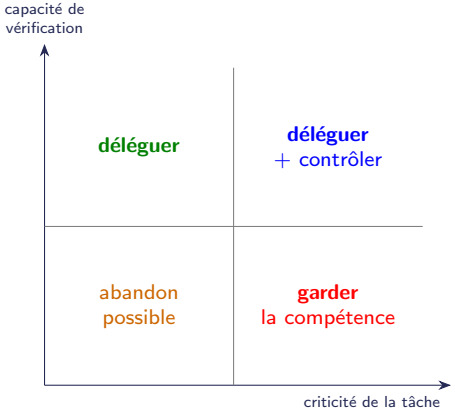
Choisir sciemment les compétences que l'on abandonne

Des précédents

- Calcul mental → calculatrice
- Orthographe → correcteur
- Orientation → GPS
- Mémorisation → moteur de recherche

Critères de choix

- Puis-je **vérifier** le résultat sans la compétence ?
- Que se passe-t-il en cas de **panne** ?
- Est-ce une compétence **fondatrice** pour d'autres ?

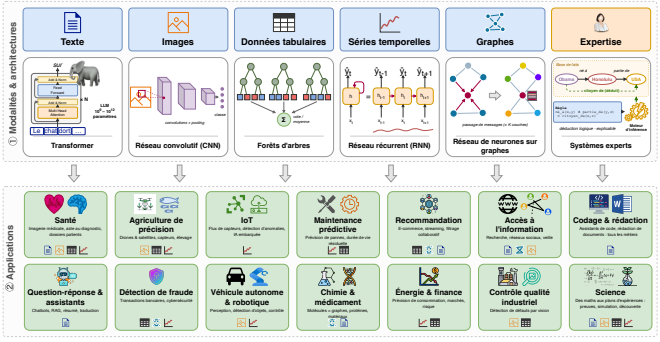


Comment ne pas subir l'oubli ?

Faire un choix explicite, métier par métier (et le réviser).

Relever le défi de la formation à l'IA

Former à l'IA : ne pas se tromper d'IA



(cf. panorama de l'introduction)

- IA $\neq$ chatGPT
- Prévision, classification, optimisation, vision, robotique...
- Beaucoup de métiers ont surtout besoin **d'apprentissage statistique classique** sur leurs propres données

Former au *prompting*, c'est former à **un outil**, pas à **un domaine**.

Former à l'IA : niveau & pérennité

Expert IA $\geq$ 1 an
adapter / entraîner des modèles, contraintes métier, systèmes hybrides

Développeur "léger" 1 semaine – 3 mois
scikit-learn, API, chaînes de traitement standards, protocoles d'évaluation

Utilisateur 2–4 h, puis pratique
interfaces conversationnelles, agents ; savoir vérifier, protéger ses données

- Chaque niveau a ses **risques propres** : sur-confiance (utilisateur), protocole biaisé (développeur), ...
- Une organisation a besoin des **trois**, en proportions très différentes
- Former un ingénieur $>$ 2020 = au moins niveau 2 en IA

Former à l'IA : niveau & pérennité

2022 *Prompt engineering*

2023 RAG, *fine-tuning* léger

2024 Multimodalité, appels d'outils, MCP

2025 Agents, modèles de raisonnement, harnais

2026 ...

Et c'est encore pire pour les experts en IA!

Demi-vie des savoir-faire $\approx$ 1 à 2 ans

- Ancrer la formation sur les **fondamentaux** (statistique, évaluation, données) : ils durent
- Prévoir du **temps de formation continue** dans les fiches de poste, y compris pour les experts
- Former **les formateurs**... en continu