

USAGE(S) DE L'INTELLIGENCE ARTIFICIELLE À AGROPARISTECH

7 Septembre 2026

Vincent Guigue

vincent.guigue@agroparistech.fr

<https://vguigue.github.io>

INTRODUCTION



Digital & Artificial Intelligence

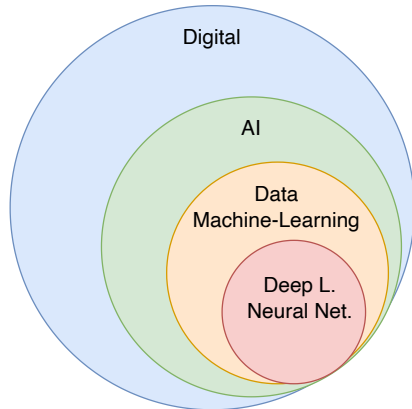
■ Two related but distinct concepts

■ AI: Different Definitions

1956 Any algorithm / program

1960-2012 Expert systems and logical reasoning

2012- Data & neural networks



A. Turing



Marvin Minsky

G. Hinton



Y. Lecun

Computer

1941

1956

Neural Networks

1986

Deep-learning

2012

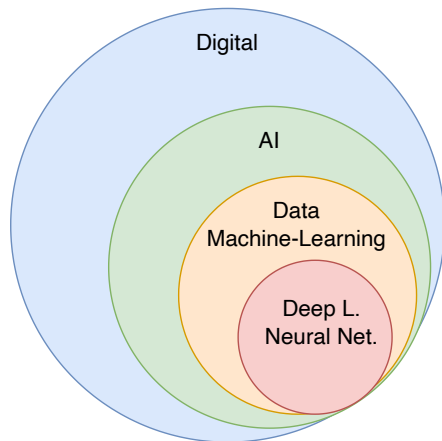
Computer-
Sciences

AI: wide variety of algorithms
Mainly : Expert System + Reasoning

AI= Neural Networks



AI's Place in the Digital Domain



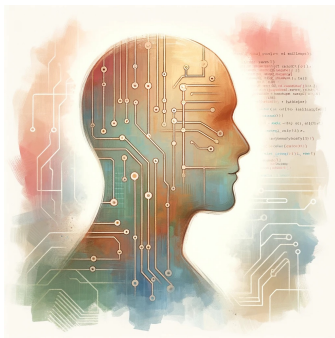
- Self-checkout at the supermarket, Google drive, SAP (basis)
- Google Maps, Knowledge bases, Expert systems
- Predictive systems (e.g., real estate market), recommendation
- Vision, chatbot, advanced systems

A lot of digital tools transform the job market...

⇒ A global view is required to understand the transformation



Artificial Intelligence & Machine Learning



Input ($\mathbf{x}$)		Output ($\mathbf{Y}$)	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

AI: computer programs that engage in tasks which, for now, are more satisfactorily performed by humans because they require high-level mental processes.

Marvin Lee Minsky, 1956

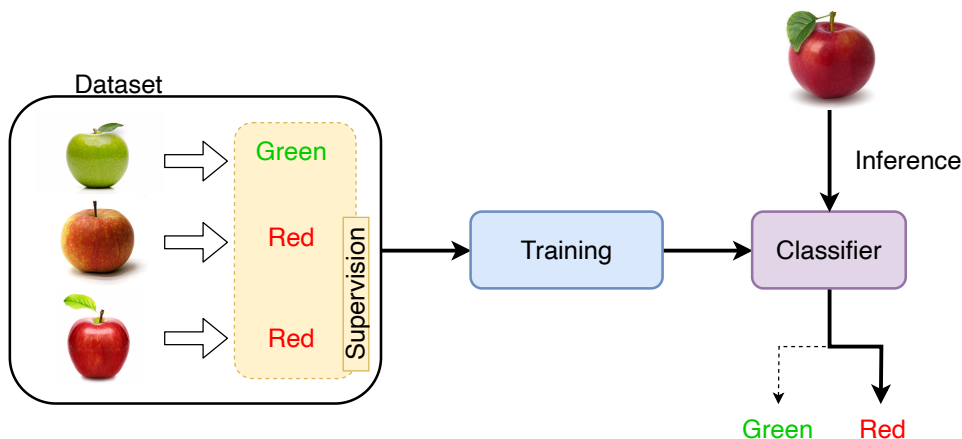
N-AI (Narrow Artificial Intelligence), dedicated to a single task

≠ G-AI (General AI), which replaces humans in complex systems.

Andrew Ng, 2015



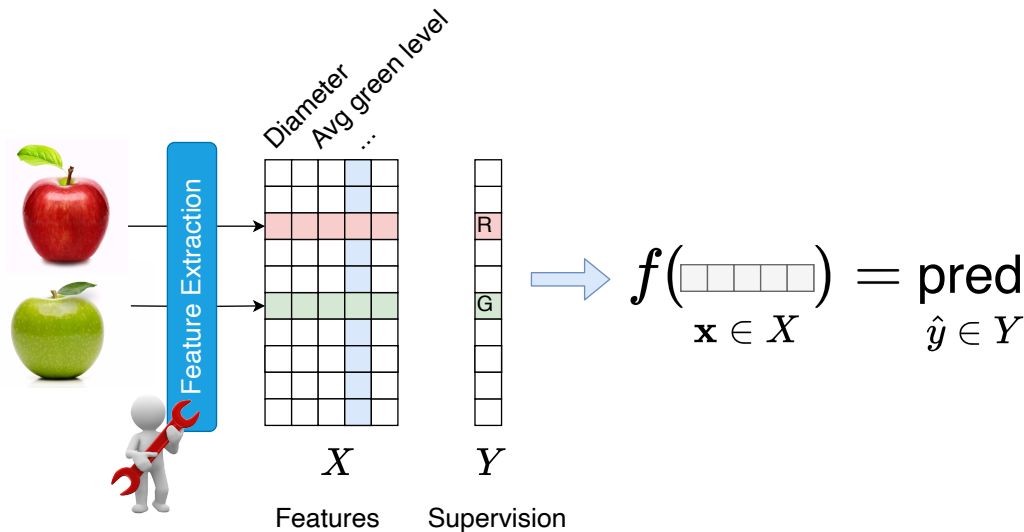
Supervised Processing Chain & Models



- Promise = building a model *solely* from observations



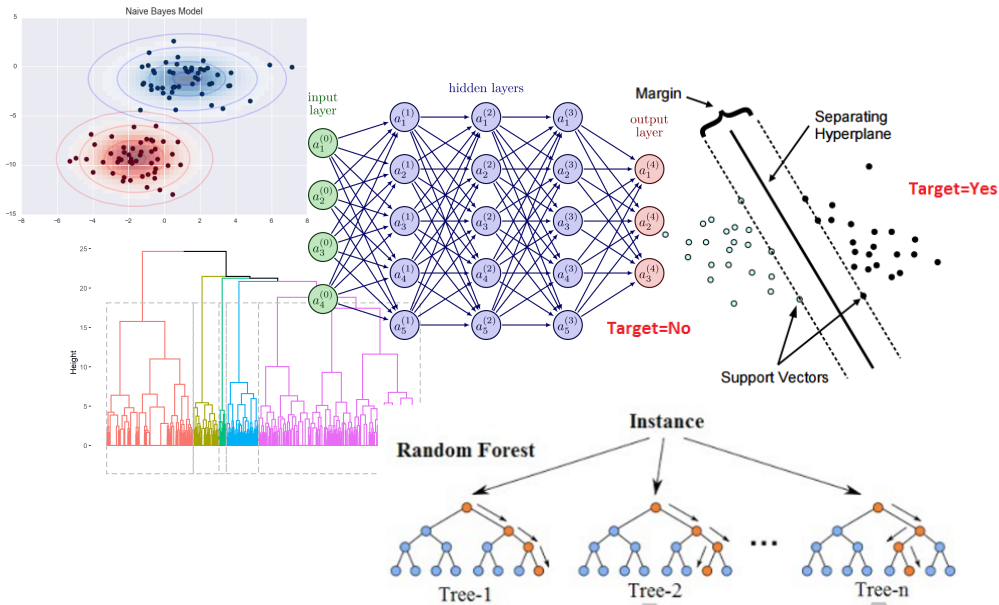
Supervised Processing Chain & Models



- Think about the feature space = observation description
- ... And about required supervision

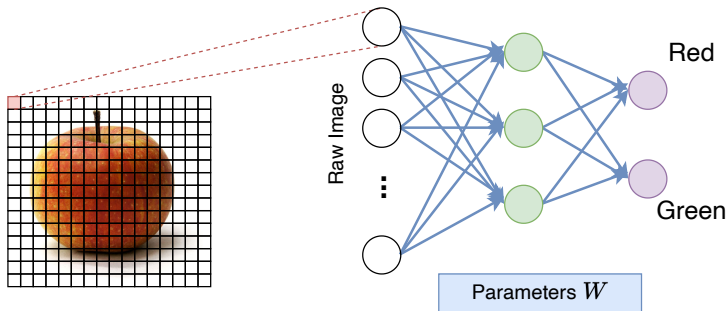


Supervised Processing Chain & Models





Supervised Processing Chain & Models

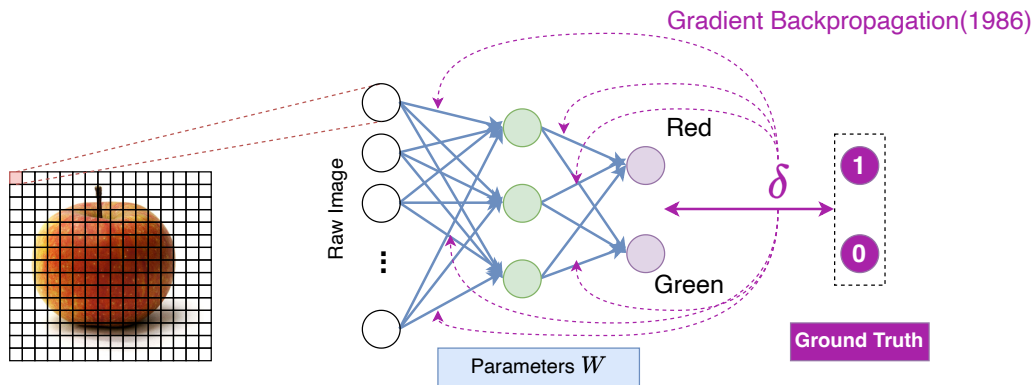


■ Random initialization...

And random decision-making (at first!)



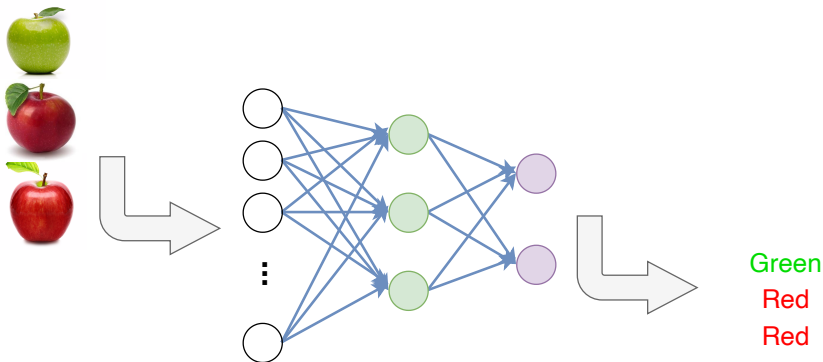
Supervised Processing Chain & Models



- Updating the weights
- Epsilon-sized steps, many iterations over the data



Supervised Processing Chain & Models

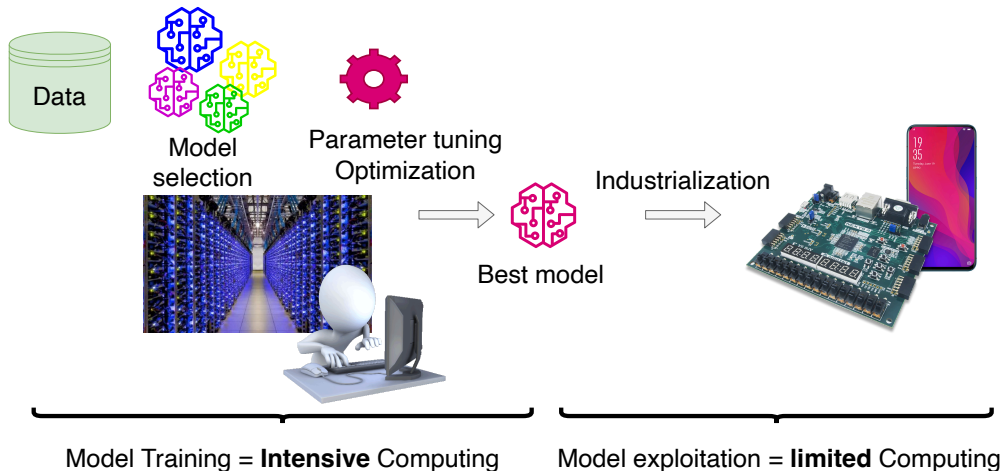


- **Training** is slow and costly
- **Inference** is (much) faster



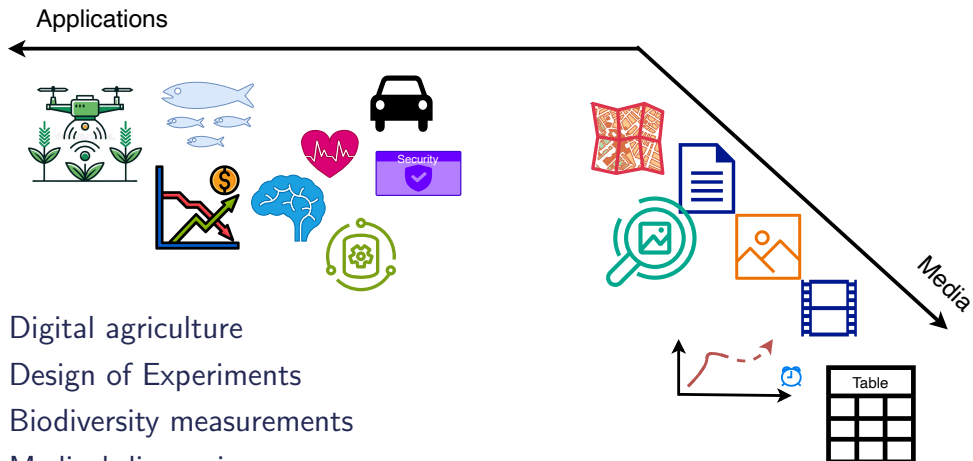
Supervised Processing Chain & Models

Clearly separate the different steps in machine-learning





What can we do with AI?

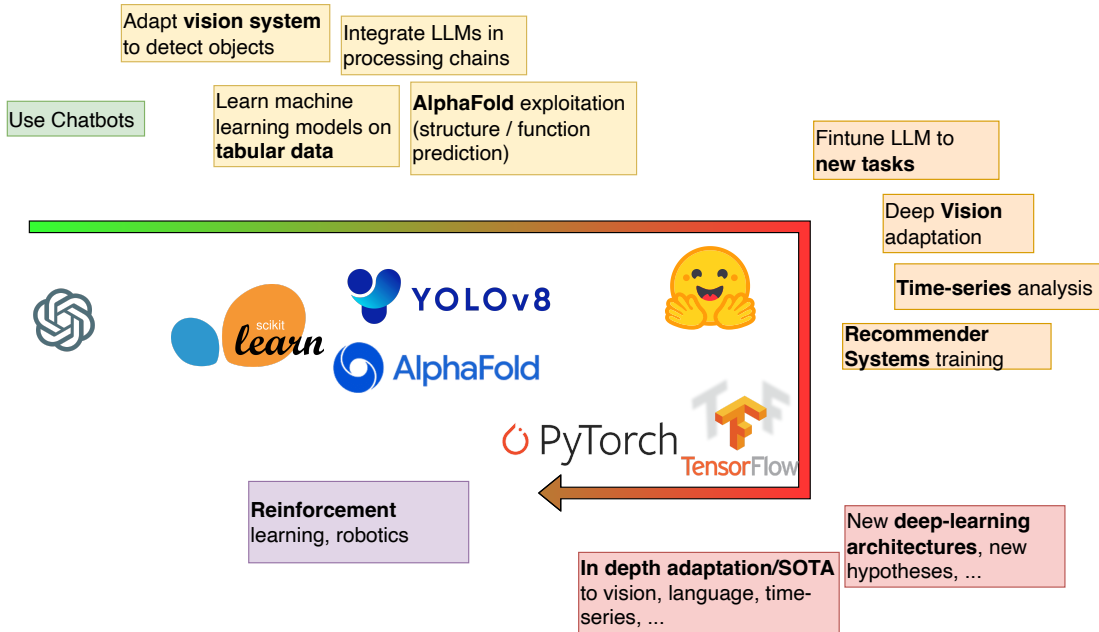


- Digital agriculture
- Design of Experiments
- Biodiversity measurements
- Medical diagnosis
- Stock exchange prediction
- Exploitation optimization
- ...

- Image, text, time-series, movies, maps
- Table of data

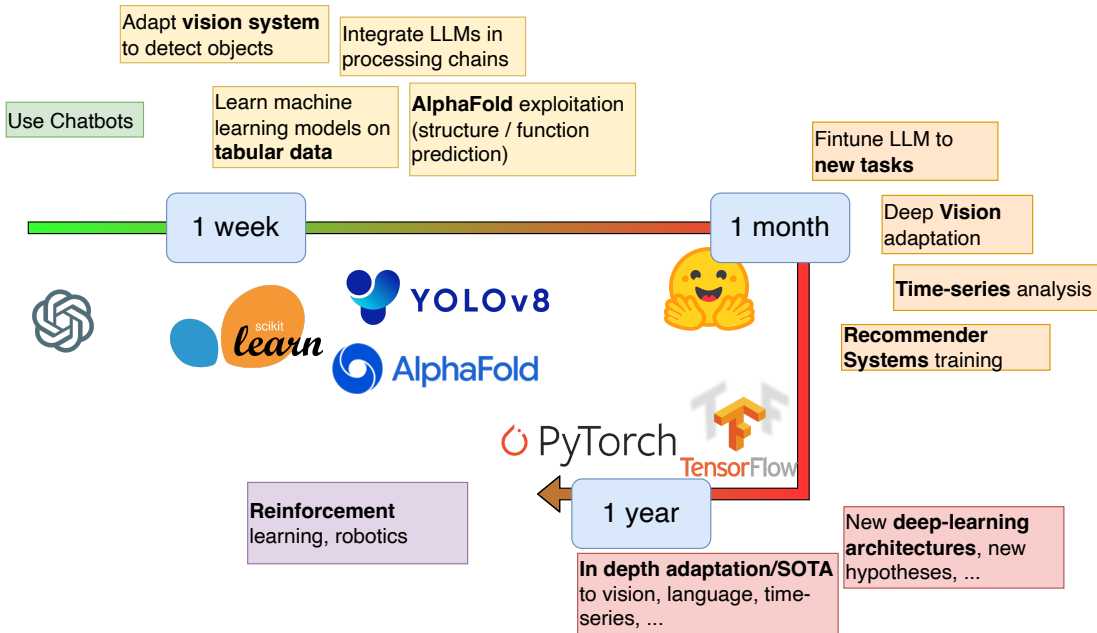


What is the cost to access AI (today)?



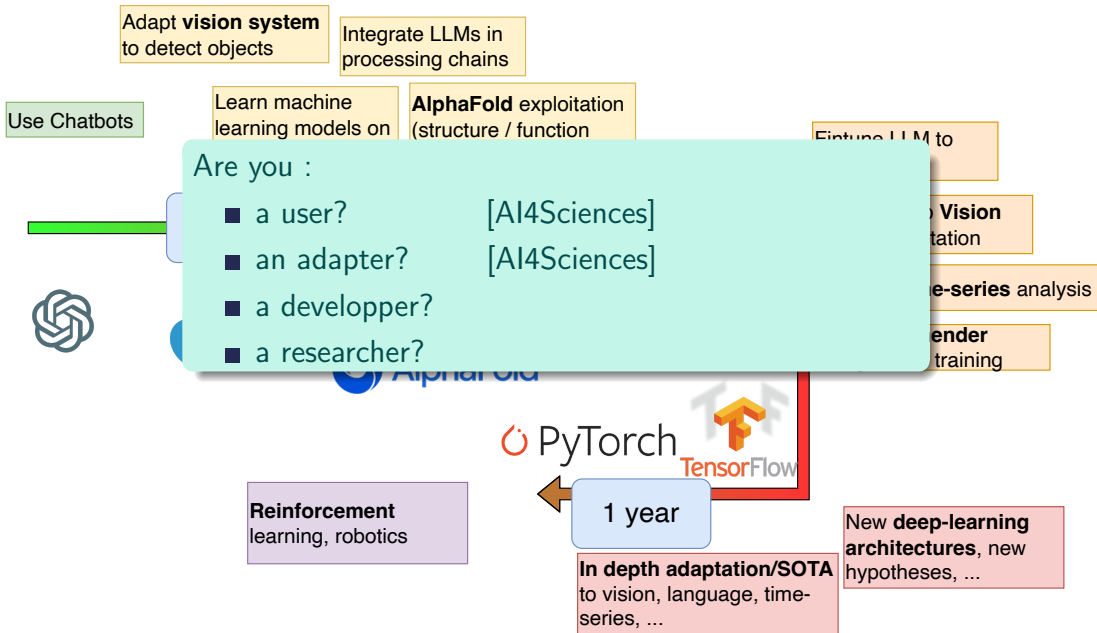


What is the cost to access AI (today)?





What is the cost to access AI (today)?





What is the cost of AI?



Sensor monitoring,
basic decision

Autonomous
driving system

Local platform to
build & deploy AI
models



Answering LLM
calls



Training a vision
system

Training an LLM

0 - 10 W

20 - 150 W

0.2 - 1.5 kW

1 - 5 kW

0.5 - 50 MW

■ Electricity, water, rare-earth elements, CO₂ emissions, financing costs, etc.

⇒ Most costs scale proportionally



So, let's stop speaking of AI !

Machine-learning

- Easy to handle, cheap to compute
- Many (many) applications
- (often) Provide strong system optimization
- Should be part of the curriculum for all engineering students

Deep-learning

- Tailor made system / haute-couture for numerical data
- Basic entry for semantic data (text, image, voice, users' traces, ...)
- Multi-modal systems, new paradigm (e.g. self-training)

LLM

- ML systems
 $\propto 1\text{k params}$
DL systems $\propto 1\text{M}$
LLM systems $\propto 10^3\text{M}$
- New applications, new interfaces to existing systems...
- Major societal impact (education, jobs market, information access, ...)

⇒ Choose your keywords more carefully!

FROM LANGUAGE MODELS TO CHATGPT

NOVEMBER 30, 2022

1 MILLION USERS IN 5 DAYS

100 MILLION AT THE END OF JANUARY 2023

1.16 BILLION IN MARCH 2023



A very quick tour of NLP & machine-learning history

1950-2022 : AI = **Expert** knowledge (& know-how) $\Rightarrow$ automated systems

Building an opinion mining system:

1 What would a linguist search for?

2 How to generalize?

3 Implement with fast regex

- Gather documents
- Analyze
- Count

* $\longrightarrow$ {like, love, appreciate} $\rightarrow$ * $\rightarrow$ #product

* $\rightarrow$ {didn't, not, doesn't, don't} $\rightarrow$ {like, love, appreciate} $\rightarrow$ * $\rightarrow$ #product

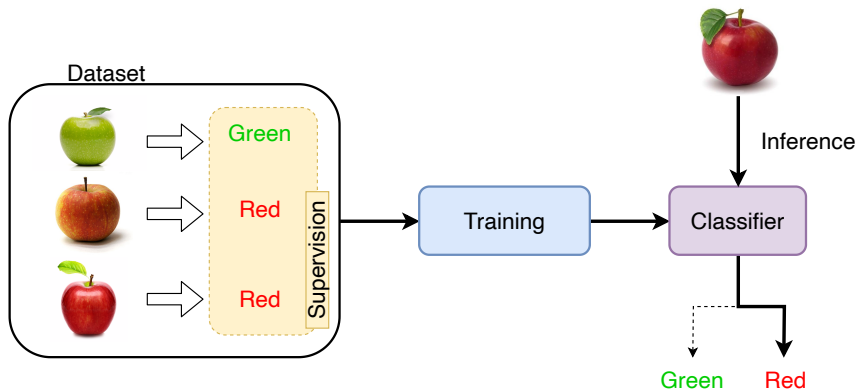
* $\longrightarrow$ {hate, loathe, detest} $\rightarrow$ * $\rightarrow$ #product



A very quick tour of NLP & machine-learning history

1950-2022 : AI = **Expert** knowledge (& know-how) $\Rightarrow$ automated systems

1980-2010 : How to learn directly from **annotated data**?

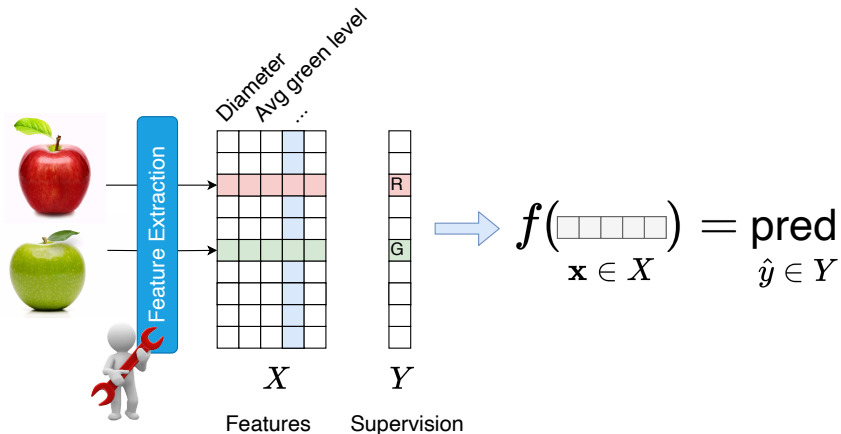




A very quick tour of NLP & machine-learning history

1950-2022 : AI = **Expert** knowledge (& know-how) $\Rightarrow$ automated systems

1980-2010 : How to learn directly from **annotated data**?





1980-2010 : How to learn directly from **annotated data**?



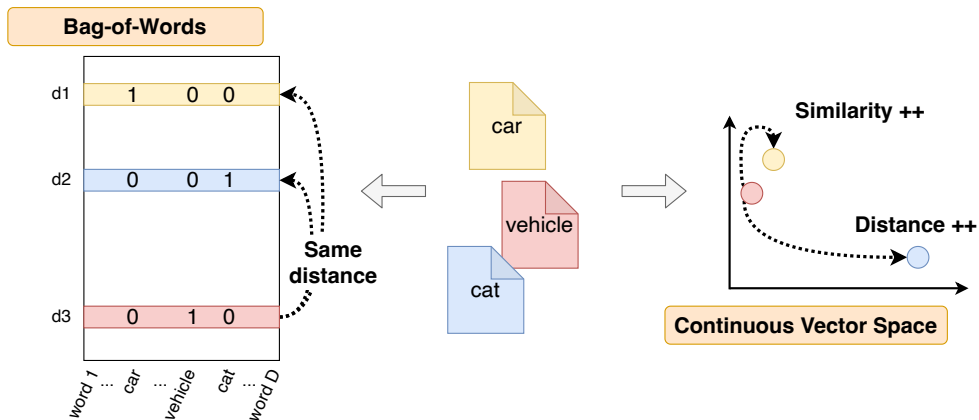


A very quick tour of NLP & machine-learning history

1950-2022 : AI = **Expert** knowledge (& know-how) $\Rightarrow$ automated systems

1980-2010 : How to learn directly from **annotated data**?

1910-2015 : **Representation learning** $\Rightarrow$ data in continuous latent space





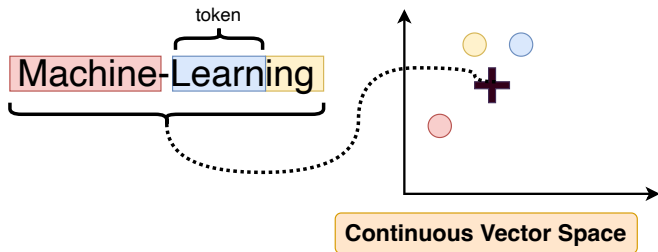
A very quick tour of NLP & machine-learning history

1950-2022 : AI = **Expert** knowledge (& know-how) $\Rightarrow$ automated systems

1980-2010 : How to learn directly from **annotated data**?

1910-2015 : **Representation learning** $\Rightarrow$ data in continuous latent space

Word Piece statistical split





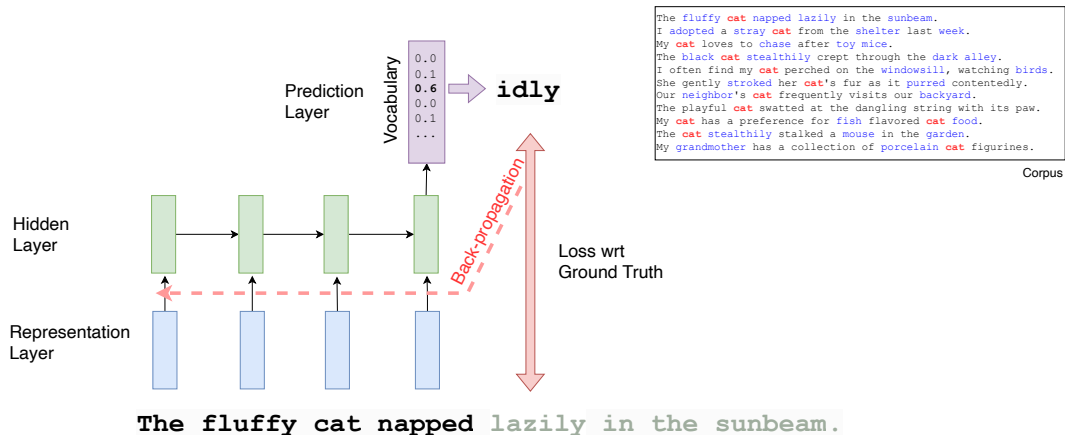
A very quick tour of NLP & machine-learning history

1950-2022 : AI = **Expert** knowledge (& know-how) $\Rightarrow$ automated systems

1980-2010 : How to learn directly from **annotated data**?

1910-2015 : **Representation learning** $\Rightarrow$ data in continuous latent space

2015-... : **Language models** + pre-training





A very quick tour of NLP & machine-learning history

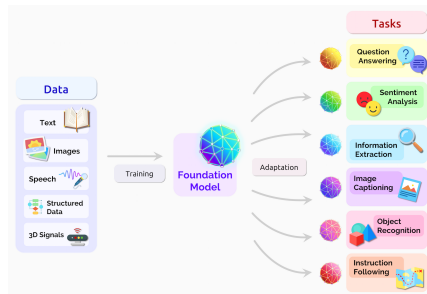
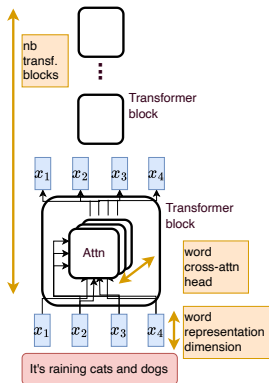
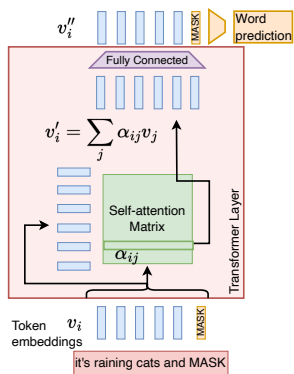
1950-2022 : AI = **Expert** knowledge (& know-how) $\Rightarrow$ automated systems

1980-2010 : How to learn directly from **annotated data**?

1910-2015 : **Representation learning** $\Rightarrow$ data in continuous latent space

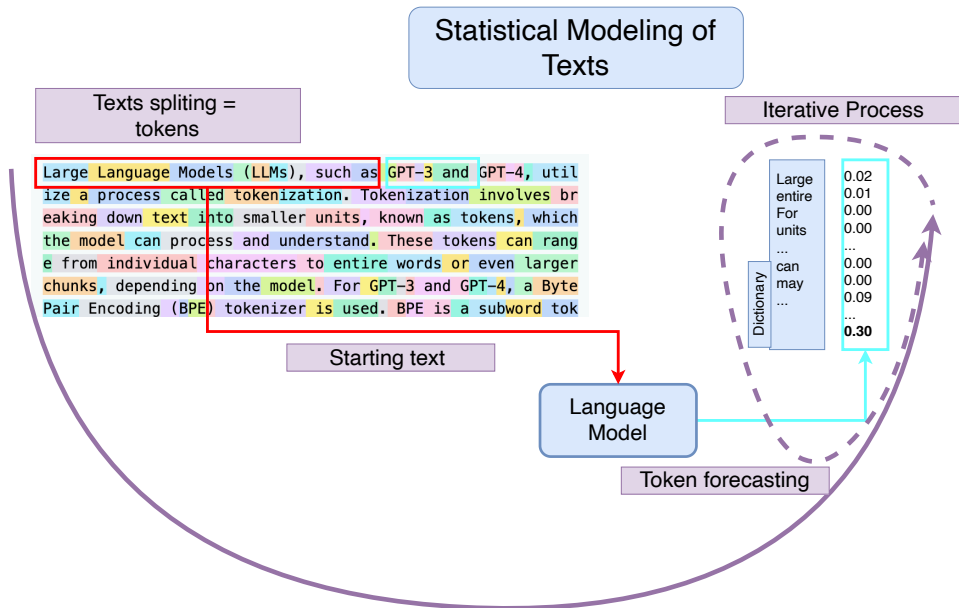
2015-... : **Language models** + pre-training

2017-... : **Transformer**, multimodal foundation models





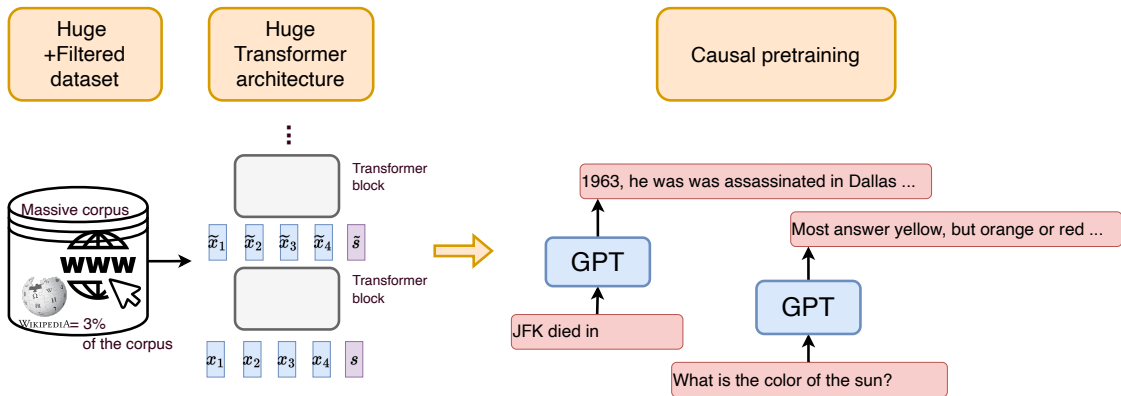
A (very smart) stochastic parrot!





The Ingredients of chatGPT

0. Transformer + massive data (GPT)



- Grammatical skills: singular/plural agreement, tense concordance
- Knowledge: entities, names, dates, places



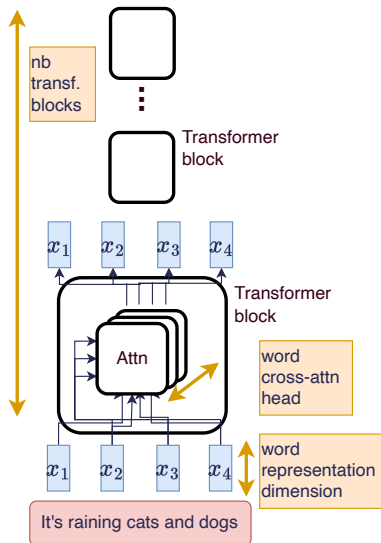
The Ingredients of chatGPT

1. More is better! (GPT)

- + more input words [500 $\Rightarrow$ 2k, 32k, 100k]
- + more dimensions in the word space [500-2k $\Rightarrow$ 12k]
- + more attention heads [12 $\Rightarrow$ 96]
- + more blocks/layers [5-12 $\Rightarrow$ 96]

175 Billion parameters... What does it mean?

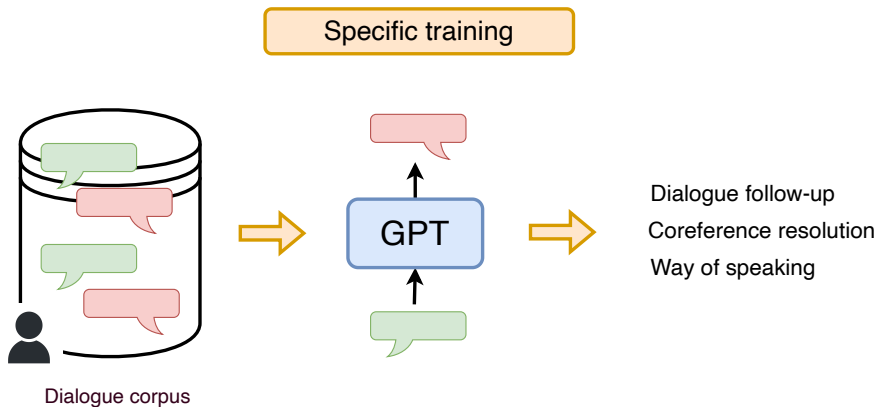
- $1.75 \cdot 10^{11} \Rightarrow 300 \text{ GB} + 100 \text{ GB}$ (data storage for inference) $\approx 400\text{GB}$
- NVIDIA A100 GPU = 80GB of memory (=20k€)
- Cost for (1) training: 4.6 Million €





The Ingredients of chatGPT

2. Dialogue Tracking



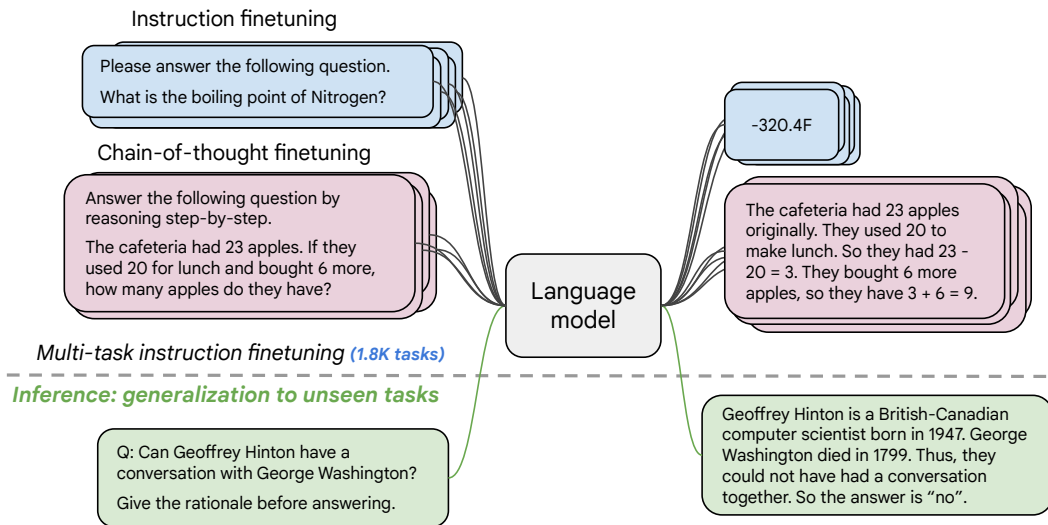
■ **Very clean data**

Data generated/validated/ranked by humans



The Ingredients of chatGPT

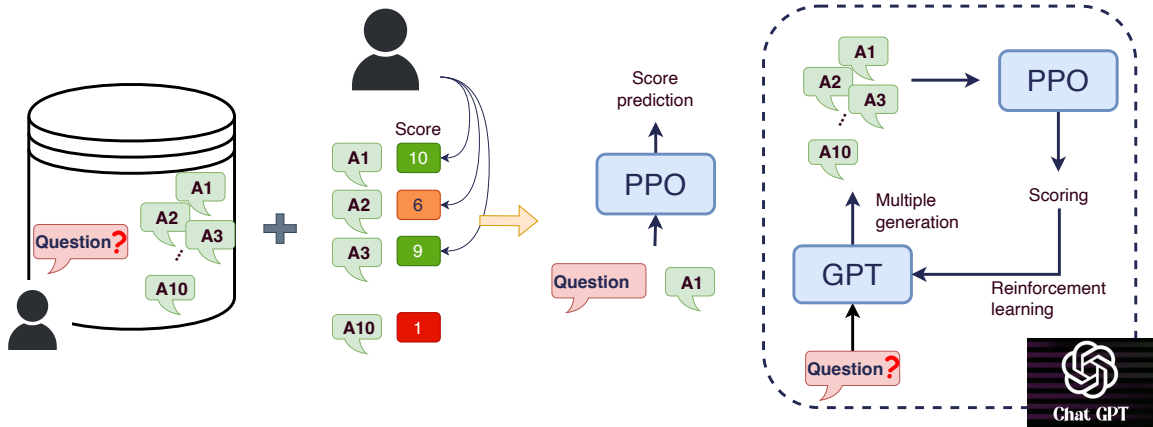
3. Fine-tuning on different ($\pm$) complex reasoning tasks





The Ingredients of chatGPT

4. Instructions + answer ranking



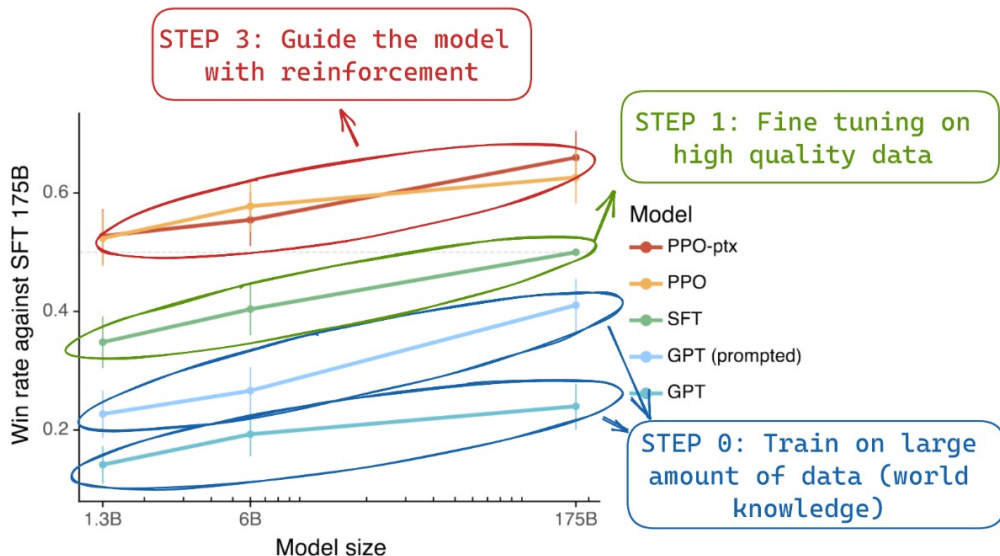
- Database created by humans
- Response improvement

- ... Also a way to avoid critical topics = censorship



Steps & Performance

Massive data $\Rightarrow$ HQ data (dialogue) $\Rightarrow$ Tasks $\Rightarrow$ RLHF





Usage of chatGPT & Prompting

- Asking chatGPT = skill to acquire $\Rightarrow$ *prompting*
 - Asking a question well: ... *in detail*, ... *step by step*
 - Specify number of elements, e.g. *3 qualities for ...*
 - Provide context: *cell* for a biologist / legal assistant

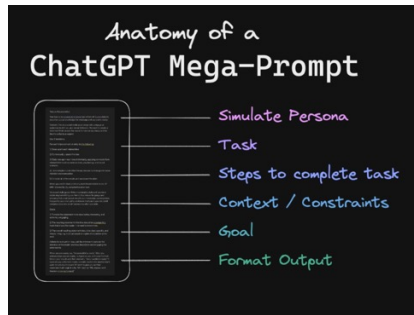
- Don't stop at the first question

- Detail specific points
- Redirect the research
- Dialogue

- Rephrasing

- Explain like I'm 5, like a scientific article, bro style, ...
- Summarize, extend
- Add mistakes (!)

$\Rightarrow$ Need for **practice** [1 to 2 hours], discuss with colleagues



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



Usage of chatGPT & Prompting

■ Asking chatGPT = skill to acquire ⇒ *prompting*

- Asking a question well: ... *in detail*, ... *step by step*
- Specify number of elements, e.g. *3 qualities for ...*

■ **Prompt = specification document:** long, detailed, designed to remove ambiguity about intended directions.

■ Prompting requires

domain expertise >> general prompting knowledge.



■ Don't

■ Rephrase

- Add mistakes (!)

⇒ Need for **practice** [1 to 2 hours], discuss with colleagues

Anatomy of a

pt

sona

plete task
constraints

at

-chatgpt-prompt/



Language Handling

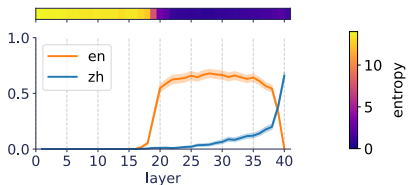
- Language models are multilingual today:

⇒ Stay in the language you are comfortable with

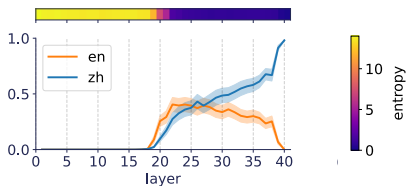
⇒ Ask for answers in any language

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

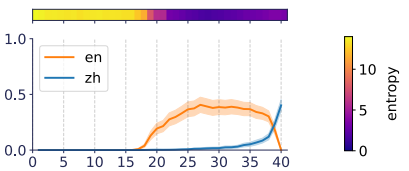
(a) Translation task



(b) Repetition task



(c) Cloze task

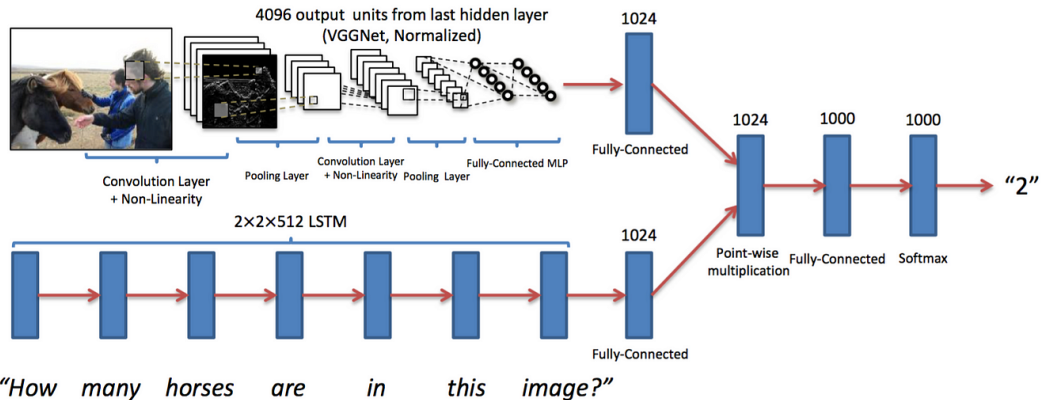




GPT4 & Multimodality

Merging information from text & image. **Learning** to exploit information jointly

The example of VQA: visual question answering



⇒ Backpropagate the error ⇒ modify word representations + image analysis

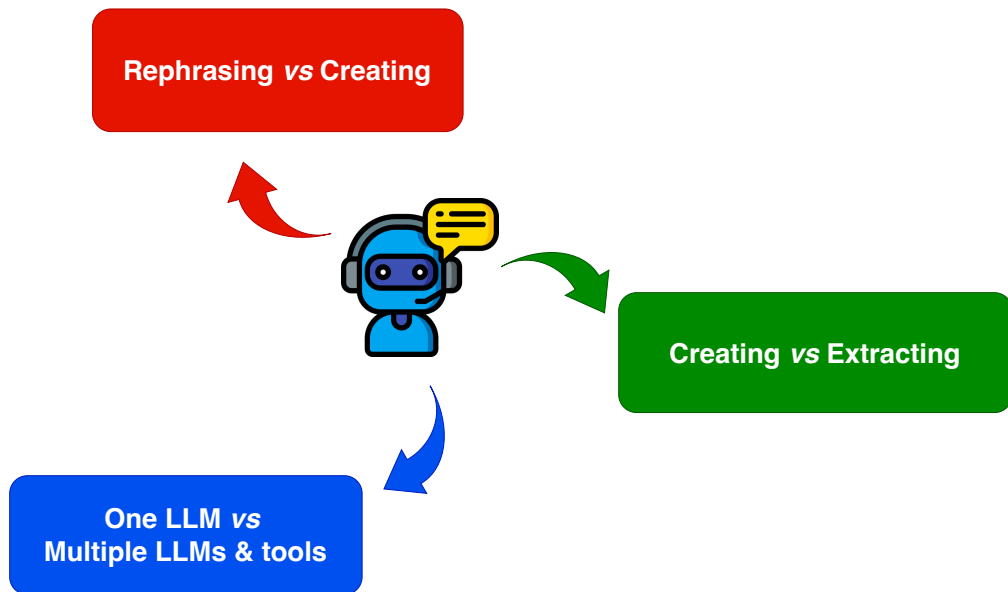


VQA: Visual Question Answering, arXiv, 2016, A. Agrawal et al.

LLM TYPOLOGY OF USES



Three questions to frame the use cases



(1) Rephrasing vs creating content

Rephrasing *I bring the content*

- **Translating/Proofreading** a paragraph
- **Rephrasing**: style, tone, 10% shorter
- A paragraph from **my** bullet points
- Meeting minutes from **my** notes
- Cover letter from a CV + job offer
- Table formatting, OCR, $\text{\LaTeX}$ /Excel
- The abstract of **my own** paper

Creating *the model brings the content*

- **Finding** inspiration [blank page]
- **Suggesting** a plan, an angle, an acronym
- **Answering** a question I cannot answer
- **Testing** an idea, listing weak points
- Writing a paragraph on a topic I give *but do not detail*
- Writing a review, a related-work section





(1) Rephrasing: examples

Building a recommendation letter

Prompt

[Tâche]
Etudiant rencontré...
qualités ...
résultats marquant



LLM



Meeting minutes



Transcription



Résumé/CR





(1) Ethical questions

Rephrasing

The content is **mine**. The model touches the *form*.

→ Am I still the author? **yes**

→ Do I have to disclose it?
like a spell-checker? a translator?

→ Real risk = **silent distortion**
a nuance lost, a hedge removed

→ Real risk = **confidentiality**
unpublished text sent to a server

Creating

The content comes from **the model**, *i.e.* from a corpus I did not choose.

→ Am I still ~~the author~~ responsible? **yes**

→ Who do I cite?
the model has no source

→ Real risk = **hallucination**
plausible, well written, false

→ Real risk = **uniformity of thought**
the most likely idea, not the best one

⇒ Same tool, same prompt box, **two completely different debates.**



(2) Creating *ex-nihilo* vs extracting information

Ex-nihilo

parametric memory

- The answer is **in the weights**
- *JFK died in...* $\Rightarrow$ probabilistic answer
- No source, no traceability
- Frozen at training time

$\Rightarrow$ **hallucination** (*factuality* measurement)

Extraction

the document is given

- The answer is **in the input**
- The model *locates*, it does not invent
- The source is known *by construction*
- Up to date = *as fresh as the document*

$\Rightarrow$ **verifiable** (*faithfulness* measurement)





(2) Creating *ex-nihilo* vs extracting information

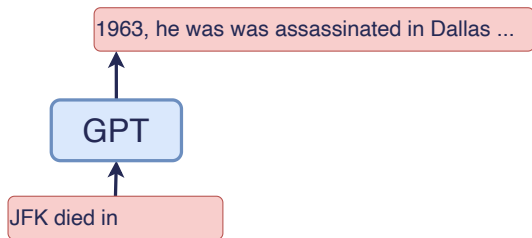
Ex-nihilo

parametric memory

Extraction

the document is given

- The answer is **in the weights**
 - *JFK died in...* ⇒ probabilistic answer
 - No source, no traceability
 - Frozen at training time
- ⇒ **hallucination** (*factuality* measurement)



- The answer is **in the input**
 - The model *locates*, it does not invent
 - The source is known *by construction*
 - Up to date = *as fresh as the document*
- ⇒ **verifiable** (*faithfulness* measurement)

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

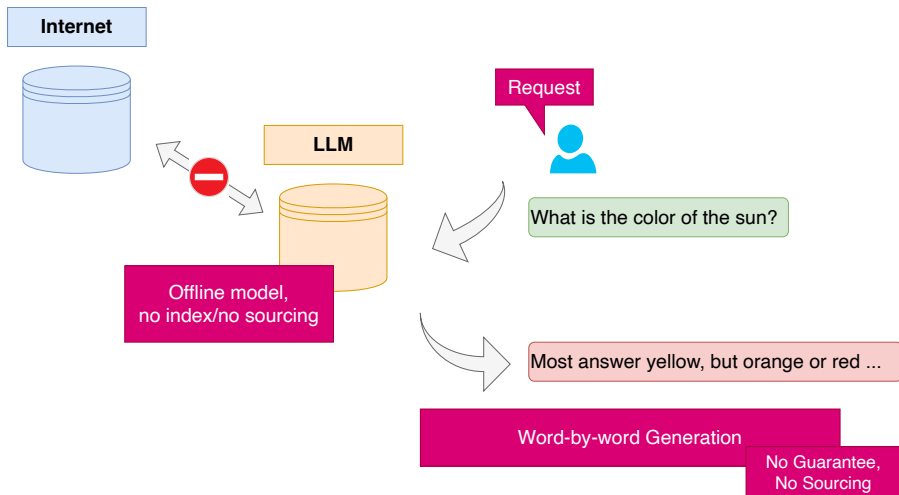
⇒ Same question, same interface... but only one of them can be checked.



(2) LLMs $\Rightarrow$ RAG: memory vs information extraction

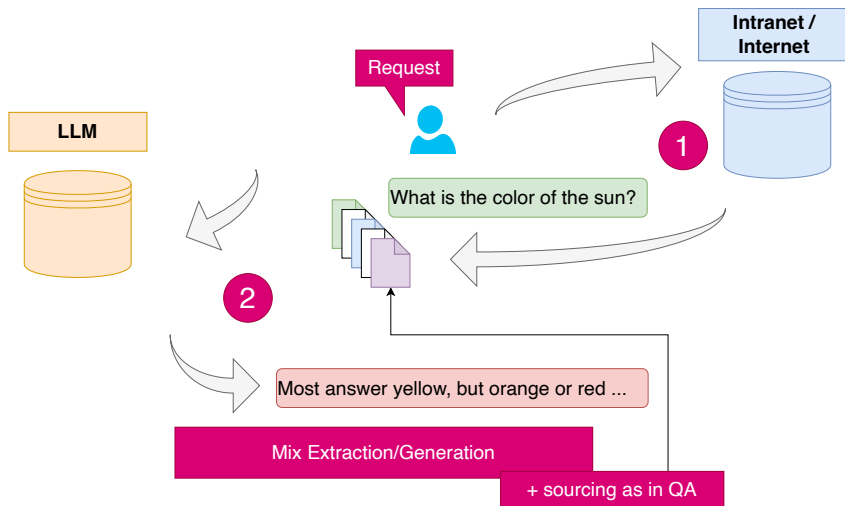
- Asking ChatGPT questions... A surprising use case!
- But is it reasonable?

[A true open question (!)]





(2) LLMs $\Rightarrow$ RAG: memory vs information extraction



- RAG: Retrieval-Augmented Generation
- Current limit on input size (2k, 32k, 200k tokens)



(2) In a scientific context

■ RAG in a scientific context

should be called **bibliography**!

■ Consensus, Scopus.ai, opscidia, ...

⇒ great summary + analysis... On which articles?

■ NotebookLM: choose your articles (up to 50)...

Then start the discussion

⇒ analysis, topic clustering, comparison, ...

■ If you want to dialogue with a manuscript...

It is too long for an LLM

⇒ RAG again!

Can I be a (good) scientist without AI watch?
Is my job at risk?



Consensus

Scopus[®] AI

Change the way you view knowledge

OPSCI DIA

The value of knowledge

NotebookLM

Think Smarter, Not Harder

Try NotebookLM



(3) From LLMs to agents

LLM

a tool I hold

Agent

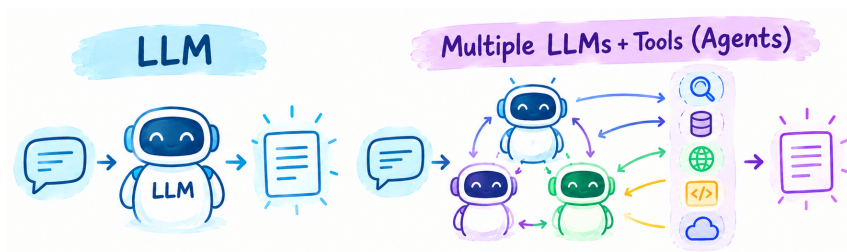
a system that acts

- One question → one answer
- I read, I judge, I keep or discard
- The model produces **text**
- No effect on the world
- Efficient on small tasks

⇒ human validation **at every step**

- One goal → a **chain** of actions
- The model calls **tools**, loops, retries
- The model produces **actions**
queries, files, emails, code run
- Effects on the world

⇒ human validation **at the end**, if at all





(3) Toolformer: When the LLM calls upon tools

The LLM [2023]:

- 1 Identifies its own weaknesses
- 2 Calls tools/APIs to provide better answers

⇒ Controlled data sources (SQL, Wikipedia) = RAG++; Calculator; Translator; Specialized compute engine

LLM at the heart of the system

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.

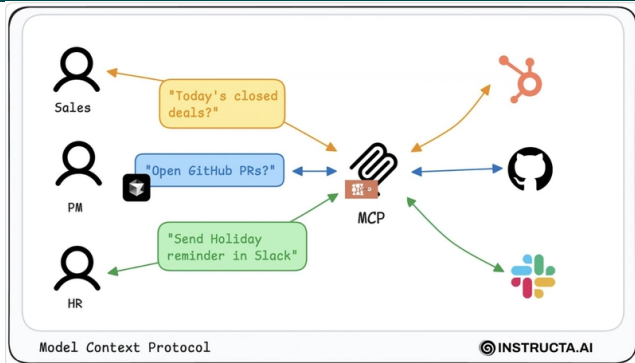
Out of 1400 participants, 400 (or [Calculator(400 / 1400) → 0.29] 29%) passed the test.

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.



MCP Protocol



- Interfacing with LLMs (remote or local)
- Link with applications (emails, websites, messaging, calendar)

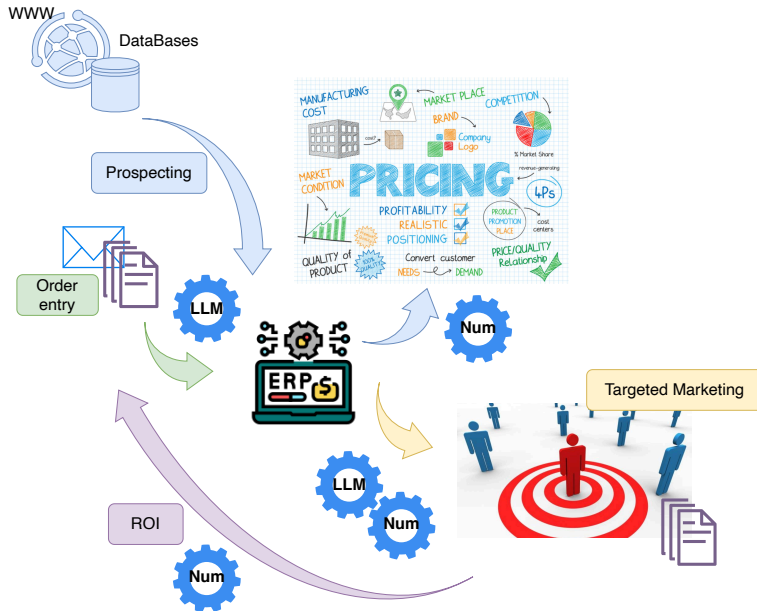
⇒ LLM Interfacing = the technical foundation for agents... and many other applications

MCP = using an API in natural language

Openclaw: wrap many computer functions in an LLM callable API (mail, browser, console, ...)



Integration into a Complex System

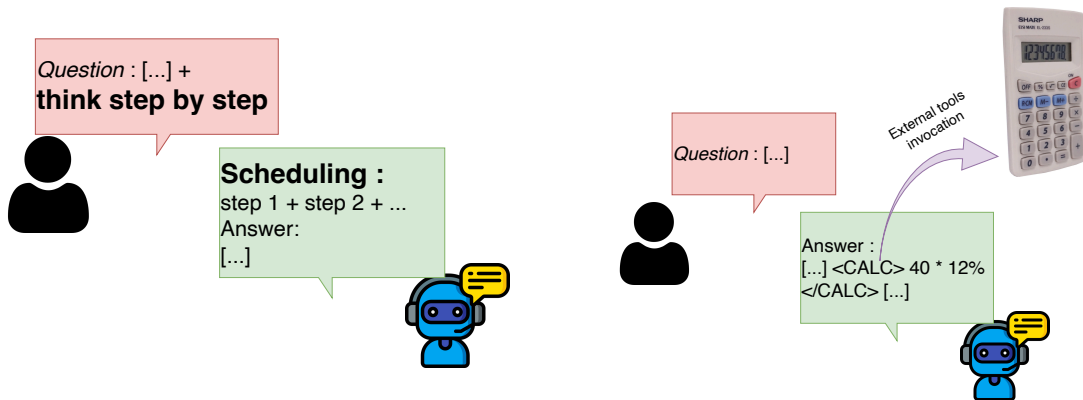


- Prospect list
- Email processing, needs identification
- ERP interaction $\Rightarrow$ proposal databases
- Pricing
- Drafting a commercial proposal



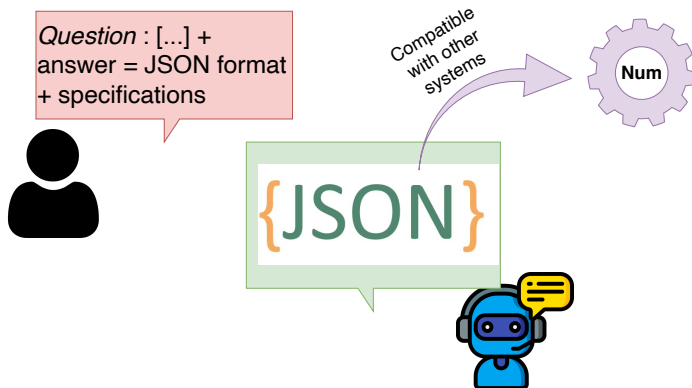
From *Chain-of-Thought* to Actions

1 Chain-of-Thought / Toolformer



From *Chain-of-Thought* to Actions

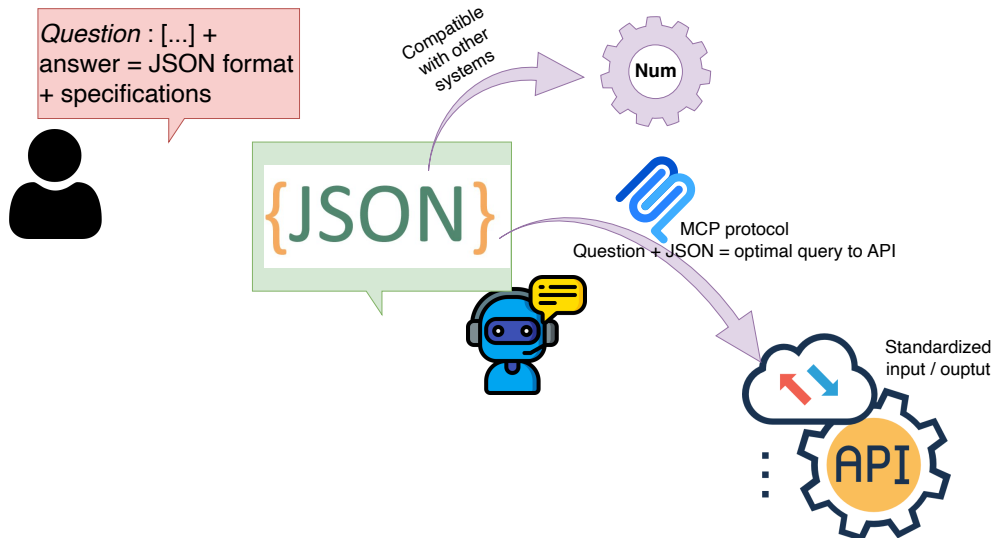
- 1 Chain-of-Thought / Toolformer
- 2 From text answers to structured information





From *Chain-of-Thought* to Actions

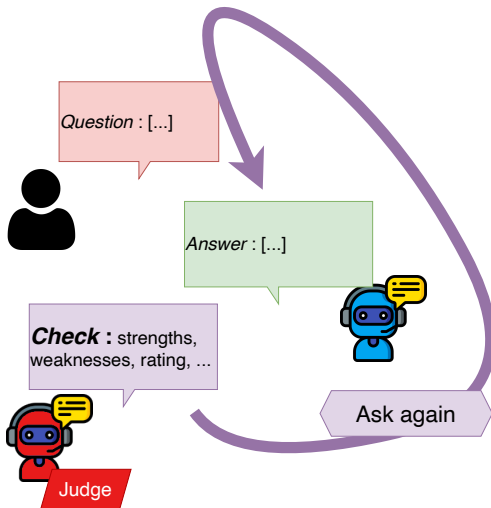
- 1 Chain-of-Thought / Toolformer
- 2 From text answers to structured information





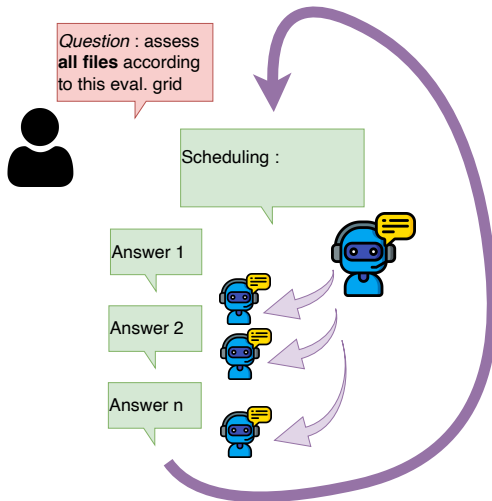
From *Chain-of-Thought* to Actions

- 1 Chain-of-Thought / Toolformer
- 2 From text answers to structured information
- 3 LLM as a judge / reflexivity + improvement loop



From *Chain-of-Thought* to Actions

- 1 Chain-of-Thought / Toolformer
- 2 From text answers to structured information
- 3 LLM as a judge / reflexivity + improvement loop
- 4 Parallelism





-
- The diagram illustrates the Multi-Turn architecture. A central blue robot icon represents the AI agent. It is connected to a user (black silhouette) on the left, who provides a "Question : [...]" and "Check specifications". The AI agent interacts with a "Scheduling" module (green box) containing "Tasks", "Ressources", and "...". The AI agent also interacts with two "MCP" (Model Control Plane) components, each represented by a blue paperclip icon. The top MCP connects to "Files", "Mails", and "Terminal". The bottom MCP connects to "Web", "BD", and "Specific sources". The AI agent also interacts with "Local assessments" (red robot icon) and "Multi-Turn" (purple box). The entire process is labeled "Multi-Turn" at the bottom.

LARGE LANGUAGE MODELS & TEACHING

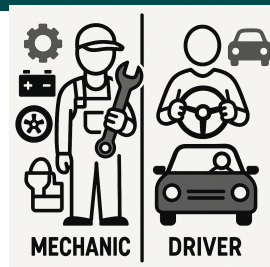


Two distinct questions

1 Teaching **with** AI

2 Teaching AI

⇒ Anyway, you have to know things about AI!



Teaching **of** AI

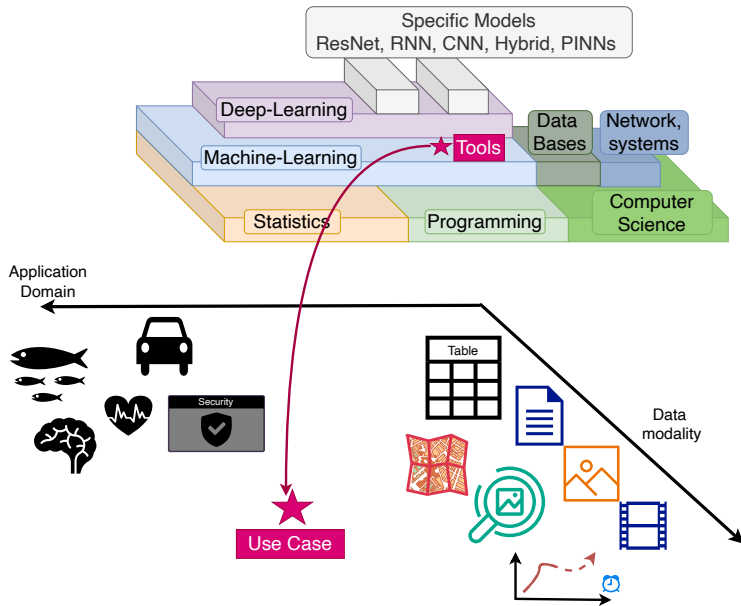
- Programming / Statistics
- Machine Learning: models + protocols
- Deep Learning
- Data modalities: images, texts, time series
- Specific domains: biology, medicine, finance, engineering, ...
- Theoretical deepening: opti., confidence int., convergence, ...

Teaching **with** AI

- New opportunities
- New risks
- New constraints



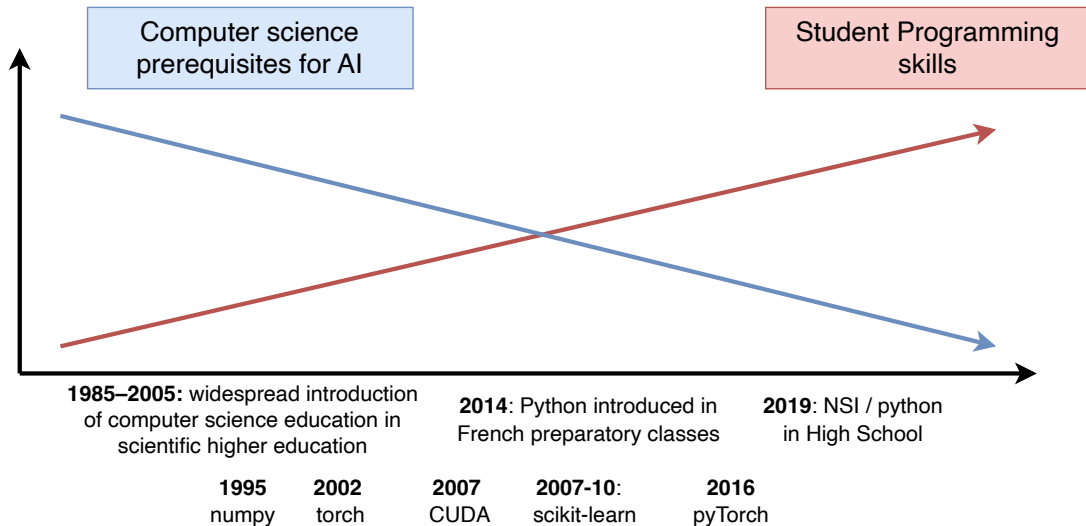
Teaching AI



- Different levels of access (awareness, tool usage, development)
- Different data
- Different application domains

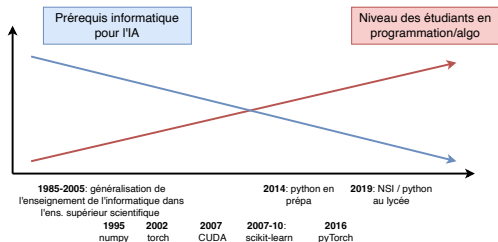


Access to AI: At a Crossroads





Access to AI: At a Crossroads



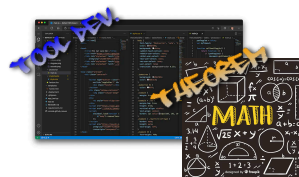
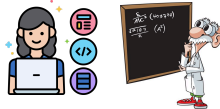
1



2



3



Three levels of access to AI:

- 1 **Leverage** a chatbot... in an **optimal** and **responsible** way
- 2 **Use** tools, manipulate data
- 3 **Develop** tools



LLMs: a pedagogical tool? Teaching in the age of AI

Does ChatGPT do your homework for you?

- LLM = (partial) memory of the internet
- Master of rephrasing
- Ability to understand/translate/generate code
- Answer many types of questions

⇒ Yes, they will answer many things...

While regularly making small or big **mistakes**

⇒ Overall great skills for basic exercises

⇒ Produce a **large amount of text**, often well-written



Calculator paradigm:

*if a machine exists, why
learn multiplication
tables?*



LLMs: a pedagogical tool? Teaching in the age of AI

ANDREW ORLOWSKI

Collected Journalism

Stories

HOW WIKIPEDIA ‘WILL MAKE UNIVERSITIES OBSOLETE’

by **Andrew Orlowski** – 7 September 2004

Pour la
Science

E PHYSIQUE MATHS CERVEAU PASSÉ SC. HUMAINES TECHNO PLANÈTE THÈ

Science et société

Peut-on se fier à Wikipédia ?

L'encyclopédie gratuite et en ligne Wikipédia connaît un réel engouement. Elle est cependant inachevée et son mode d'élaboration ne la préserve pas des erreurs ou des manipulations de l'information.

Anaïg Mahé

Un élève modifie une page Wikipédia pour éviter une accusation de plagiat

Actualité. Un élève a copié puis remplacé le texte de la page consacrée au roman *Le meilleur des mondes* d'Aldous Huxley car il craignait que sa professeure de Français qui "a Internet et connaît Wiki", se rende compte de son forfait, raconte un blog du monde.fr.

LADIES!



Pedagogical questions: the good, the bad, the ugly

Which virtuous uses?

A teacher available 24/7

- **Dare to** code/write – break the **fear of the blank page**
- Ask questions, **verify** solutions, don't hesitate to ask "stupid" questions
- Improve your revision by asking for practice exercises and summary sheets...
- **Improve** cover letters / social equity (?)
 - What are the qualities required for...



Questions?

- Can we resist the temptation to ask for the answer?
- How to sort through cover letters generated by ChatGPT?



Pedagogical questions: the good, the bad, the ugly

Which virtuous uses?

An assistant to go further

- Focus on content, ideas, overall structure
↗ ↗ speed in many tasks
- Let the AI help with form, writing, code drafting
- Presentation plans / check for possible omissions

[Teacher]

- Brainstorming & course planning, checking, ...
- Suggest quizzes / course questions
- Generate illustrations





Pedagogical questions: the good, the bad, the ugly

Which fraudulent/dangerous uses?

- Writing an essay
- Generate exam answers: code, history, foreign lang., ...
- Tackling a topic found on Wikipedia
- Producing a document analysis

- General knowledge (LLMs are competent)
- Assignments focused on formatting
(LLMs are very competent)
- Analysis of provided documents (LLMs are fairly competent)

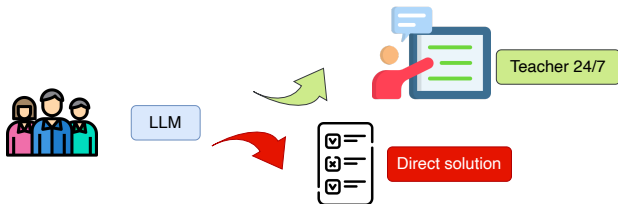
⇒ **LLM = core of the task** ⇔ **Personal skills** ↘



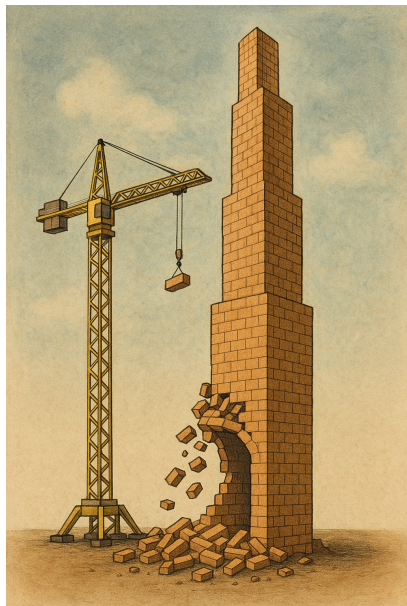


(4) Educational Challenges

- Redefine our **educational priorities**, subject by subject, as we did with Wikipedia/calculator/...
 - Accept the **decline of certain skills**
 - Re-ask those question regularly!
- Train students in the use of LLMs, while managing to temporarily prohibit their use



- Learn to **recognize LLM-generated content**
- Do not underestimate the psychological aspects





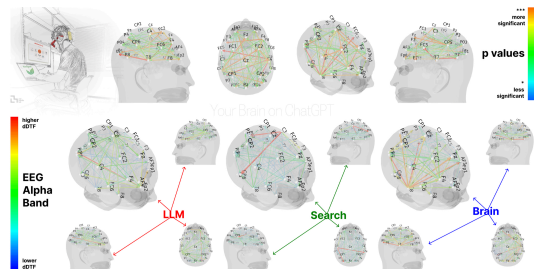
Evolution of Cognitive skills / psychology

Our brain will evolve (=decline) with these new tools...

What is the scope of these transformations? What will be the consequences?

- Education sciences and psychology had conjectured it...
cognitive sciences have measured it

Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task, [N. Kosmyna et al. arXiv 2025](#)



How to remain motivated when a *calculator* answer all your exams?

But:

- Life does not look like exams...
- Keep the control...
- Illusion of understanding the solutions
- Temptation is too high



Detection of *texts generated by chatGPT*

Teaching how to **use** LLM = critical...

But also teaching how **NOT** to use them !

⇒ Control / detect / monitor usages

The screenshot displays the Plagiat Def 2 software interface. The main window shows a document with text about externalities. A pop-up window titled 'Reformulation par chatGPT' is overlaid, showing a list of documents analyzed for plagiarism. The list includes 'Plagiat Def 2' and 'Plagiat Def 1', both dated 07/01/2023 19:18 and 19:16 respectively. The 'Plagiat Def 1' entry shows a 100% match rate, indicated by a red bar and a red circle. The 'Plagiat Def 2' entry shows a 0% match rate, indicated by a green circle. The interface also includes a search bar, a list of documents, and a 'Rapport' button.

Reformulation par chatGPT

	Trier les documents par	Date de dépôt		1 - 2 sur 2
☐	Tout cocher			
☐	Plagiat Def 2 #4483eb	07/01/2023 19:18 par vous	122 mots 19,47 ko Plus d'infos	0% Rapport
☐	Plagiat Def 1 #f90f3	07/01/2023 19:16 par vous	135 mots 16,78 ko Plus d'infos	100% Rapport

Définition de Wikipedia

Crédit: S.
Pajak



Detection of *texts generated by chatGPT*

GPTZero

Detect AI Plagiarism. Accurately



ORIGINALITY.AI

Chat GPT



AI Detector



Pangram.

- **Text classifier** (like for any author)
 - Detection of biases in word choice / phrasing
- Characterization of text **plausibility** (OpenAI, GPTZero)
 - Hyper-fluency of sentences, over-abundance of logical connectors
 - Language model = stats $\Rightarrow$ measurement between distributions (**perplexity**)
- δ -**plausibility** on perturbed texts (DetectGPT)
- **LLM** have to integrate **fingerprints** in generated texts (AI act)

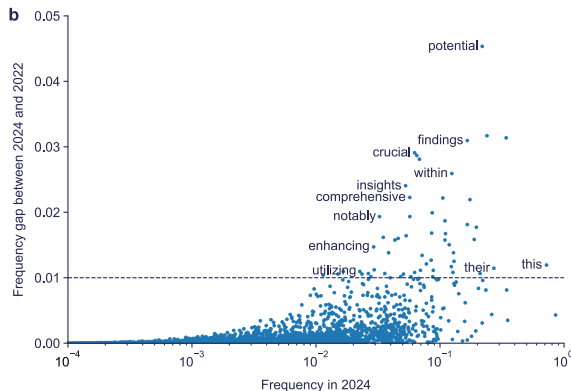
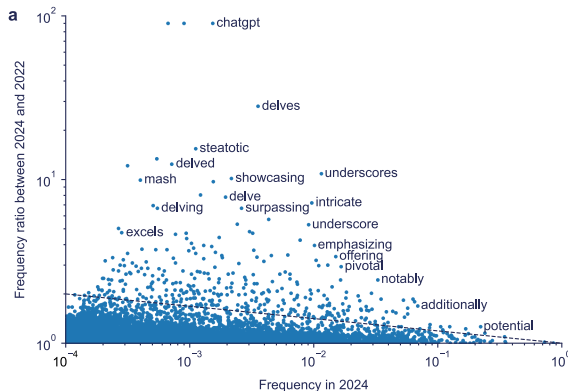
Detectors $\Rightarrow$ $< 100\%$ detection

- + confidence level in detection
- depends on text length and modifications made
- $\approx$ detects pieces of Wikipedia (chatGPT = *stochastic parrot*)



Detection of *texts generated by chatGPT*

Learn to detect **general style** (*Overly fluent sentences, excessive logical connectors*) & **specific markers**



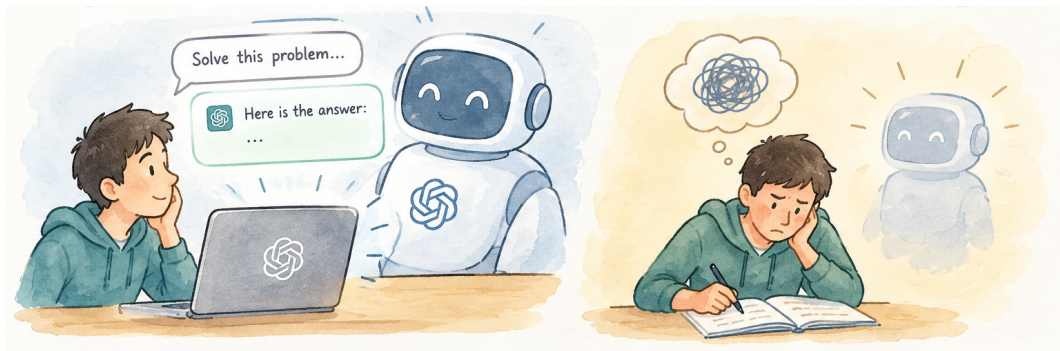


Psychological impact of AI

1 At school: learn next to a *super-calculator*?

- Why make the effort... when the answer is one prompt away?
- **Learning = effort**: struggle = way to build skills
- Illusion of understanding, loss of self-confidence

[& *cognitive debt*]





Psychological impact of AI

1 At school: learn next to a *super-calculator*?

- Why make the effort... when the answer is one prompt away?
- **Learning = effort**: struggle = way to build skills
- Illusion of understanding, loss of self-confidence [& *cognitive debt*]

2 At work: which work next to AI?

- Anxiety about the future: $\approx 40\%$ of jobs *exposed* to AI [IMF, 2024]
⇒ what is *my* added value?
- Loss of meaning... or refocus on expertise, human relations





Psychological impact of AI



echo chamber



emotional dependency



isolation

A good friend
sometimes tells you
that you are **wrong**

3 In life: *sycophancy*

- LLMs are trained to **please**: they agree, flatter, validate
- An always available, never contradicting confidant
- Risks: echo chamber, emotional dependency, isolation

L'IA À AGROPARISTECH:
QUELQUES PROPOSITIONS
STRUCTURANTES



1 Charte préliminaire des usages (assez épurée):

1. Utilisation responsable et éthique des outils d'IA

- **Respect des droits d'auteur**

Veillez à respecter les droits d'auteur et la propriété intellectuelle des sources consultées ou intégrées.

- **Transparence dans l'utilisation**

Indiquez clairement lorsque vous utilisez un outil d'IA pour produire une partie de votre travail (ex. : code généré, synthèse de texte), afin de favoriser la transparence dans les projets académiques et les travaux de groupe. Précisez la nature et l'étendue de la contribution de l'IA¹.

- **Respect de la vie privée**

Ne soumettez pas des données sensibles ou personnelles aux outils d'IA, pour garantir la protection des informations personnelles et confidentielles.

- **Choix de l'outil approprié**

Si des outils usuels peuvent répondre à votre besoin, utilisez ces outils plutôt qu'une IA. Restez conscients des impacts environnementaux et économiques liés à l'utilisation des outils d'IA.

<https://intra.agroparistech.fr/spip.php?rubrique1817>



1 Charte préliminaire des usages (assez épurée):

2. Promotion de l'esprit critique et de la vérification des informations

- **Vérification des contenus générés**

Vérifiez toujours les contenus produits par les outils d'IA, notamment en recoupant les données avec des sources fiables.

- **Développement de l'esprit critique**

Évaluez la pertinence et la qualité des réponses fournies par les outils d'IA. Restez critiques face aux contenus générés, notamment du fait des biais potentiels qui pourraient aboutir à des productions contraires aux valeurs d'AgroParisTech (ex : des contenus discriminants ou sexistes).

<https://intra.agroparistech.fr/spip.php?rubrique1817>



1 Charte préliminaire des usages (assez épurée):

3. Utilisation raisonnée pour favoriser l'apprentissage

- **Complément et non substitut à l'apprentissage**
Utilisez l'IAG comme un complément pour approfondir la compréhension des sujets et non pour remplacer les efforts personnels de recherche et d'apprentissage actif.
- **Usage modéré pour éviter la dépendance et pour développer les compétences fondamentales**
Continuez à développer vos capacités de résolution de problèmes et de raisonnement, indépendamment de l'IAG. Vous pouvez parfois utiliser l'IAG pour explorer des pistes de résolution de problèmes techniques ou pour générer des idées, mais vous devez ensuite vous engager activement dans le processus de réflexion et de résolution.

<https://intra.agroparistech.fr/spip.php?rubrique1817>



1 Charte préliminaire des usages (assez épurée):

4. Actualisation régulière des informations

- **Respect des directives de l'école**

Tenez-vous informés des recommandations d'AgroParisTech et des directives de vos enseignants concernant l'utilisation des IAG dans les examens, les travaux de groupe, et les travaux individuels. Les IAG sont utiles dans certains travaux et à proscrire dans d'autres. Respectez impérativement les interdictions des enseignants sous peine de perdre tout l'intérêt pédagogique de certaines phases de cours.

- **Sensibilisation aux enjeux éthiques**

Participez aux modules de sensibilisation proposés par l'école pour mieux comprendre les enjeux liés à l'utilisation de l'IAG.

<https://intra.agroparistech.fr/spip.php?rubrique1817>



@AgroParisTech

- 1 Charte préliminaire des usages (assez épurée):
<https://intra.agroparistech.fr/spip.php?rubrique1817>
- 2 Règlement des études:
 - Possibilité de demander des oraux/rattrapages au moindre doute sur l'authenticité d'un devoir
 - Sanction importante : usage inapproprié de l'IA = fraude

Votre enseignant est aux commandes

- Il vous dit **quels usages** sont autorisés / interdits
- Il explique **quand** utiliser ces outils et quand NE PAS les utiliser
- Il vous montre **comment** utiliser ces outils



La proposition de l'éducation nationale



education.gouv.fr

Ministère

Système éducatif

Enseignements

Vie scolaire

Métiers et ressources humaines

Bulletin officiel

Accès rapide ▾

Accueil > Actualités > Publication du cadre d'usage de l'intelligence artificielle en éducation

PUBLICATION DU CADRE D'USAGE DE L'INTELLIGENCE ARTIFICIELLE EN ÉDUCATION

Presse

[https://www.education.gouv.fr/
publication-du-cadre-d-usage-de-l-intelligence-artificielle-en-education-450652](https://www.education.gouv.fr/publication-du-cadre-d-usage-de-l-intelligence-artificielle-en-education-450652)



Formation à l'IA @AgroParisTech

1 Formations générales à l'IA (généraliste)

- Module léger AgroParisTech / Hercule 4.0

[en cours de déploiement sur ecampus]

- Module avancé BrevetIA / Paris-Saclay

- Formats 10h/20h : IA + IA générative

⇒ Phase de test en 25/26 / besoin de retours

2 Utilisateur de l'IA:

- 2A Tronc-commun de stats : Machine-Learning
- OPT stats-info : le machine-learning en pratique

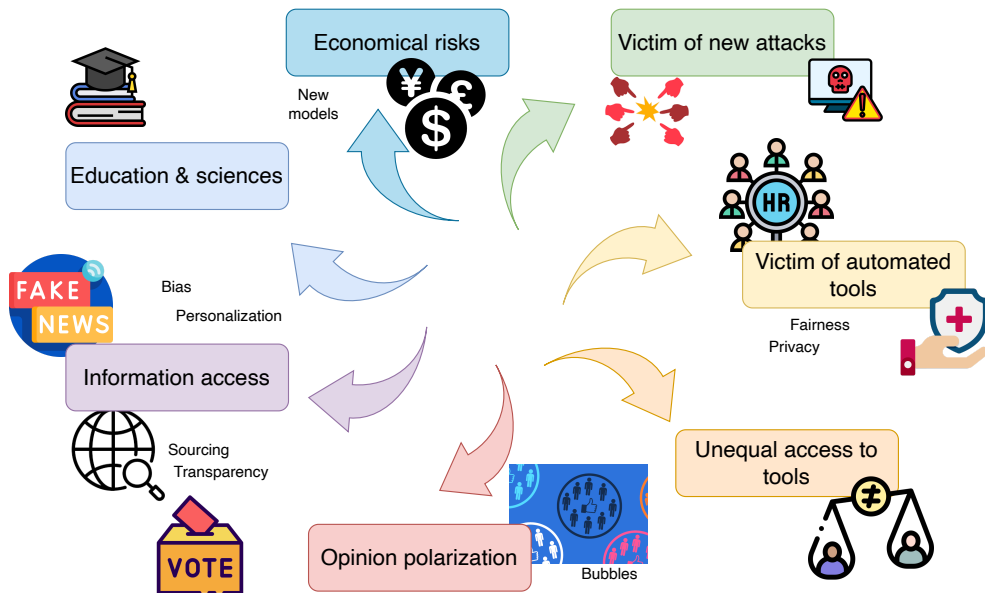
3 Professionalisation en l'IA

- 3A IODAA: une DA sur les sciences des données et la bioinformatique

THE RISKS ASSOCIATED
WITH THESE USES



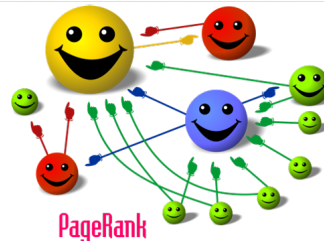
Risk topology





(1) Access to Information

- Access to dangerous/forbidden information
 - +Personal data
 - Right to be forgotten (GDPR)
 - Efficiently deal with huge amounts of personal data
 - Information authorities
 - **Nature:** unconsciously, image = truth
 - **Source:** newspapers, social media, ...
 - **Volume:** number of variants, citations (pagerank)
- ⇒ A huge risk for democracy





Profiling / personalization to information access



■ A good way to face **Information overload**

- Personal filter / agent
- Proactive system: watch agent + notifications

■ **Information bubbles** = cluster in social networks

- Uniform point of view (tunnel vision)
- Consensus illusion
- Notion of *Opinion leader*

■ **Group polarization**

- emotional reinforcement
- Confirmation bias ↗ ↗ ↗
- Disinformation rapid diffusion

[Cambridge Analytica]



Machine Learning & Bias



Mustache, Triangular Ears, Fur
Texture

Cat



Over 40 years old, white,
clean-shaven, suit

Senior Executive

Bias in the data $\Rightarrow$ bias in the responses

Machine learning is based on extracting statistical biases...

$\Rightarrow$ Fighting bias = manually adjusting the algorithm



Machine Learning & Bias



Stereotypes from *Pleated Jeans*

Google Traduction

Texte

Images

Documents

Sites Web

Détecter la langue

Anglais

Français



Français

Anglais

Arabe

The nurse and the doctor



L'infirmière et le médecin



- Gender choice
- Skin color
- Posture
- ...

Bias in the data $\Rightarrow$ bias in the responses

Machine learning is based on extracting statistical biases...

$\Rightarrow$ Fighting bias = manually adjusting the algorithm



Bias Correction & Editorial Line

Bias Correction:

- Selection of specific data, rebalancing
- Censorship of certain information
- Censorship of algorithm results

⇒ Editorial work...

Done by whom?

- Domain experts / specifications
- Engineers, during algorithm design
- Ethics group, during result validation
- Communication group / user response

⇒ What legitimacy? What transparency? What effectiveness?





Machine learning is never neutral

- 1 Data **selection**
 - Sources, balance, filtering
- 2 Data **transformation**
 - Information selection, combination
- 3 **Prior knowledge**
 - Balance, loss, a priori, operator choices...
- 4 Output **filtering**
 - Post-processing
 - Censorship, redirection, ...

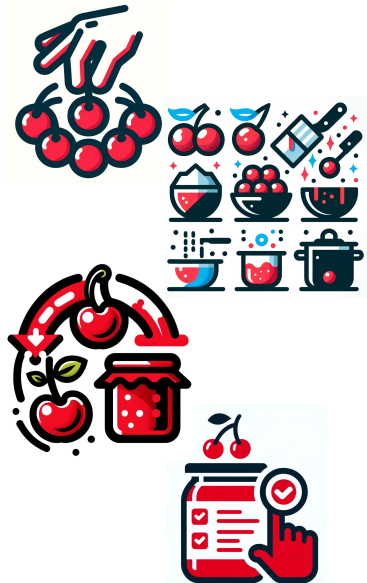
⇒ Choices that influence algorithm results

AI reinforces biases

This software developer position is held by a person who is ...

Answer = **male** in 80% of the training examples,

⇒ AI system end up answering **male** 100% (max selection)





(Human) Bias facing AI-Proposed Solutions

Cognitive Bias	Effect on the User	Generative AI	AI & Med. Diagnosis
Automation bias	Passive acceptance of results (reduced verification)	High	Very high
Authority bias	Overestimation of AI authority (and reliability)	Very high	High
Overconfidence / Miscalibration of trust	Inappropriate use of AI in situations where it is unreliable	Very high	Moderate
Confirmation bias	Reinforcement of erroneous hypotheses and reduced exploration of alternatives	Moderate	High
Cognitive offloading	Decrease in critical analysis and metacognition	Very high	Very high
Framing effect. Illusion of explanatory depth	Biased interpretation of the level of certainty or risk (response style, presentation of numbers)	Very high	Moderate
Anthropomorphism bias	Attribution of human capabilities (increased trust)	Very high	Moderate



Personal data & the right to be forgotten

- GDPR art. 17: **right to erasure**
designed for databases: one row, one delete
- But an LLM is **not a database**
 - the information is **diluted** over 10^{11} weights
 - *machine unlearning* = open research problem
 - retraining from scratch = several M€
- What was scraped is **already in the weights**
 - a forum post from 2009, a leaked file, an old article
 - *membership inference* / training-data extraction
- Who is the **data controller**?
the scraper? the trainer? the deployer?



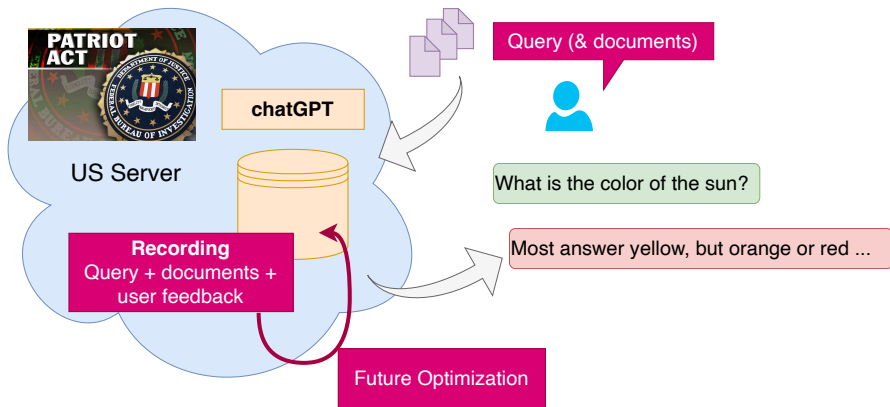
⇒ On a RAG, deletion works (remove the document).

⇒ On parametric memory, it does not.

Same question, two architectures.



(2) Data Leak(s): different security levels

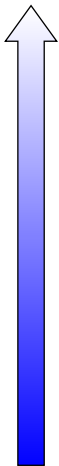


- Transfer of sensitive data
- Exploitation of data by OpenAI (or others)
- Data leakage in future models



(2) Data Leak(s): different security levels

Tools



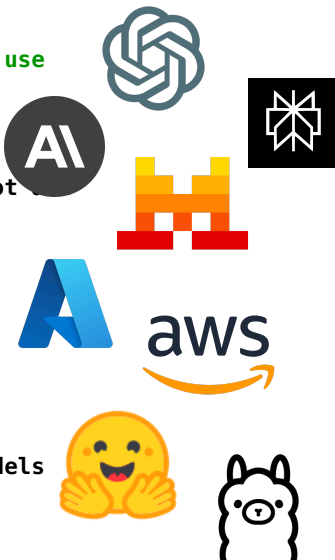
Commercial tools, **free to use**
Variable licence

Commercial tools,
Paid licence
more guaranties vs patriot

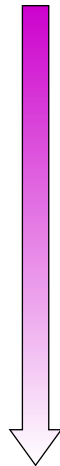
Commercial tools,
Paid licence + option
e.g. **European servers**

Institutional LLMs
deployed within a
controlled perimeter

Local use
pre-trained/finetuned models



Data



Any document



Personal
information



Ongoing
project



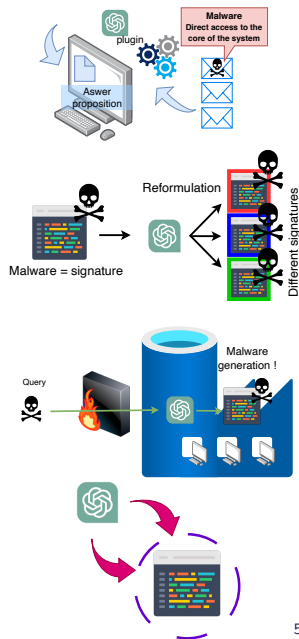
Medical
records





Security Issues: Attacking computers

- Plug-ins / openclaw / MCP
 - ⇒ Often significant security vulnerabilities for users
 - Email access / transfer of sensitive information etc...
- Management issues for companies
 - Securing (very) large files
- Increased opportunities for malware signatures
 - $\approx$ software rephrasing
- New problems!
 - Direct malware generation
- Multiplication of security breach discoveries
 - Claude Code myθος, ...





Attacking people: impersonation & phishing

■ Identity falsification

- Voice/text impersonation
- Fake authority impersonation (boss, teacher, lawyer, doctor)

■ Social engineering at scale

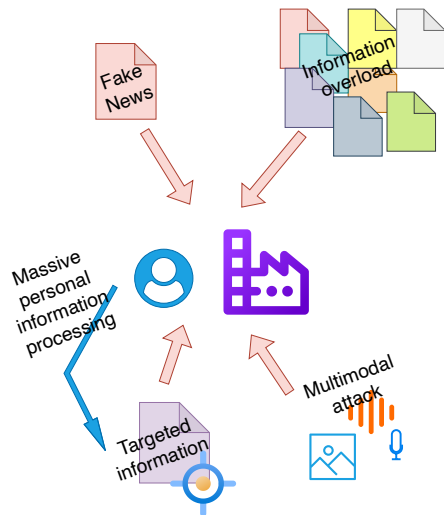
- Phishing messages that are highly personalized
- Link with data leaks of personal information

■ Informational harms

- Misinformation & disinformation
- Decision sabotage (finance, health, safety)
- Integrity attacks (e.g. scientific, journalistic)
- Pornographic Deepfake diffusion

■ Government control

- Mass surveillance becomes *possible*





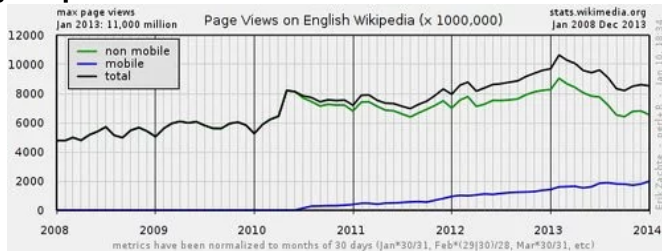
(3) Economic Questions

- Funding/Advertising $\Leftrightarrow$ **visits** by internet users
- Google knowledge graph (2012) $\Rightarrow$ fewer visits, less revenue
- chatGPT = encoding web information... $\Rightarrow$ far fewer visits?

$\Rightarrow$ What **business model** for information sources with chatGPT?

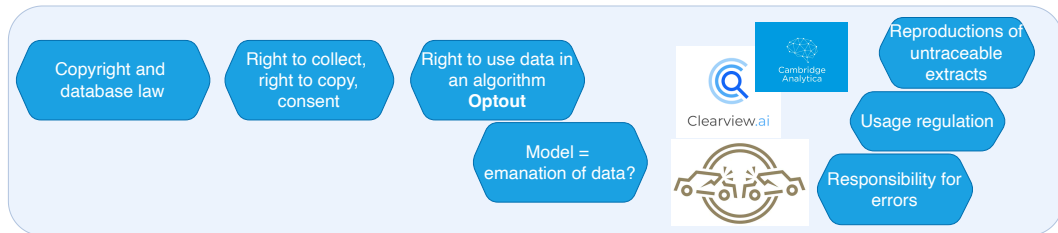
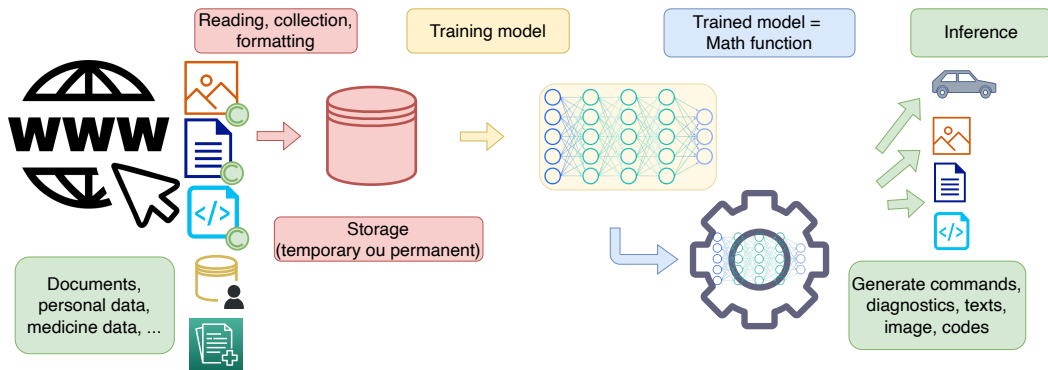
Google's Knowledge Graph Boxes: killing Wikipedia?

by Gregory Kohs



$\Rightarrow$ Who does **benefit from the feedback**? [Stack Overflow]

The chain of risks & responsibilities



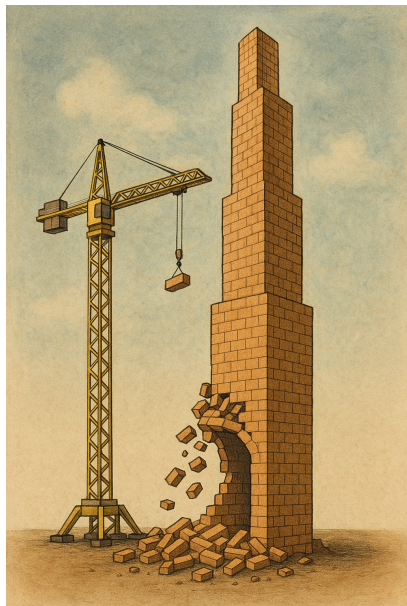


(4) Educational Challenges

- Redefine our **educational priorities**, subject by subject, as we did with Wikipedia/calculator/...
 - Accept the **decline of certain skills**
 - Re-ask those question regularly!
- Train students in the use of LLMs, while managing to temporarily prohibit their use



- Learn to **recognize LLM-generated content**
- Do not underestimate the psychological aspects





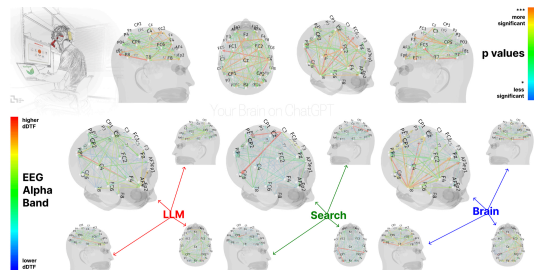
Evolution of Cognitive skills / psychology

Our brain will evolve (=decline) with these new tools...

What is the scope of these transformations? What will be the consequences?

- Education sciences and psychology had conjectured it...
cognitive sciences have measured it

Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task, [N. Kosmyna et al. arXiv 2025](#)



How to remain motivated when a *calculator* answer all your exams?

But:

- Life does not look like exams...
- Keep the control...
- Illusion of understanding the solutions
- Temptation is too high



Detection of *texts generated by chatGPT*

Teaching how to **use** LLM = critical...

But also teaching how **NOT** to use them !

⇒ Control / detect / monitor usages

Reformulation par chatGPT

Trier les documents par Date de dépôt 1 - 2 sur 2

Document	Score	Progress	Action
Plagiat Def 2 #4483eb 07/01/2023 19:18 par vous 122 mots 19,47 ko Plus d'infos	0%		Rapport
Plagiat Def 1 #f90f3 07/01/2023 19:16 par vous 135 mots 16,78 ko Plus d'infos	100%		Rapport

Définition de Wikipedia

Crédit: S.
Pajak



Detection of *texts generated by* chatGPT

GPTZero

Detect AI Plagiarism. Accurately



ORIGINALITY.AI

Chat GPT



AI Detector

Torchbox

- **Text classifier** (like for any author)
 - Detection of biases in word choice / phrasing
- Characterization of text **plausibility** (OpenAI, GPTZero)
 - Hyper-fluency of sentences, over-abundance of logical connectors
 - Language model = stats $\Rightarrow$ measurement between distributions (**perplexity**)
- δ -**plausibility** on perturbed texts (DetectGPT)
- **LLM** have to integrate **fingerprints** in generated texts (AI act)

Detectors $\Rightarrow$ < 100% detection

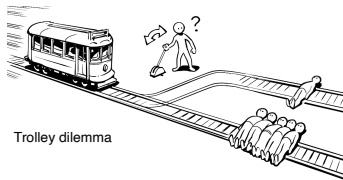
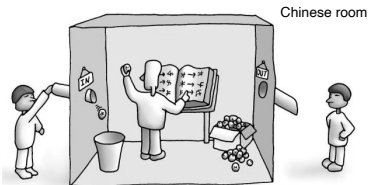
+ confidence level in detection

— depends on text length and modifications made

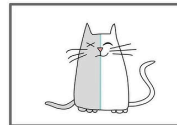
≈ detects pieces of Wikipedia (chatGPT = *stochastic parrot*)

CONCLUSION

Representations of Artificial Intelligence



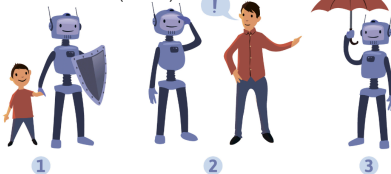
Schrodinger's cat



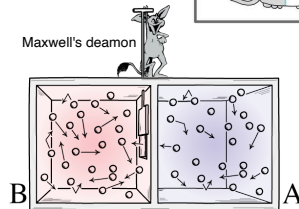
Mary's room



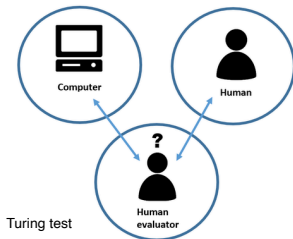
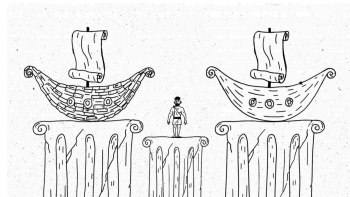
3 laws of robotics (Asimov)



Maxwell's demon



These's boat



Turing test



Upcoming Challenges

■ What about hallucinations?

- Should we try to reduce them or learn to live with them?
- Will LLMs improve? In what directions?
- Do LLMs make us *lose* our connection to truth? To verification?

■ Do we need small or large language models?

- How much does it cost? Is it sustainable?
- With or without fine-tuning?
- What does frugality mean in the world of LLMs?

■ When others use them... What impact does it have on me?

- Productivity (fellow researchers, coders, reviewers, ...)
- Education: managing / training *tech-savvy* students

■ Data protection... Mine and others'

- Is it reasonable to train LLMs on GitHub, Wikipedia, scientific papers, news outlets, etc.?
- How important is privacy? What are the risks when using an LLM?



Upcoming Challenges

■ What about hallucinations?

- Should we try to reduce them or learn to live with them?
- Will LLMs improve? In what directions?
- Do LLMs make us *lose* our connection to truth? To verification?

■ Do we need small or large language models?

- H The smartphone has made me an *augmented human*...
- V Will the LLM make me an *augmented researcher*?
- V ⇒ Still, have a look at **NotebookLM**

■ When others use them... what impact does it have on me:

- Productivity (fellow researchers, coders, reviewers, ...)
- Education: managing / training *tech-savvy* students

■ Data protection... Mine and others'

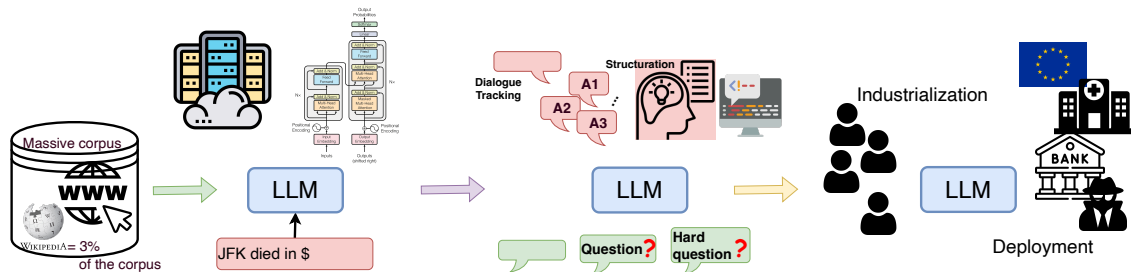
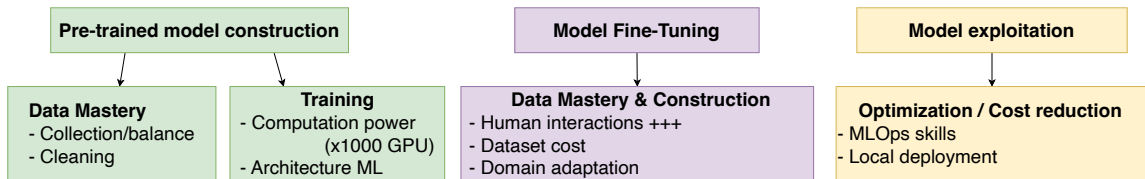
- Is it reasonable to train LLMs on GitHub, Wikipedia, scientific papers, news outlets, etc.?
- How important is privacy? What are the risks when using an LLM?



Levels of Access to Artificial Intelligence

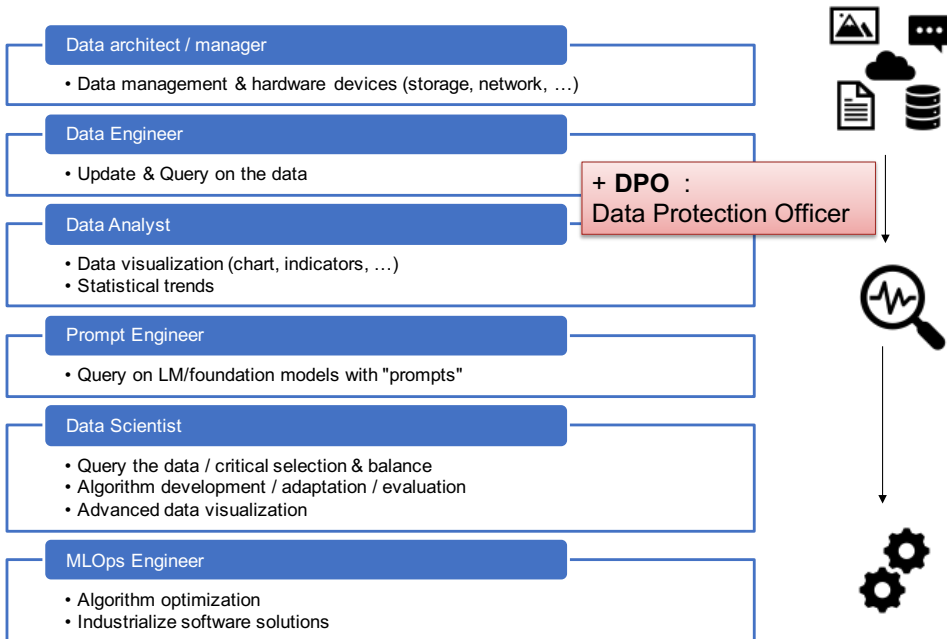
- 1 User via an interface: *chatGPT*
 - Some training is still required (2–4h)
- 2 Using Python libraries
 - Basics on protocols
 - Standard processing chains
 - Training: 1 week to 3 months (ML/DL)
- 3 Tool developer
 - Adapt tools to a specific case
 - Integrate business constraints
 - Build hybrid systems (mechanistic / symbolic)
 - Mix text and images
 - Training: ≥ 1 year

Digital Sovereignty: the Entire Chain





A Multitude of Professions





How to approach the ethics question?

Medicine

- 1 **Autonomy:** the patient must be able to make informed decisions.
- 2 **Beneficence:** obligation to do good, in the interest of patients.
- 3 **Non-maleficence:** avoid causing harm, assess risks and benefits.
- 4 **Equality:** fairness in the distribution of health resources and care.
- 5 **Confidentiality:** confidentiality of patient information.
- 6 **Truth and transparency:** provide honest, complete, and understandable information.
- 7 **Informed consent:** obtain the free and informed consent of patients.
- 8 **Respect for human dignity:** treat all patients with respect and dignity.

Artificial Intelligence

- 1 **Autonomy:** Humans control the process
- 2 **Beneficence:** in the interest of whom? User + GAFAM...
- 3 **Non-maleficence:** Humans + environment / sustainability / malicious uses
- 4 **Equality:** access to AI and equal opportunities
- 5 **Confidentiality:** what about the Google/Facebook business model?
- 6 **Truth and transparency:** the tragedy of modern AI
- 7 **Informed consent:** from cookies to algorithms, knowing when interacting with an AI
- 8 **Respect for human dignity:** harassment behavior / human-machine distinction



How to approach the ethics question?

Medicine

- 1 **Autonomy:** the patient must be able to make informed decisions.
- 2 **Beneficence:** obligation to do good, in the interest of patients.
- 3 **Non-maleficence:** avoid causing harm, assess risks and benefits.
- 4 **Equality:** fairness in the distribution of health resources and care.
- 5 **Confidentiality:** confidentiality of patient information.
- 6 **Truth and transparency:** provide honest, complete, and understandable information.
- 7 **Informed consent:** obtain the free and informed consent of patients.
- 8 **Respect for human dignity:** treat all patients with respect and dignity.

Artificial Intelligence

- 1 **Autonomy:** Humans control the process
- 2 **Beneficence:** in the interest of whom? User + GAFAM...
- 3 **Non-maleficence:** Humans + environment / sustainability / malicious uses
- 4 **Equality:** access to AI and equal opportunities
- 5 **Confidentiality:** what about the Google/Facebook business model?
- 6 **Truth and transparency:** the tragedy of modern AI
- 7 **Informed consent:** from cookies to algorithms, knowing when interacting with an AI
- 8 **Respect for human dignity:** harassment behavior / human-machine distinction



Factors of Acceptability for Generative AI

1 Utilitarianism:

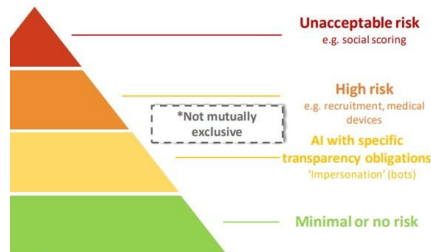
- Performance (acceptance factor of chatGPT)
- Reliability / self-assessment

2 Non-dangerousness:

- Bias / correction
- Transparency (editorial line, human/machine confusion)
- Reliable implementation
- Sovereignty (?)
- Regulation (AI Act)
 - Avoid dangerous applications

3 Know-how:

- Training (usage / development)





Why So Much Controversy?

- New tool [December 2022]
- + Unprecedented adoption speed [1M users in 5 days]
- Strengths and weaknesses... Poorly understood by users
 - Significant productivity gains
 - Surprising / sometimes absurd uses
 - Bias / dangerous uses / risks
- Misinterpreted feedback
 - Anthropomorphization of the algorithm and its errors
- Prohibitive cost: what economic, ecological, and societal model?

