MIA
PARIS-SACLAY
EKINOCS

LES MODÈLES DE LANGUE
USAGES ET ENJEUX SOCIÉTAUX

Mercredi 8 Octobre 2025
Séminaire CNRS, ANF-TDM-IA 2025

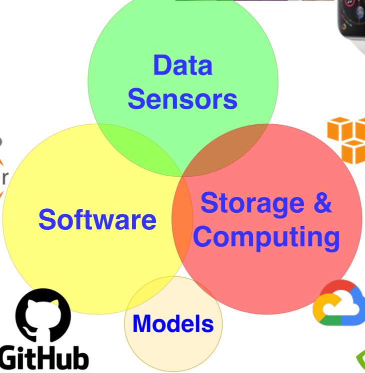
Vincent Guigue
`https://vguigue.github.io`

INRAE AgroParisTech université PARIS-SACLAY

Introduction

The Ingredients of Machine-Learning

WWW



amazon web services

Microsoft Azure

Google Cloud

NVIDIA

jupyter

python™

NVIDIA CUDA

TensorFlow

APACHE Spark™

GitHub

PyTorch

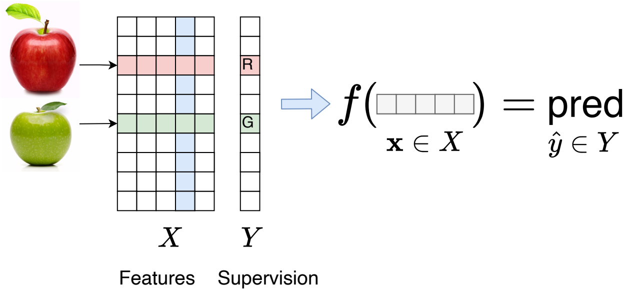
scikit learn

Hugging Face

From tabular data to text

- **Tabular data**
  - Fixed dimension
  - Continuous values

  ⇒ A perfect playground for machine learning



- **Textual data**
  - Various length
  - Discrete values

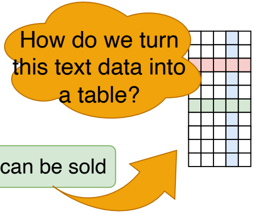
  ⇒ Complex for machine learning

This new iPhone, what a marvel

An iPhone, What a scam!

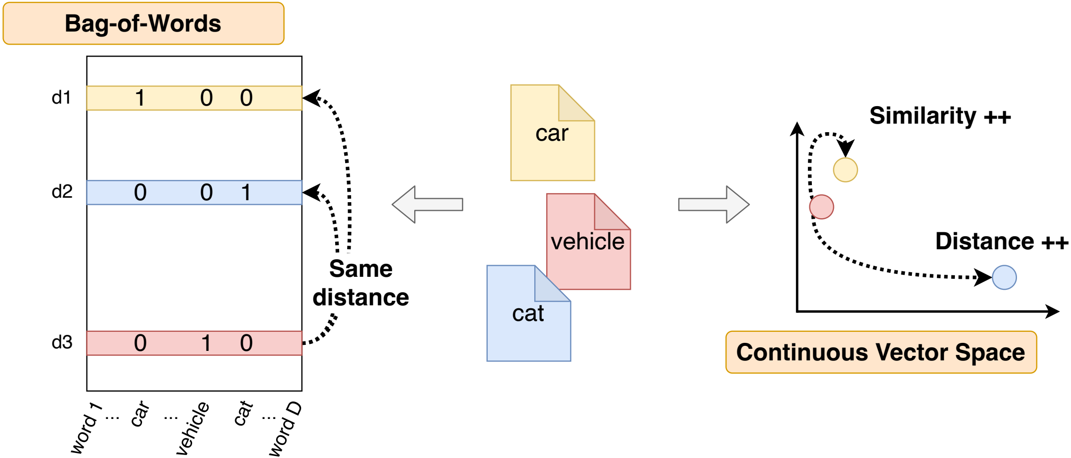
Half the price is for the logo

Apple once again proves that perfection can be sold


How do we turn this text data into a table?

Deep/Representation Learning for Text Data

From Bag of Words to Vector Representations [2008, 2013, 2016]

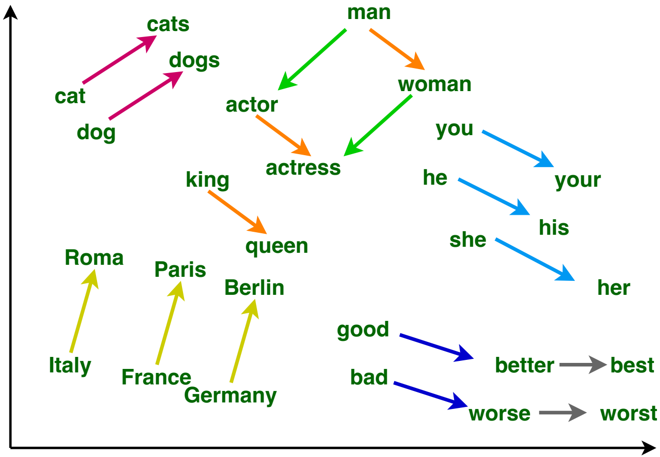


	word 1	…	car	…	vehicle	cat	…	word D
d1			1		0	0		
d2			0		0	1		
d3			0		1	0		

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

Deep/Representation Learning for Text Data

From Bag of Words to Vector Representations [2008, 2013, 2016]



- Semantic Space:

  similar meanings

  ⇔

  close positions

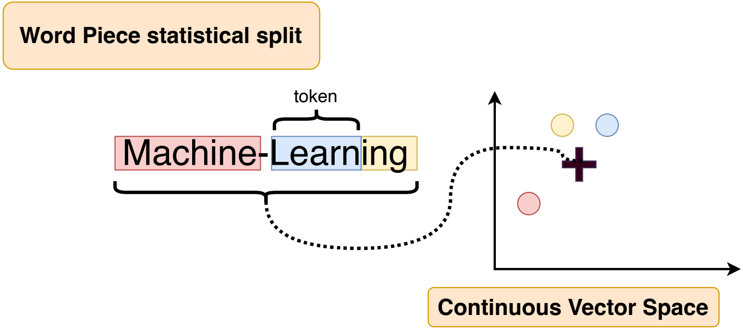
- Structured Space: grammatical regularities, basic knowledge, ...

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

Deep/Representation Learning for Text Data

From Bag of Words to Vector Representations [2008, 2013, 2016]

From Words to Tokens



Word Piece statistical split

token

Machine-Learning

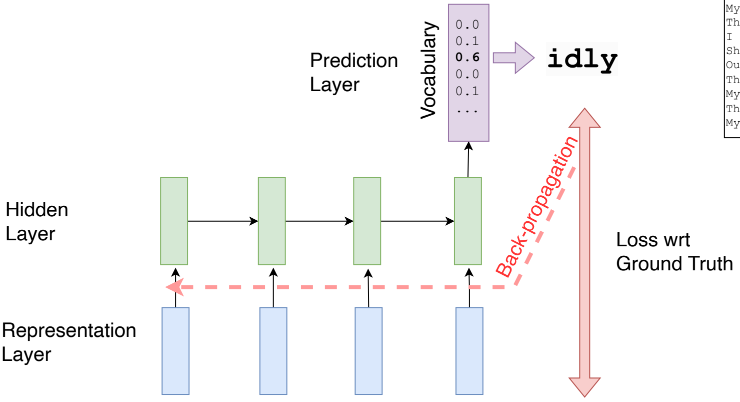
Continuous Vector Space

- Representation of unknown words
- Adaptation to technical domains
- Resistance to spelling errors

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

Aggregating word representations: towards generative AI

- Generation & Representation
- New way of learning word positions



The fluffy cat napped lazily in the sunbeam.

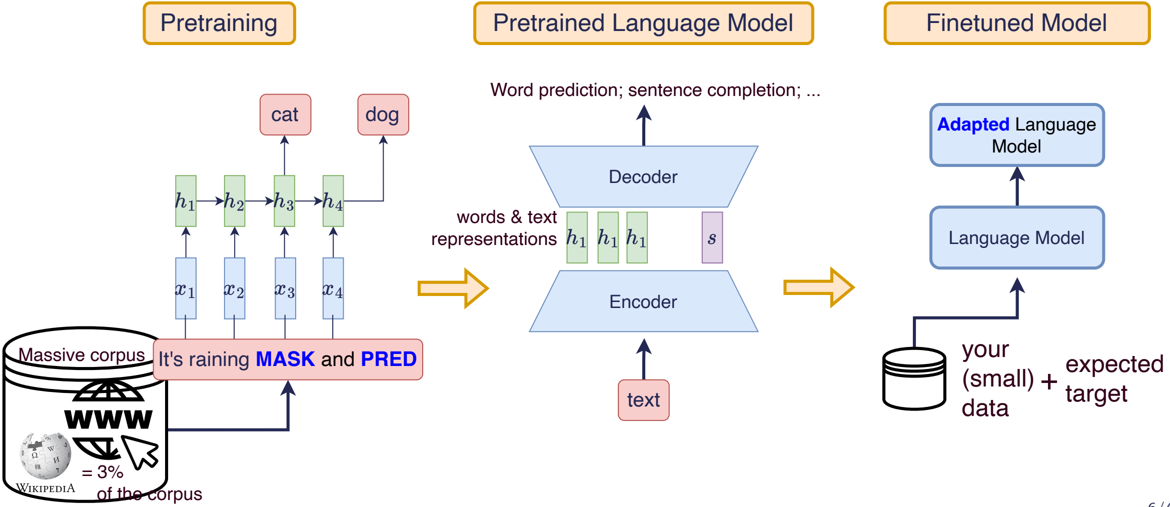
```
The fluffy cat napped lazily in the sunbeam.
I adopted a stray cat from the shelter last week.
My cat loves to chase after toy mice.
The black cat stealthily crept through the dark alley.
I often find my cat perched on the windowsill, watching birds.
She gently stroked her cat's fur as it purred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

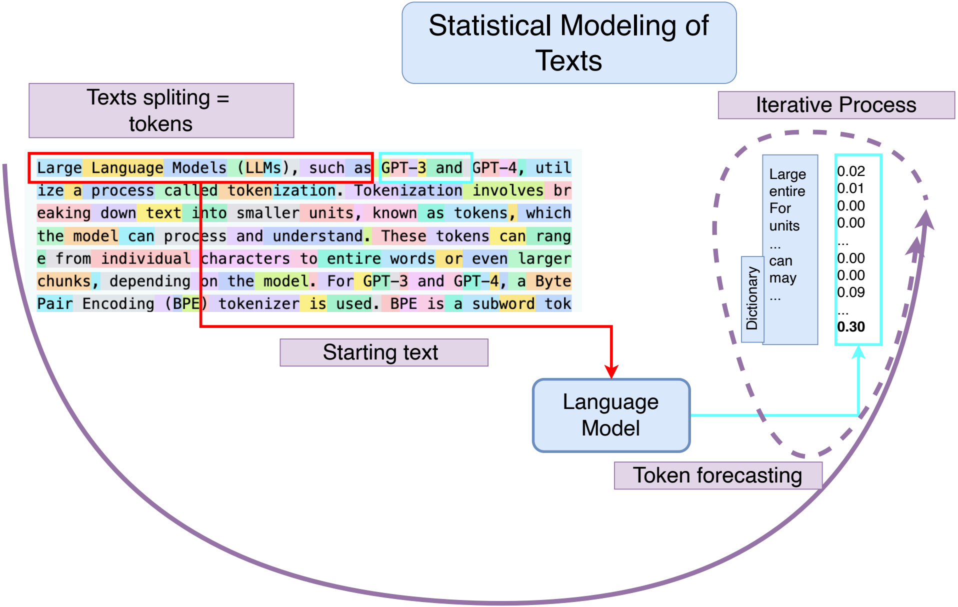
Sequence to Sequence Learning with Neural Networks, Sutskever et al. NeurIPS 2014

A new developpement paradigm since 2015

- Huge dataset + huge archi. ⇒ unreasonable training cost
- Pre-trained architecture + 0-shot / finetuning



At the end of the day: a stochastic parrot :)



Large Language Models (LLMs), such as GPT-3 and GPT-4, util
ize a process called tokenization. Tokenization involves br
eaking down text into smaller units, known as tokens, which
the model can process and understand. These tokens can rang
e from individual characters to entire words or even larger
chunks, depending on the model. For GPT-3 and GPT-4, a Byte
Pair Encoding (BPE) tokenizer is used. BPE is a subword tok

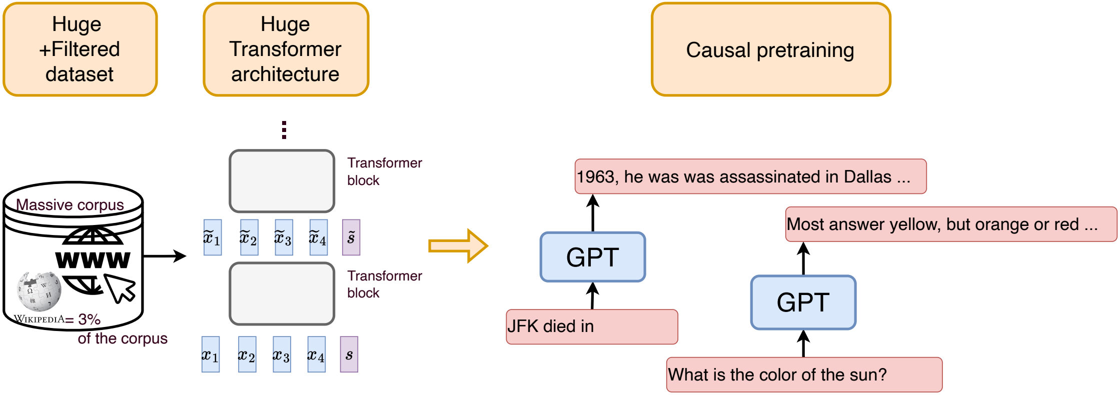
chatGPT

November 30, 2022

1 million users in 5 days
100 million by the end of January 2023
1.16 billion by March 2023

The Ingredients of chatGPT

0. Transformer + massive data (GPT)



Huge +Filtered dataset

Huge Transformer architecture

Causal pretraining

Massive corpus

WWW

WIKIPEDIA= 3% of the corpus

Transformer block

Transformer block

$\tilde{x}_1$ $\tilde{x}_2$ $\tilde{x}_3$ $\tilde{x}_4$ $\tilde{s}$

x_1 x_2 x_3 x_4 s

1963, he was was assassinated in Dallas ...

GPT

JFK died in

Most answer yellow, but orange or red ...

GPT

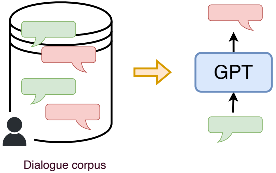
What is the color of the sun?

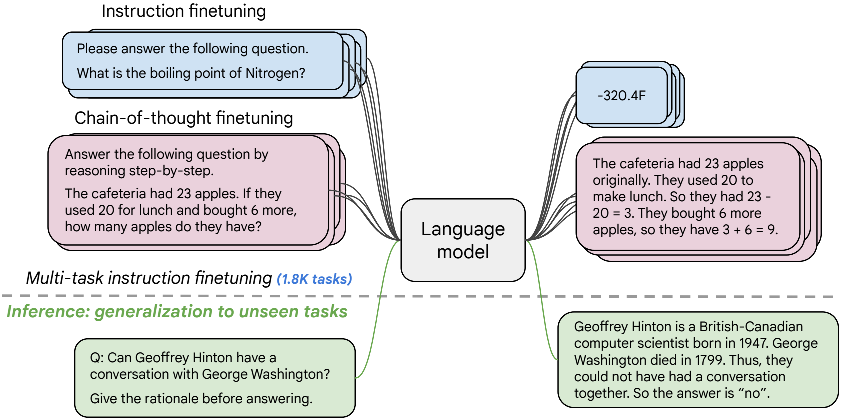
- Grammatical skills: singular/plural agreement, tense concordance
- (Parametric) Knowledge: entities, names, dates, places

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

The Ingredients of chatGPT

1. Dialogue + Tasks

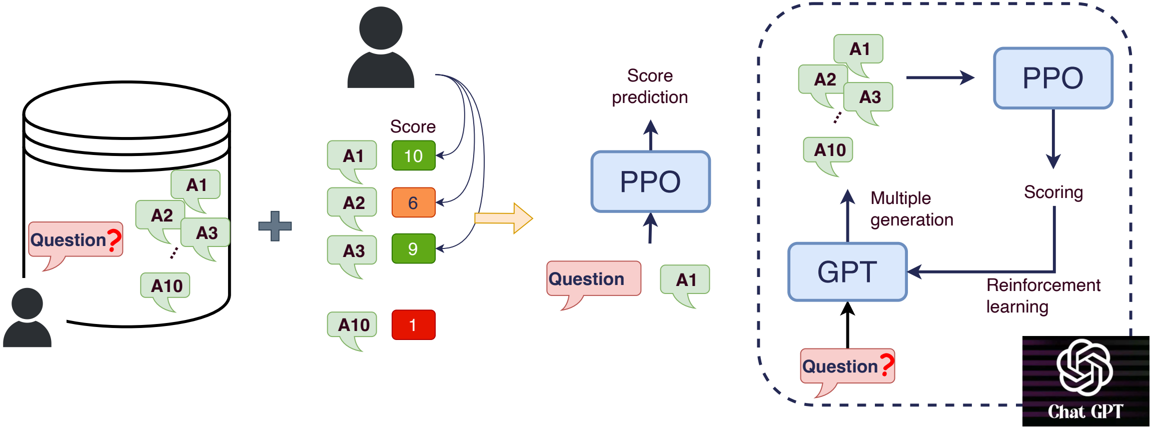



Instruction finetuning
Please answer the following question.
What is the boiling point of Nitrogen?
Chain-of-thought finetuning
Answer the following question by reasoning step-by-step.
The cafeteria had 23 apples. If they used 20 for lunch and bought 6 more, how many apples do they have?
Multi-task instruction finetuning (1.8K tasks)
Inference: generalization to unseen tasks
Q: Can Geoffrey Hinton have a conversation with George Washington?
Give the rationale before answering.
Language model
-320.4F
The cafeteria had 23 apples originally. They used 20 to make lunch. So they had 23 - 20 = 3. They bought 6 more apples, so they have 3 + 6 = 9.
Geoffrey Hinton is a British-Canadian computer scientist born in 1947. George Washington died in 1799. Thus, they could not have had a conversation together. So the answer is "no".

- **Very clean** data Data generated/validated/ranked by humans

The Ingredients of chatGPT

3. Instructions + answer ranking



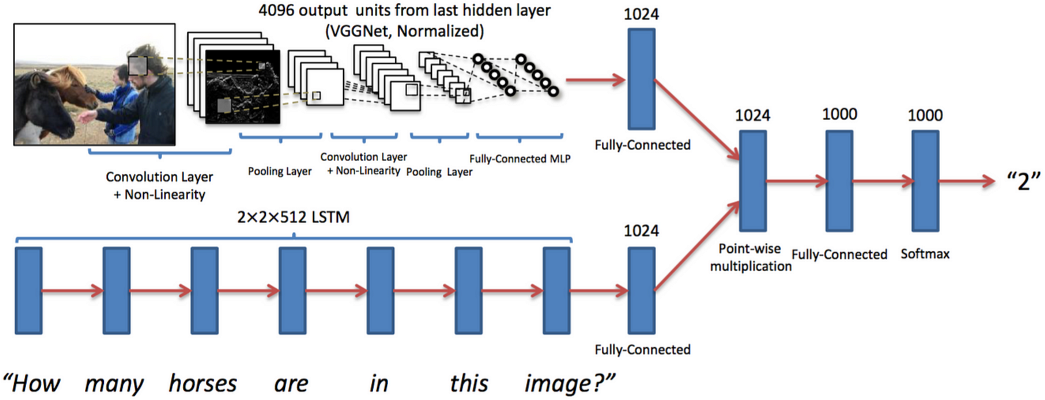
- Database created by humans
- Response improvement
- ... Also a way to avoid critical topics = censorship

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

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.

Machine Learning Limits

chatGPT and the relationship with truth

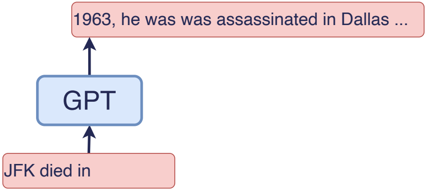
1. **Likelyhood** $=$ grammar, agreement,
   tense concordance, logical sequences...
   $\Rightarrow$ Repeated knowledge

2. Predict the most **plausible** word...
   $\Rightarrow$ produces **hallucinations**

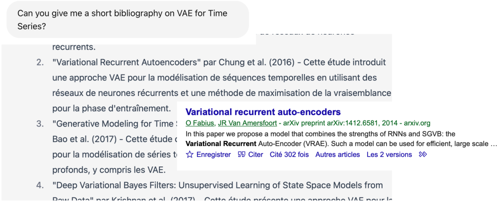
3. **Offline** functioning

4. chatGPT $\neq$ **knowledge graphs**

5. Brilliant answers...
   And silly mistakes!
   $+$ we cannot predict the errors



Example: producing a bibliography


Can you give me a short bibliography on VAE for Time Series?

recurrents.

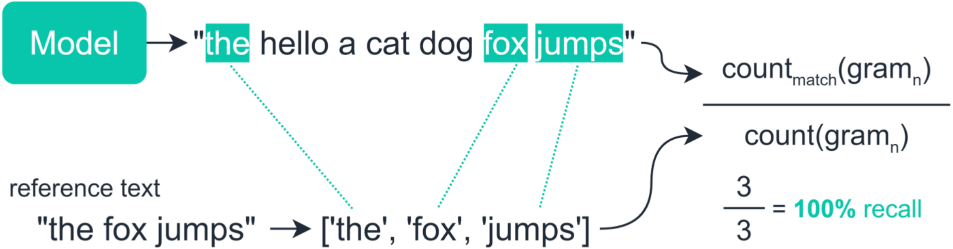
2. "Variational Recurrent Autoencoders" par Chung et al. (2016) - Cette étude introduit une approche VAE pour la modélisation de séquences temporelles en utilisant des réseaux de neurones récurrents et une méthode de maximisation de la vraisemblance pour la phase d'entraînement.
3. "Generative Modeling for Time S
   Bao et al. (2017) - Cette étude c
   pour la modélisation de séries t
   profonds, y compris les VAE.
4. "Deep Variational Bayes Filters: Unsupervised Learning of State Space Models from

Variational recurrent auto-encoders
O Fabius, JR Van Amersfoort - arXiv preprint arXiv:1412.6581, 2014 - arxiv.org
In this paper we propose a model that combines the strengths of RNNs and SGVB: the **Variational Recurrent** Auto-Encoder (VRAE). Such a model can be used for efficient, large scale …

Generative AI: how to evaluate performance?

The critical point today

- How to evaluate against ground truth?
- How to evaluate system confidence / plausibility of generation?

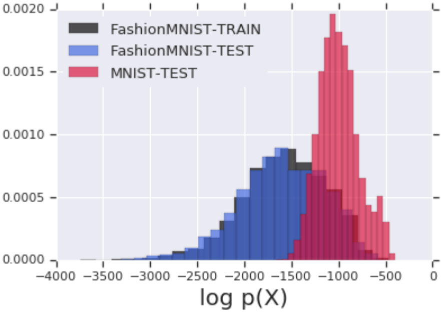


The Ultimate Performance Metric in NLP, J. Briggs, Medium 2021

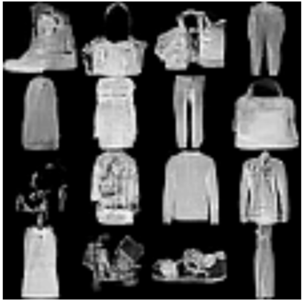
Generative AI: how to evaluate performance?

The critical point today

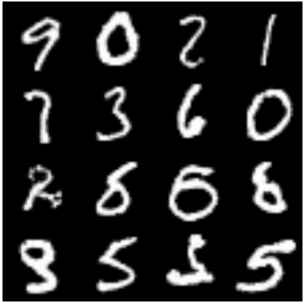
- How to evaluate against ground truth?
- How to evaluate system confidence / plausibility of generation?



Plausibility



Train



Test

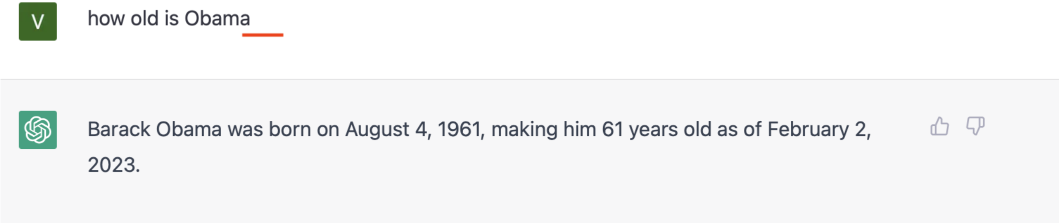
Do Large Language Models Know What They Don't Know?, Yin et al. , ACL, 2023

Do Deep Generative Models Know What They Don't Know?, Nalisnick et al. , ICLR, 2019

Stability/predictability

- Difficult to bound a behavior
- Impossible to predict good/bad answers

⇒ Little/no use in video games



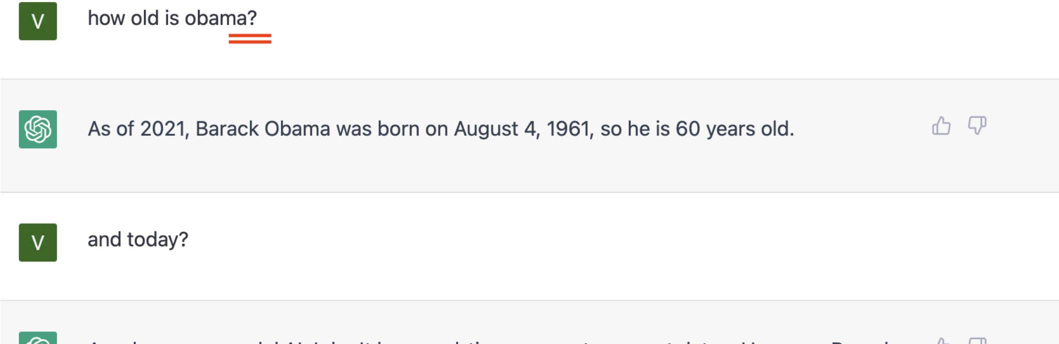
V how old is Obama

Barack Obama was born on August 4, 1961, making him 61 years old as of February 2, 2023.

Stability/predictability

- Difficult to bound a behavior
- Impossible to predict good/bad answers

⇒ Little/no use in video games

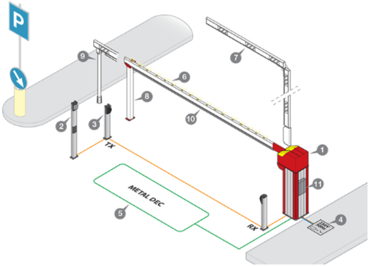


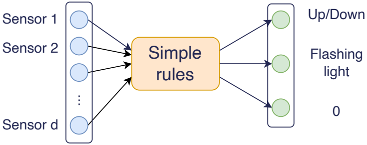
V how old is obama?

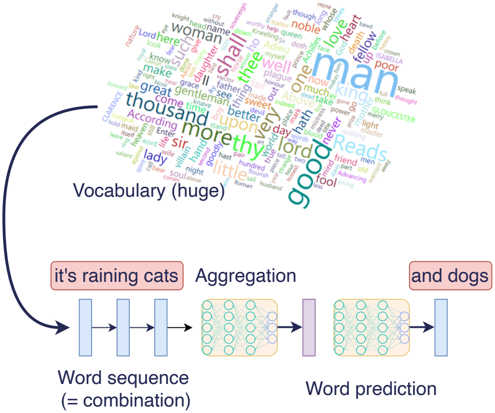
As of 2021, Barack Obama was born on August 4, 1961, so he is 60 years old.

V and today?

Explainability... And complexity







- Simple system
- Exhaustive testing of inputs/outputs
- **Predictable** & **explainable**

- Large dimension
- Complex non-linear combinations
- **Non-predictable** & **non-explainable**

Explainability... And complexity

Interpretability *vs* Post-hoc Explanation

Neural networks = **non-interpretable** (almost always)

too many combinations to anticipate

Neural networks = **explainable a posteriori** (almost always)

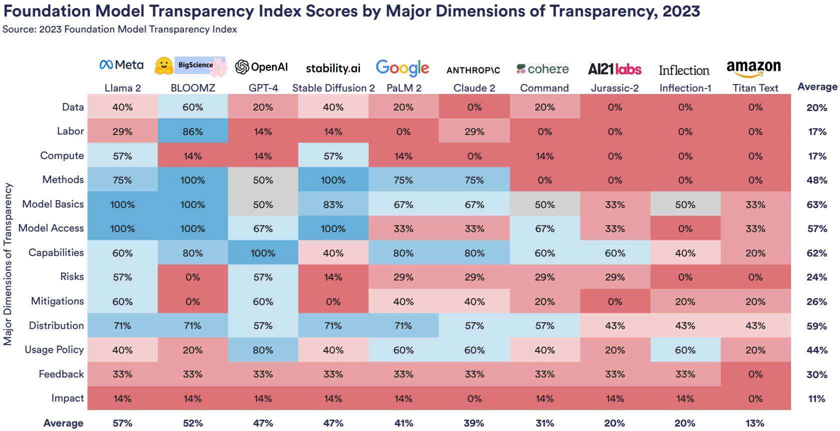
[Uber Accident, 2018]

- Simple system
- Exhaustive testing of inputs/outputs
- **Predictable** & **explainable**

- Large dimension
- Complex non-linear combinations
- **Non-predictable** & **non-explainable**

Transparency : open source / open weight

- Can I modify it? Adaptation
- What training data was used? Data contamination / skills
- What editorial stance / censorship is involved? Access to information
- Why this answer? Explainability / interpretability



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

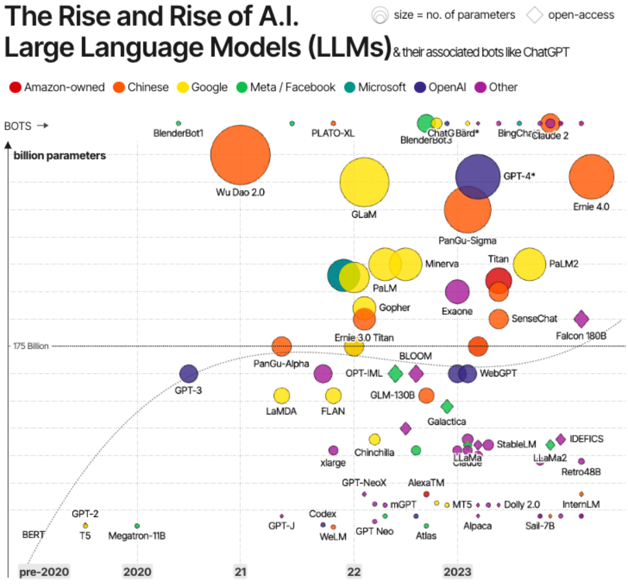
Source: 2023 Foundation Model Transparency Index

Major Dimensions of Transparency

	Meta	BigScience	OpenAI	stability.ai	Google	ANTHROP\C	cohere	AI21labs	Inflection	amazon	
	Llama 2	BLOOMZ	GPT-4	Stable Diffusion 2	PaLM 2	Claude 2	Command	Jurassic-2	Inflection-1	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%	

https://crfm.stanford.edu/fmti/May-2024/index.html

Costs / Frugality



David McCandless, Tom Evans, Paul Barton
Information is Beautiful // UPDATED 2nd Nov 23

source: news reports, LifeArchitect.ai
* = parameters undisclosed // see the data

# Parameters

1998	LeNet-5	= 0.06M
2011	Senna	= 7.3M
2012	AlexNet	= 60M
2017	Transformer	= 65M / 210M
2018	ELMo	= 94M
2018	BERT	= 110M / 340M
2019	GPT2	= 1,500M
2020	GPT3	= 175,000M
2025	Llama-4	= 2,000,000M

Everything beyond the LLM's capabilities/training

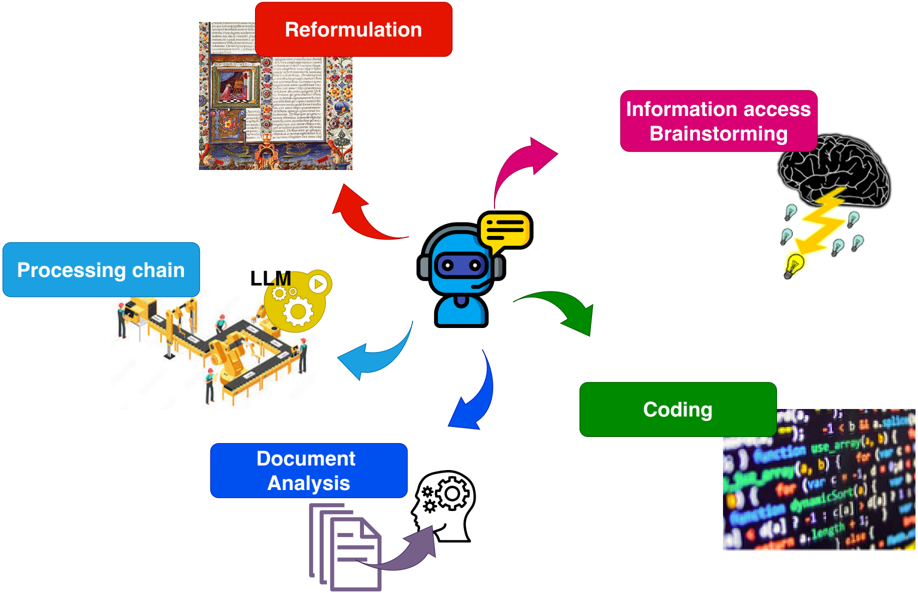
- Simple calculations (multiplication, division)
- Generating *n*-syllable animal names (in progress)
- Playing chess
- Follow (complex) causal reasoning
- ...

ATARI 2600 SCORES STUNNING VICTORY OVER CHATGPT
ChatGPT 4o
VS
WHEN YOU UNDERESTIMATE A 1977 CHESS ENGINE...
AND IT HUMBLES YOU IN FRONT OF THE WHOLE INTERNET

Do you want to play a game of chess?

Large Language Models Uses

Key uses in 5 pictures



Reformulation

Information access
Brainstorming

Processing chain

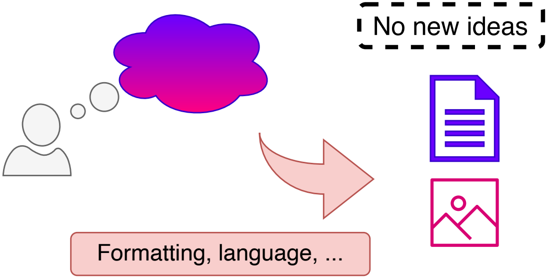
LLM

Coding

Document Analysis

(1) Formatting information

A fantastic tool for **formatting**



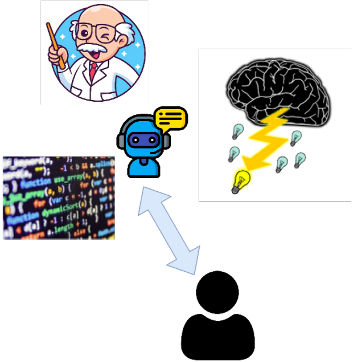
- **Personal assistant**
  - Standard letters, recommendation letters, cover letters, termination letters
  - Translations
- Meeting **reports**
  - Formatting notes
- **Writing** scientific articles
  - Writing ideas, in French, in English

No new information ⇒ just writing, improving, translating, cleaning up, ...

(2) Brainstorming / Course Planning / Statistics Review

- **Find** inspiration [writer’s block syndrome]
- **Organize** ideas quickly
- **Avoid omissions** / increase confidency
- **Search** in a targeted way, adapted to one’s needs
- **Answer** student questions (24/7)
- **Partner** in research, test/enrich ideas

⇒ Impressive answers, sometimes incomplete or partially incorrect... But often useful



- In which areas are LLMs reliable?
- What are the risks for primary information sources?
- What societal risks for information?

(3) Coding: Different Tools, Different Levels

- Providing solutions to exercises
- Learning to code or getting back into it
  - New languages, new approaches (ML?)
  - Benefit from explanations...

But how to handle mistakes?

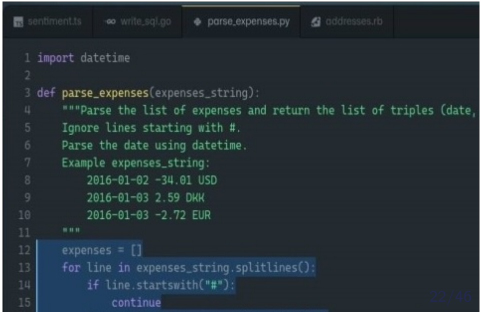
- Help with a library [*getting started*]
- Faster coding

GitHub Copilot

- What about copyrights?
  - What impact on future code processing?
- How to adapt teaching methods?
- How many calls are needed for code completion?

What about the carbon footprint?

- What is the risk of error propagation?

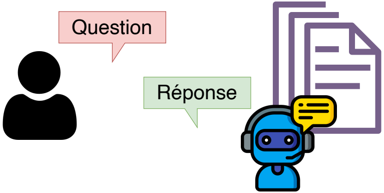


```python
import datetime

def parse_expenses(expenses_string):
    """Parse the list of expenses and return the list of triples (date,
    Ignore lines starting with #.
    Parse the date using datetime.
    Example expenses_string:
        2016-01-02 -34.01 USD
        2016-01-03 2.59 DKK
        2016-01-03 -2.72 EUR
    """
    expenses = []
    for line in expenses_string.splitlines():
        if line.startswith("#"):
            continue
```

(4) Document Analysis

- Summarizing documents / articles
- Dialoguing with a document database
- Assistance in writing reviews
- FAQs, internal support services within companies
- Technology watch
- Generating quizzes from lecture notes



NotebookLM

Think Smarter, Not Harder

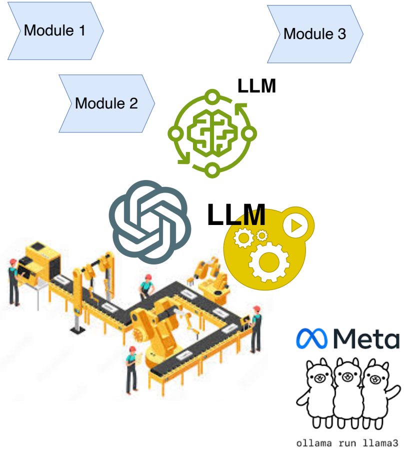
Try NotebookLM

- Will articles still be read in the future?
  - Should we make our articles NotebookLM-proof?
- How to save time while remaining honest and ethical?

(5) LLM in a Production Pipeline / Agentic AI

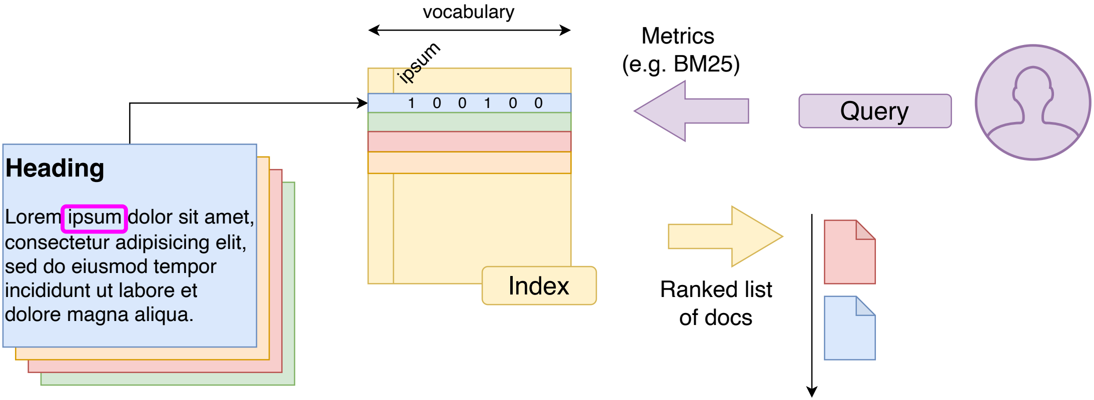
- Run LLM locally
- Extract knowledge
- Generate examples to train a model
  [Teacher/student - distillation]
- Generate variants of examples $\nearrow\nearrow$ increase dataset size
  [Data augmentation]

⇒ Integrate the LLM into a processing pipeline
= little/less supervision = **Agentic AI**

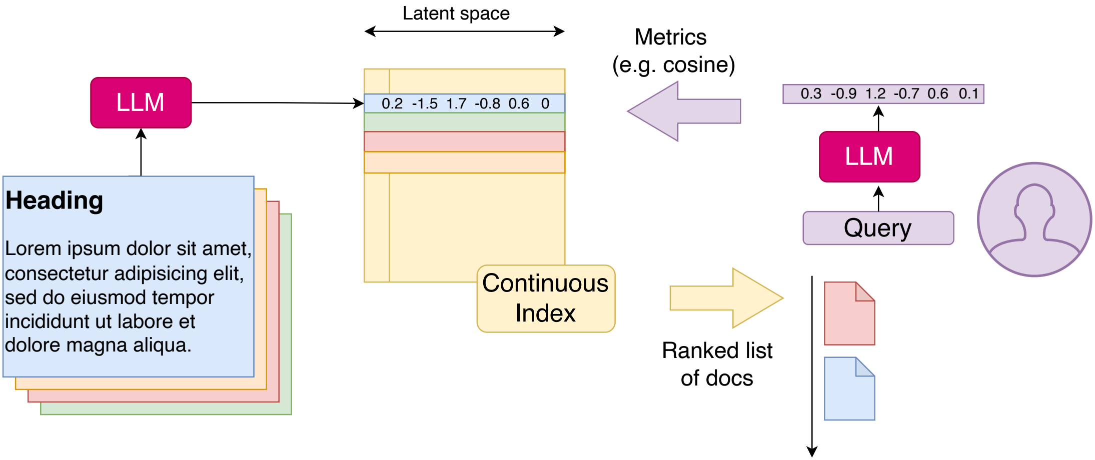


- How much does it cost? ($ + CO_2) Need for GPUs?
- How good are open-weight models?
- How to build multiple agents?

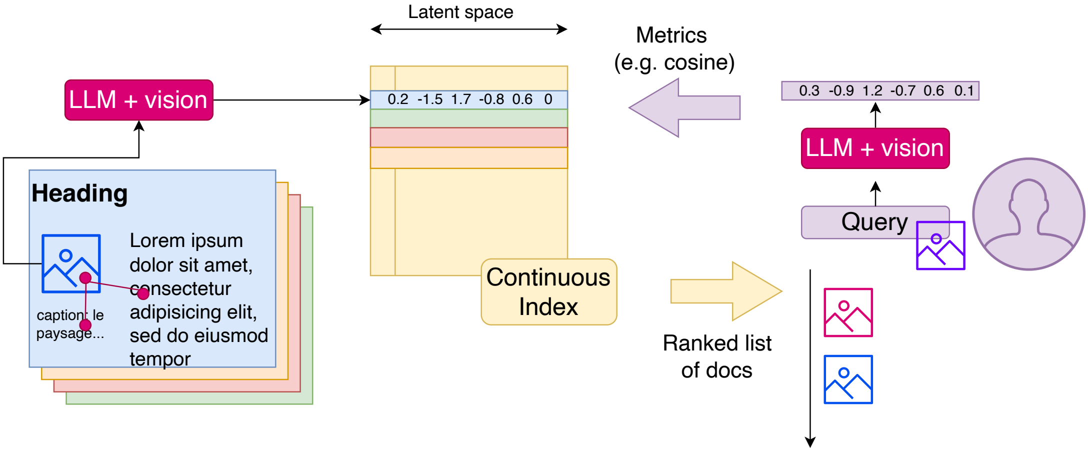
LLM *vs* Information Retrieval


vocabulary
ipsum
1 0 0 1 0 0
Metrics
(e.g. BM25)
Query
Heading
Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.
Index
Ranked list
of docs

LLM *vs* Information Retrieval



LLM *vs* Information Retrieval



Latent space

Metrics
(e.g. cosine)

0.3 -0.9 1.2 -0.7 0.6 0.1

LLM + vision

0.2 -1.5 1.7 -0.8 0.6 0

LLM + vision

Query

Heading

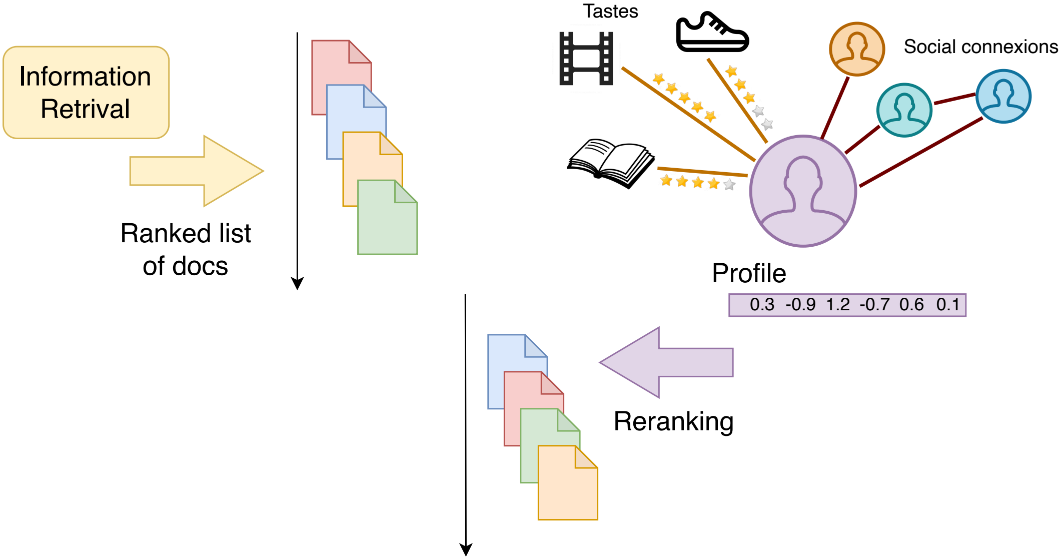
Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor

caption: le paysage...

Continuous Index

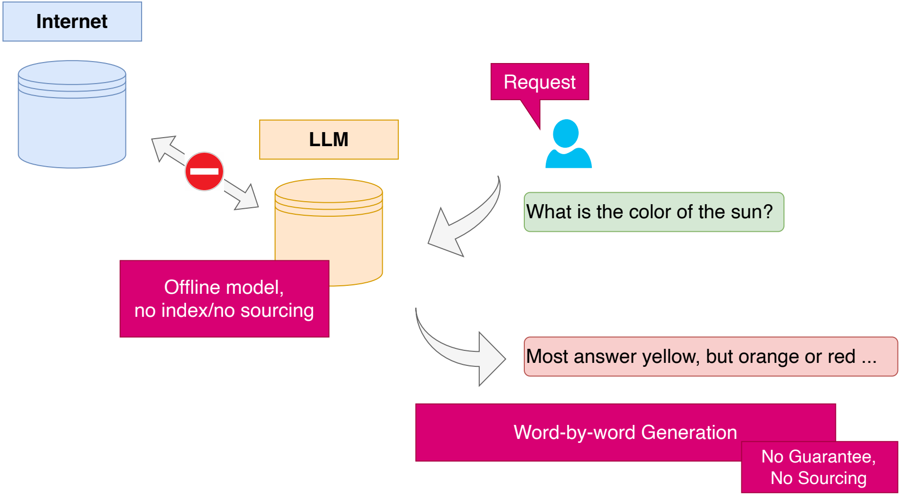
Ranked list of docs

LLM *vs* Information Retrieval

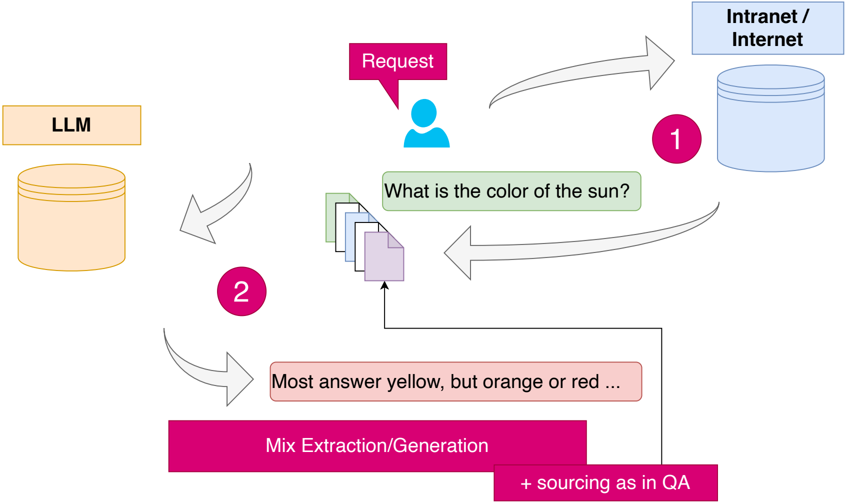


LLMs ⇒ RAG : parametric memory *vs* Info. Extraction

- Asking for information from ChatGPT... A surprising use!
- But is it reasonnable? [Real Open Question (!)]



LLMs ⇒ RAG : parametric memory *vs* Info. Extraction


Intranet / Internet
Request
LLM
1
What is the color of the sun?
2
Most answer yellow, but orange or red ...
Mix Extraction/Generation
+ sourcing as in QA

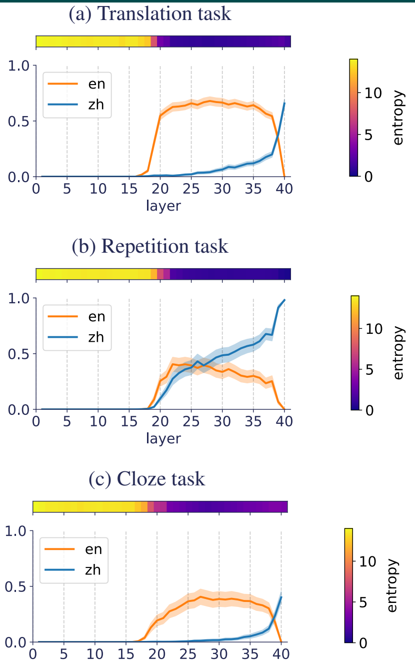
- RAG: Retrieval Augmented Generation
- (Current) limit on input size (2k then 32k tokens)

Language Handling

- Language models are (mostly) multilingual:

⇒ Think in the language you are most comfortable with
⇒ Ask for answers in the target language

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



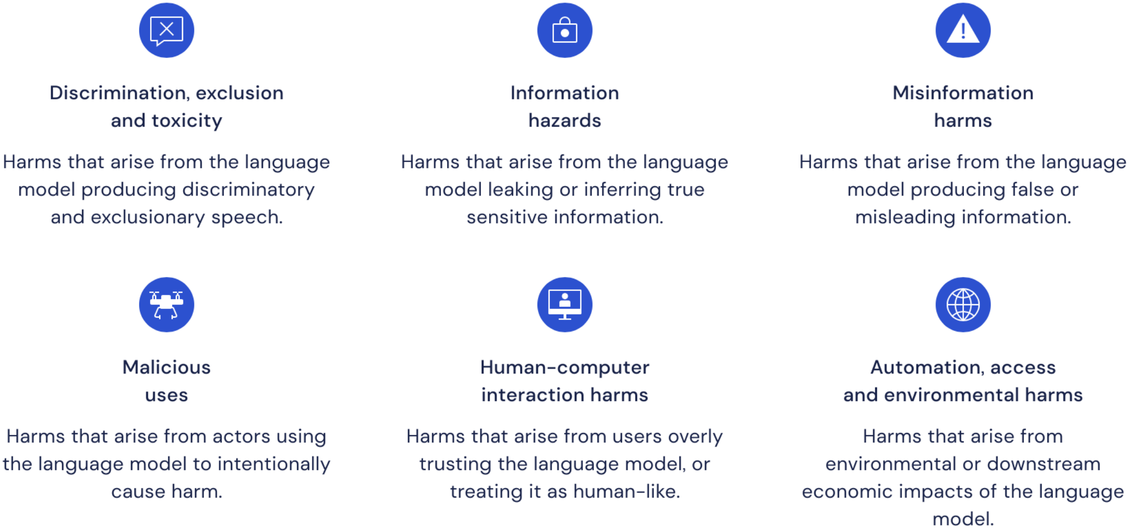
(a) Translation task

(b) Repetition task

(c) Cloze task

Risks

Typology of AI Risks in NLP (L. Weidinger)



Discrimination, exclusion and toxicity

Harms that arise from the language model producing discriminatory and exclusionary speech.

Information hazards

Harms that arise from the language model leaking or inferring true sensitive information.

Misinformation harms

Harms that arise from the language model producing false or misleading information.

Malicious uses

Harms that arise from actors using the language model to intentionally cause harm.

Human-computer interaction harms

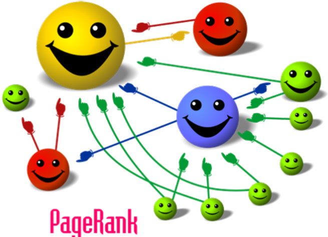
Harms that arise from users overly trusting the language model, or treating it as human-like.

Automation, access and environmental harms

Harms that arise from environmental or downstream economic impacts of the language model.

Access to Information

- Access to dangerous/forbidden information
  - +Personal data
  - Right to be forgotten (GDPR)
- Information authorities
  - Nature: unconsciously, image = truth
  - Source: newspapers, social media, ...
  - Volume: number of variants, citations (pagerank)
- Text generation: harassment...
- Risk of anthropomorphizing the algorithm
  - Distinguishing human from machine



Machine Learning & Bias

Mustache, Triangular Ears, Fur Texture

Cat

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

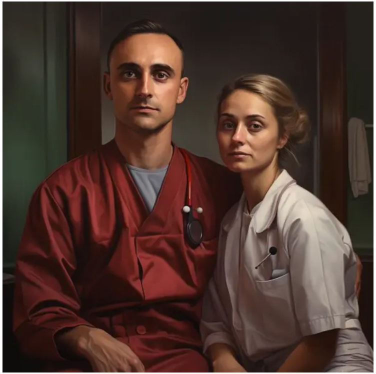
Senior Executive

Bias in the data ⇒ bias in the responses

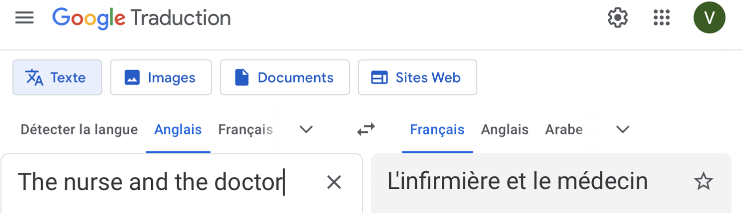
Machine learning is based on extracting statistical biases...

⇒ Fighting bias = manually adjusting the algorithm

Machine Learning & Bias



Sterreotypes from *Pleated Jeans*



The nurse and the doctor

L'infirmière et le médecin

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

Bias in the data ⇒ bias in the responses

Machine learning is based on extracting statistical biases...

⇒ 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?

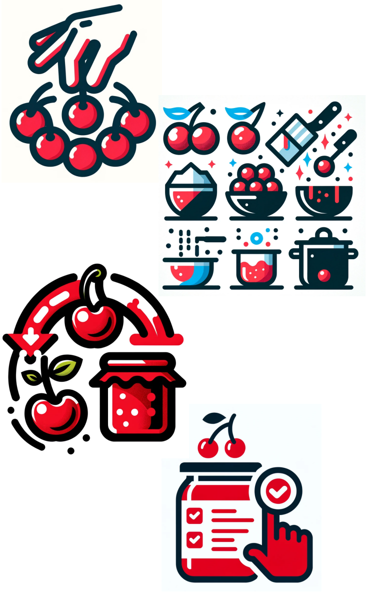
- Domain experts / specifications
- Engineers, during algorithm design
- Ethics group, during result validation
- Communication group / user response

⇒ What legitimacy? What transparency? What effectiveness?

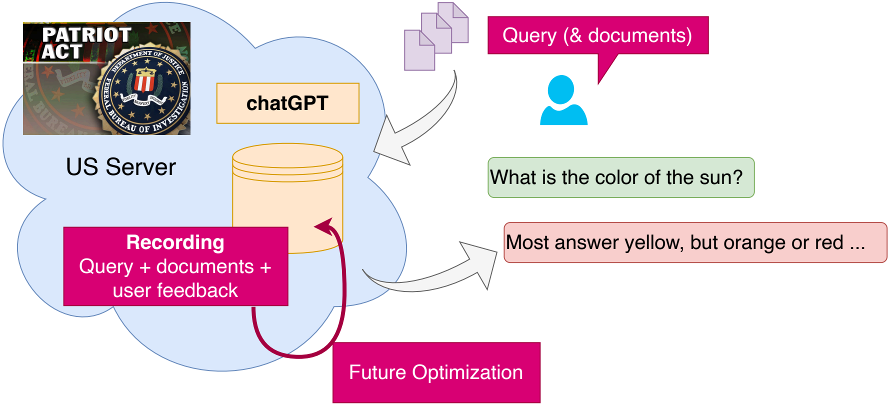
Machine learning is never neutral

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



Data Leak(s): different security levels


PATRIOT ACT
DEPARTMENT OF JUSTICE
FEDERAL BUREAU OF INVESTIGATION
FIDELITY BRAVERY INTEGRITY
chatGPT
US Server
Recording
Query + documents +
user feedback
Future Optimization
Query (& documents)
What is the color of the sun?
Most answer yellow, but orange or red ...

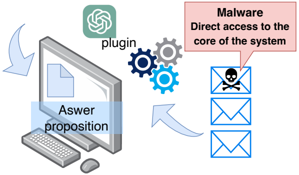
- Transfer of sensitive data
- Exploitation of data by OpenAI (or others)
- Data leakage in future models

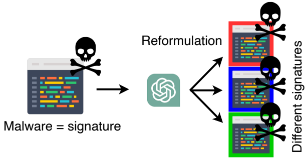
Data Leak(s): different security levels

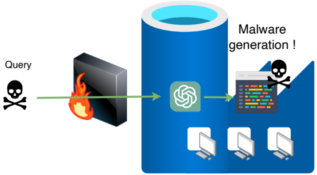
Level 1: **Commercial tools, free to use**	Variable licenses (depending on the companies and subject to change over time). Uncertain data protection, risk to personal data. *chatGPT, Mistral, Perplexity, ...*
Level 2: **Commercial tools, paid licence**	Strong contractual guarantees. Risks associated with the *Patriot Act*. Possible to enforce non-storage of queries. *chatGPT, Mistral, Perplexity, ...*
Level 3: **Local dev., Commercial tools & paid licence ++**	+ Negotiation on the server location/data security. *Microsoft Azur, Mistral, AWS, Aristote, Ragarenn...*
Level 4: **Local use**	Use of a locally operated LLM, with no data transferred over the web. *HuggingFace, Ollama, ...*

Security Issues

- Plug-ins ⇒ 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
    - ≈ software rephrasing
- New problems!
    - Direct malware generation



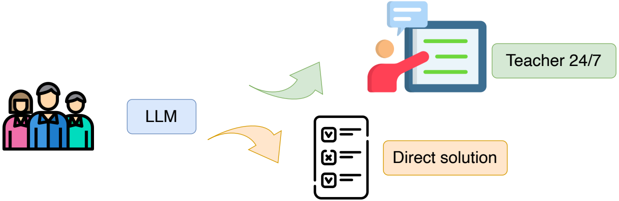




Educational Challenges

- Redefine our educational priorities, subject by subject, as we did with Wikipedia/calculator/...
  - Accept the decline of certain skills

- Train students in the use of LLMs, while managing to temporarily prohibit their use

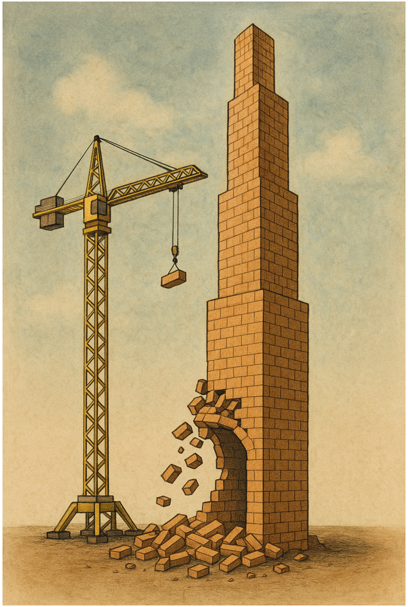


LLM

Teacher 24/7

Direct solution

- Learn to recognize LLM-generated content, use detection tools.



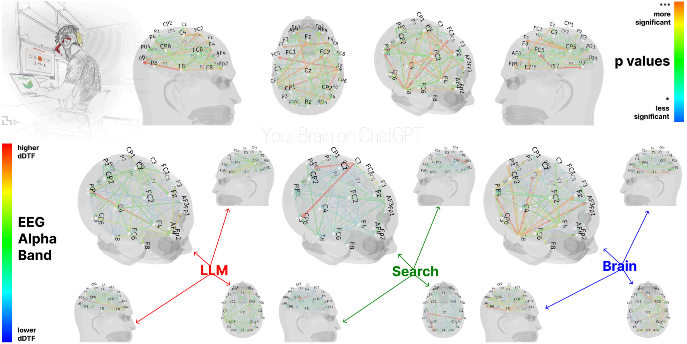
Decline / Evolution of Cognitive skills

Our brain will evolve with these new tools...

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

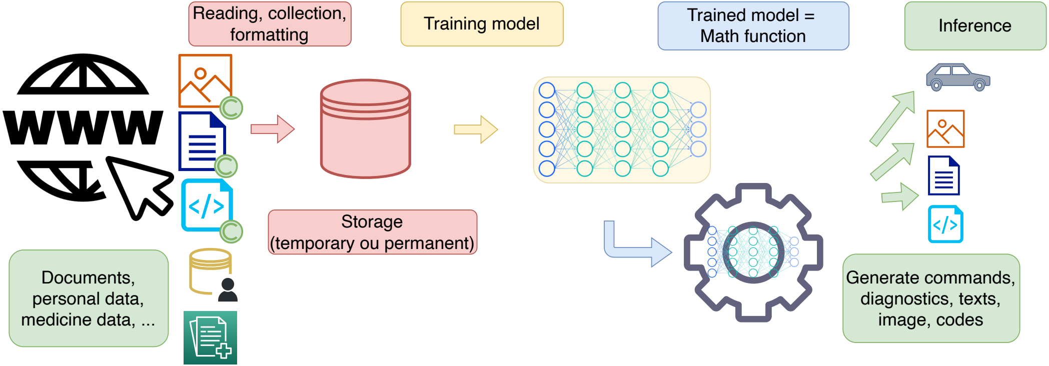
- Education sciences and psychology had conjectured it...

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

Legal Risks/Questions



Reading, collection, formatting

Training model

Trained model = Math function

Inference

Storage (temporary ou permanent)

Documents, personal data, medicine data, ...

Generate commands, diagnostics, texts, image, codes

Copyright and database law

Right to collect, right to copy, consent

Right to use data in an algorithm **Optout**

Model = emanation of data?

Clearview.ai

Cambridge Analytica

Reproductions of untraceable extracts

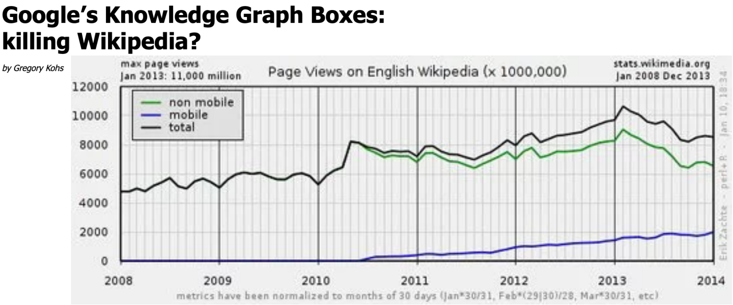
Usage regulation

Responsibility for errors

Economic Questions

- Funding/Advertising ⇔ **visits** by internet users
- Google knowledge graph (2012) ⇒ fewer visits, less revenue
- chatGPT = encoding web information... ⇒ much fewer visits?

⇒ What **business model for information sources** with chatGPT?



Google's Knowledge Graph Boxes: killing Wikipedia?

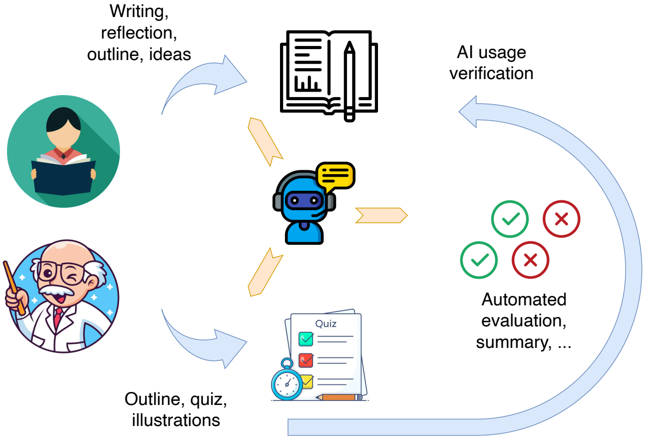
by Gregory Kohs

⇒ Who does **benefit from the feedback**? [StackOverFlow]

Risks of AI Generalization

AI everywhere =
loss of meaning?

- In the educational domain
- Transposition to HR
- To project-based funding systems



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

Conclusion

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
  - W
  - W

The smartphone has made me an *augmented human*...

Will the LLM make me an *augmented researcher*?

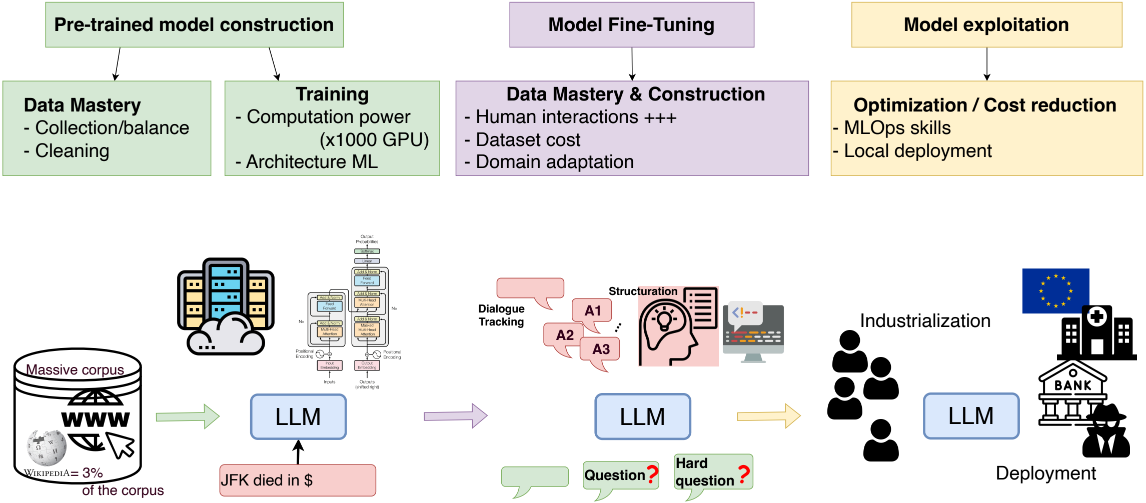
⇒ Still, have a look at NotebookLM

- **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-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: $\geq$ 1 year

Digital Sovereignty: the Entire Chain



Pre-trained model construction

Data Mastery
- Collection/balance
- Cleaning

Training
- Computation power (x1000 GPU)
- Architecture ML

Model Fine-Tuning

Data Mastery & Construction
- Human interactions +++
- Dataset cost
- Domain adaptation

Model exploitation

Optimization / Cost reduction
- MLOps skills
- Local deployment

A Multitude of Professions

Data architect / manager

- Data management & hardware devices (storage, network, …)

Data Engineer

- Update & Query on the data

+ DPO :
Data Protection Officer

Data Analyst

- Data visualization (chart, indicators, …)
- Statistical trends

Prompt Engineer

- Query on LM/foundation models with "prompts"

Data Scientist

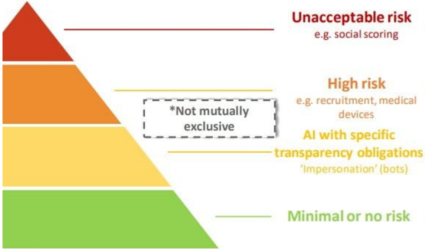
- Query the data / critical selection & balance
- Algorithm development / adaptation / evaluation
- Advanced data visualization

MLOps Engineer

- Algorithm optimization
- Industrialize software solutions

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?