USAGE(S) DE L'IA À AGROPARISTECH

Lundi 15 Septembre 2025

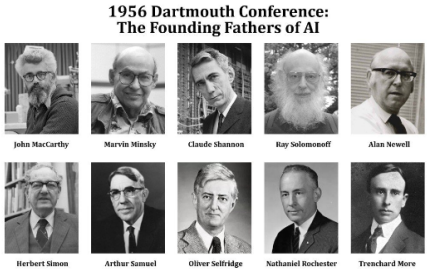
Vincent Guigue
vincent.guigue@agroparistech.fr
https://vguigue.github.io

From AI to machine-learning

A Rapid Tour of Artificial Intelligence

The Birth of Computer Science... And of Artificial Intelligence

A. Turing



Dartmouth conference creation of IA

G. Hinton

Y. Lecun
Automated cheque reading

E. Dickmanns : 1000km in autonomous vehicle

IBM Deeper Blue

First computer

DARPA

Backpropagation

1941 1950 1956 1958 1963 1965 1972 1979 1986 1987 1990 1992 1993 1994 1997

Dynamic programming

LISP Language
Perceptron

PROLOG

1st Neural Network

DARPA MUC

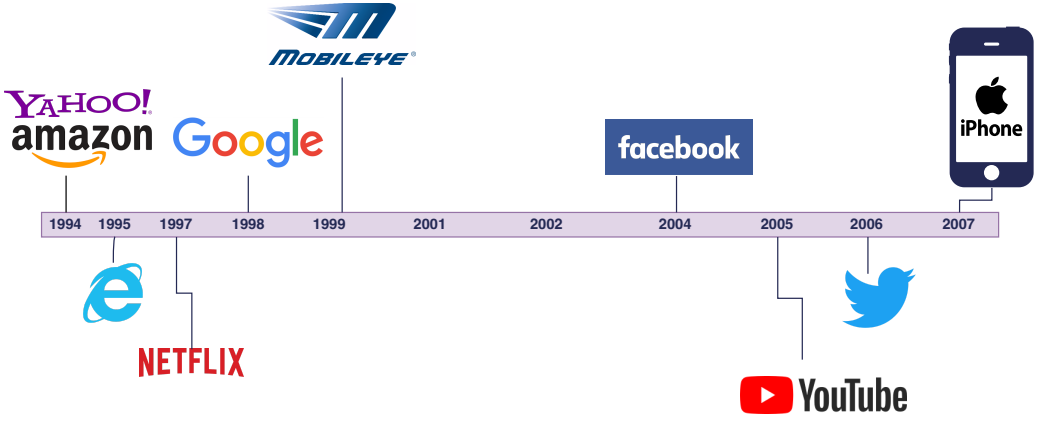
DARPA TREC

SVM

C4.5
Decision Tree

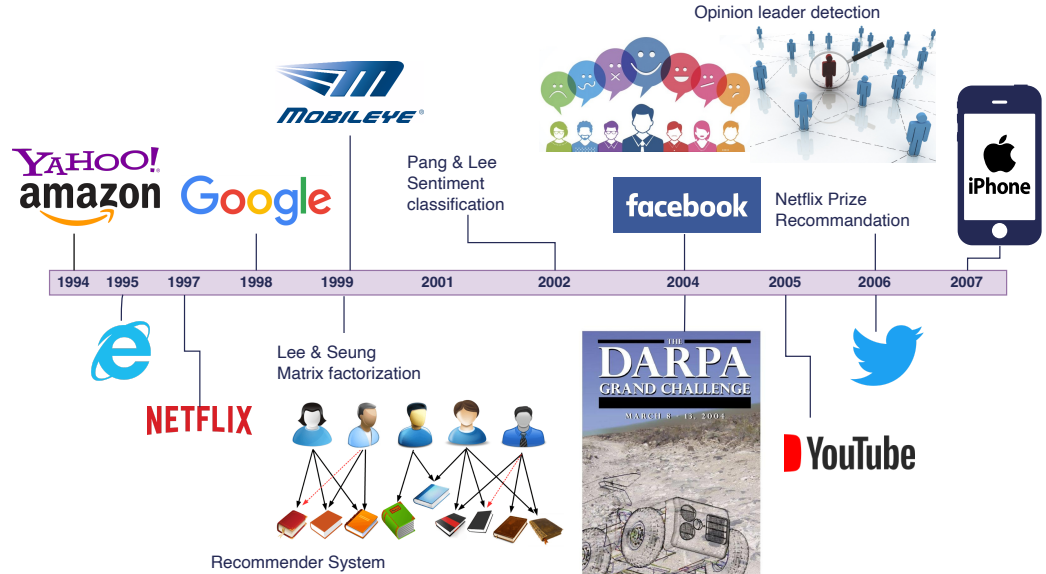
A Rapid Tour of Artificial Intelligence

Emergence (or Refoundation) of the GAFAM/GAMMA



A Rapid Tour of Artificial Intelligence

Emergence (or Refoundation) of the GAFAM/GAMMA

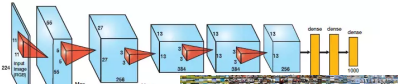


A Rapid Tour of Artificial Intelligence

A Wave of Artificial Intelligence

Thrun:
DARPA Gd Challenge
victory

Google
car



Hinton/Krizhevsky
Deep-Learning

Translate (v2)

K. Cho
Traduction auto.

AlphaGo

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

Y. Koren
Challenge
Netflix

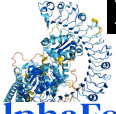
kaggle

amazon alexa

Google
DeepMind
Acquisition : $400M

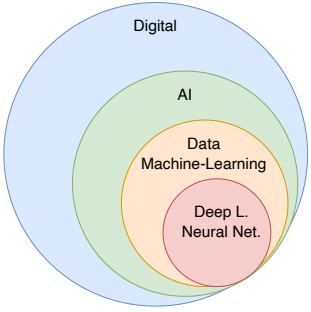
IBM Jeopardy win

Acquisition :
$15B



AlphaFold

Artificial Intelligence & Machine Learning



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

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

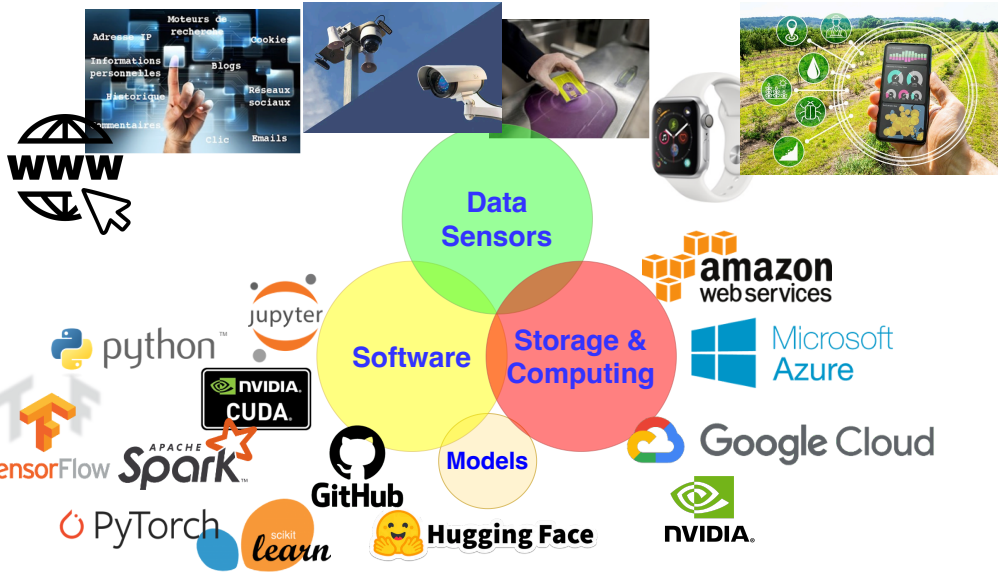
Marvin Lee Minsky, 1956

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

$\neq$ **G-AI (General AI)**, which replaces humans in complex systems.

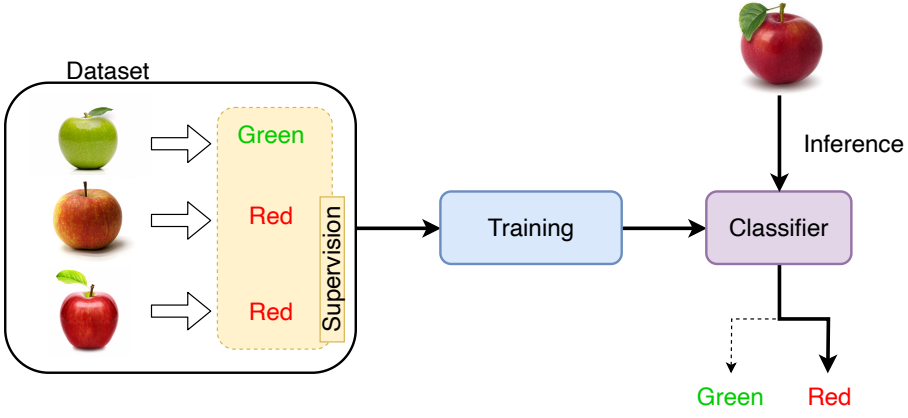
Andrew Ng, 2015

The Ingredients of Artificial Intelligence



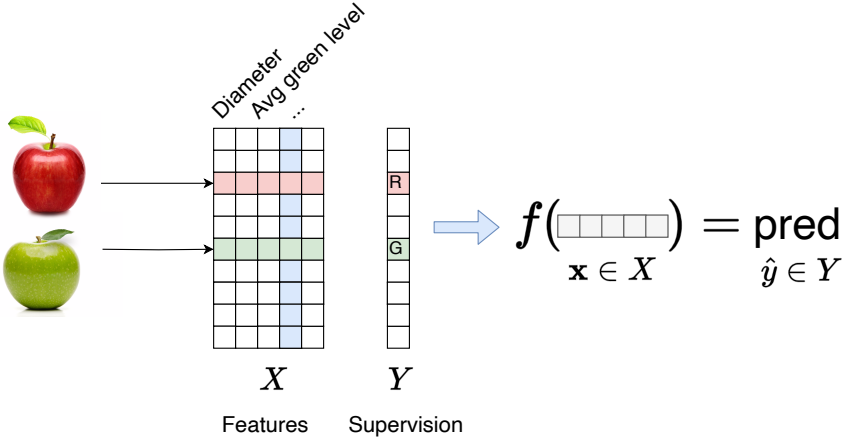
Machine Learning Definition

1. Collecting labeled **dataset**
2. Training **classifier**
3. Exploiting the model



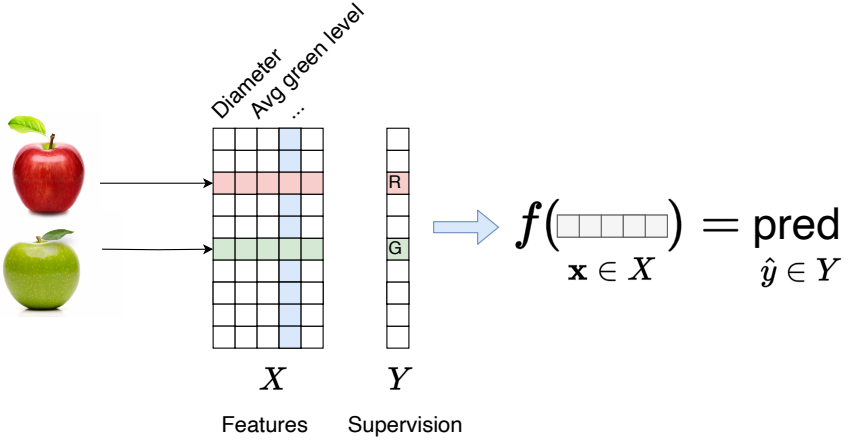
Machine Learning Definition

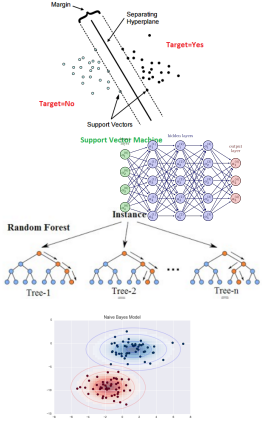
1. Collecting labeled **dataset**
2. Training **classifier**
3. Exploiting the model



Machine Learning Definition

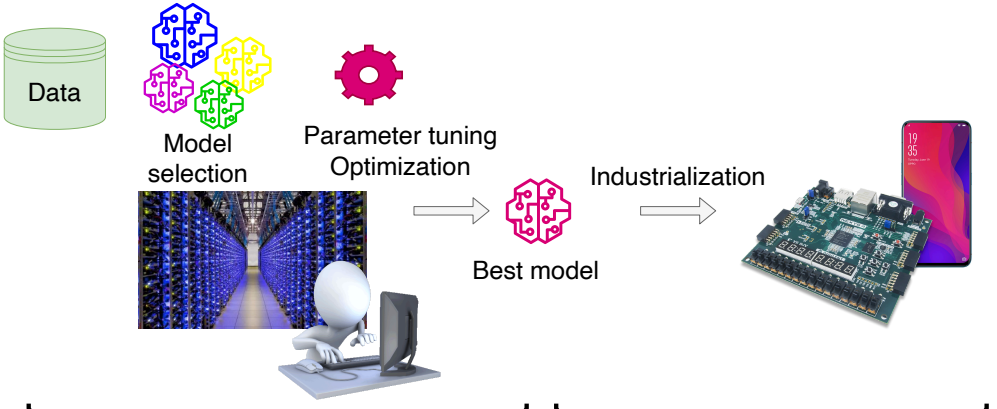
1. Collecting labeled **dataset**
2. Training **classifier**
3. Exploiting the model





Data Processing Chain

Different steps in machine-learning



Deep learning & Representation Learning

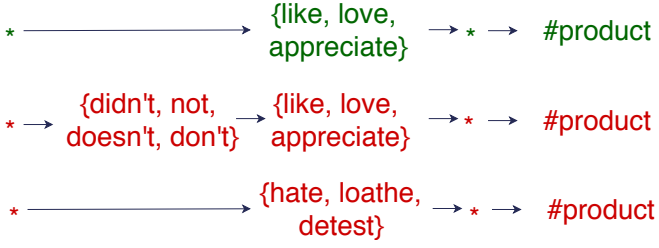
[Application to textual data]

AI + Textual Data: Natural Language Processing (NLP)

NLP = largest scientific community in AI

Linguistics [1960-2010]

Rule-based Systems:

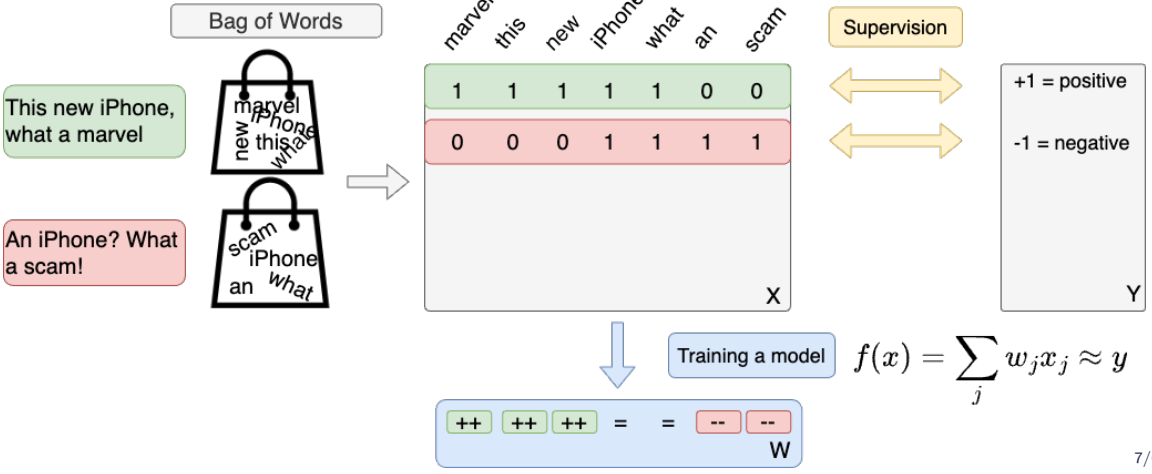


- Requires expert knowledge
- Rule extraction $\Leftrightarrow$ very clean data
- Very high precision
- Low recall
- Interpretable system

AI + Textual Data: Natural Language Processing (NLP)

NLP = largest scientific community in AI

Machine Learning [1990-2015]



AI + Textual Data: Natural Language Processing (NLP)

NLP = largest scientific community in AI

Linguistics [1960-2010]

- Requires expert knowledge
- Rule extraction $\Leftrightarrow$ very clean data

\+ Interpretable system
\+ Very high precision
− Low recall

Machine Learning [1990-2015]

- Little expert knowledge needed
- Statistical extraction $\Leftrightarrow$ robust to noisy data

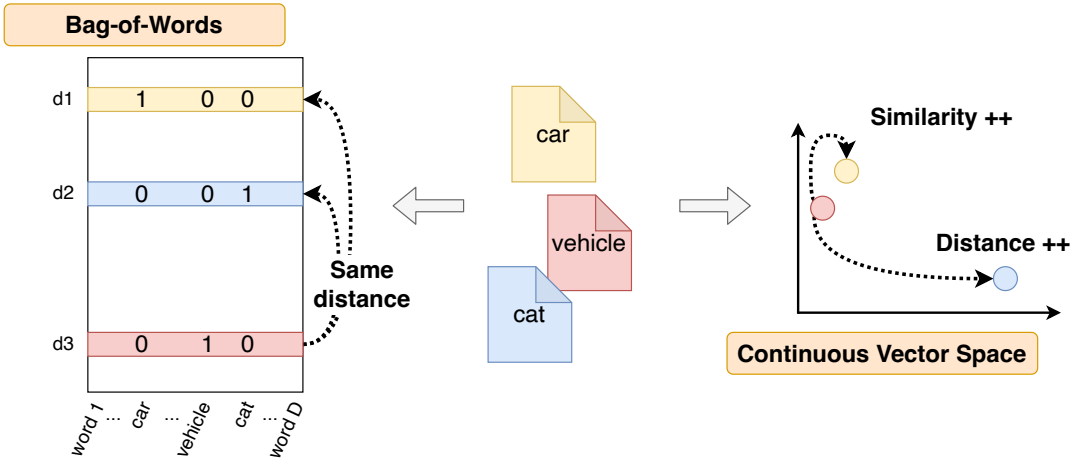
$\approx$ Less interpretable system
− Lower precision
\+ Better recall

Precision = criterion for acceptance by industry

→ Link to metrics

Deep/Representation Learning for Text Data

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

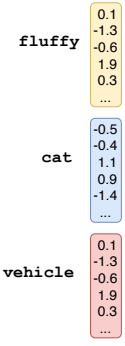


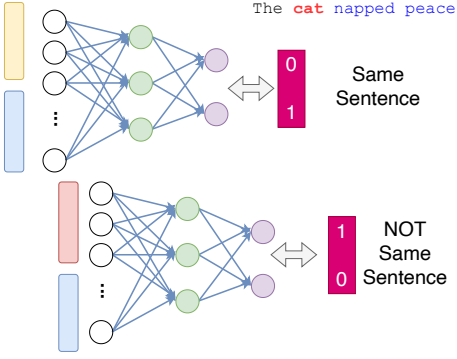
LeCun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

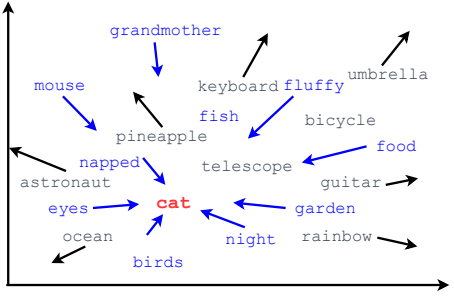
Deep/Representation Learning for Text Data

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

The fluffy cat napped lazily in the sunbeam.
I adopted a stray cat from the shelter last week.
My cat loves to chase after toy mice.
The black cat stealthily crept through the dark alley.
I often find my cat perched on the windowsill, watching birds.
She gently stroked her cat's fur as it purred contentedly.
Our neighbor's cat frequently visits our backyard.
My cat has a preference for fish flavored cat food.
The cat stealthily stalked a mouse in the garden.
My grandmother has a collection of porcelain cat figurines.
The cat napped peacefully in the warm sunlight.

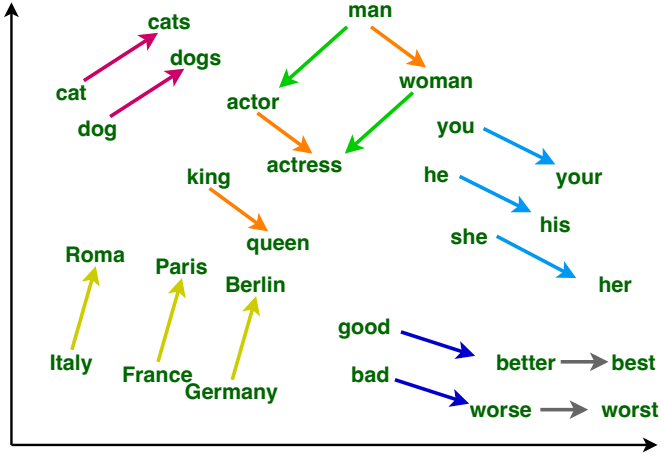






Deep/Representation Learning for Text Data

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



- Semantic Space:
  similar meaning
  $\Leftrightarrow$
  close position
- Structured Space:
  grammatical regularities,
  basic knowledge, ...

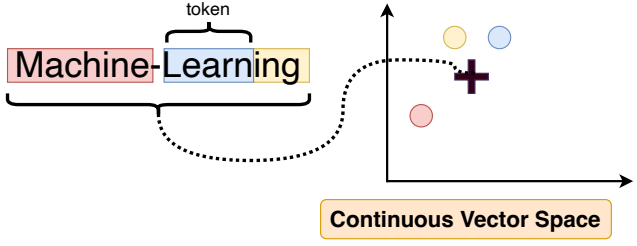
Distributed representations of words and phrases and their compositionality, Mikolov et al. NeurIPS 2013

Deep/Representation Learning for Text Data

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

From Words to Tokens

Word Piece statistical split

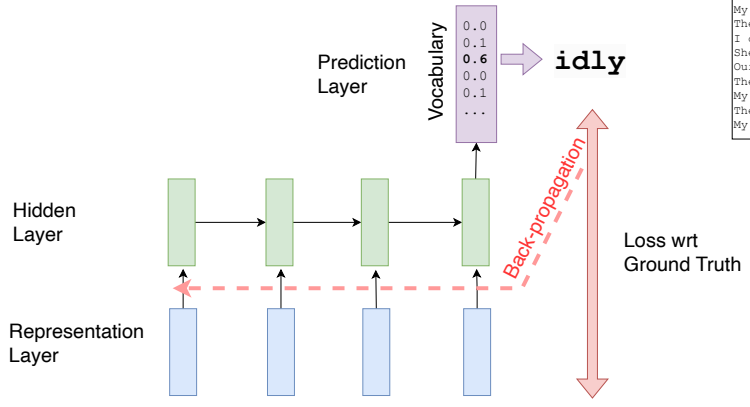


- Representation of unknown words
- Adaptation to technical domains
- Resistance to spelling errors

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

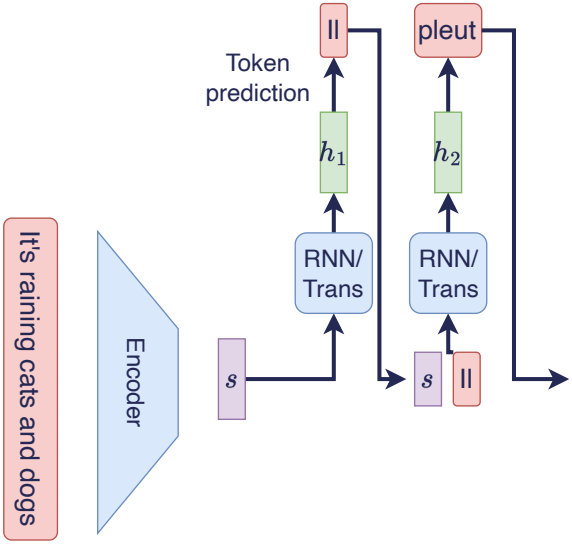


The fluffy cat napped lazily in the sunbeam.
I adopted a stray cat from the shelter last week.
My cat loves to chase after toy mice.
The black cat stealthily crept through the dark alley.
I often find my cat perched on the windowsill, watching birds.
She gently stroked her cat's fur as it purred contentedly.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

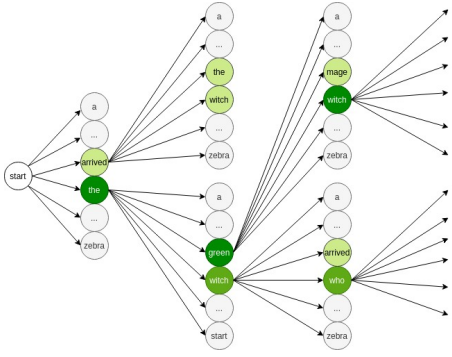
Corpus

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

Inference & Beam Search



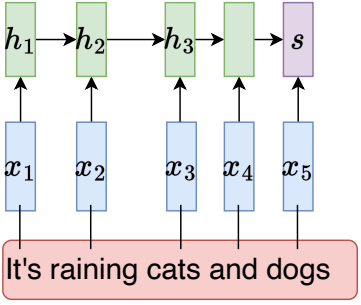
- High cost $\approx$ 1 call / token
- Max. likelihood principle
- NLP historical task =
  - specific classif./scoring archi.
  - constraint and/or post processing on generative archi.



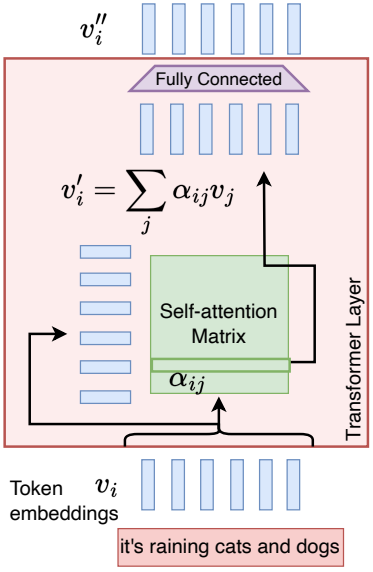
Transformer architecture: state-of-the-art aggregation

Recurrent Neural Network:

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



Transformer:

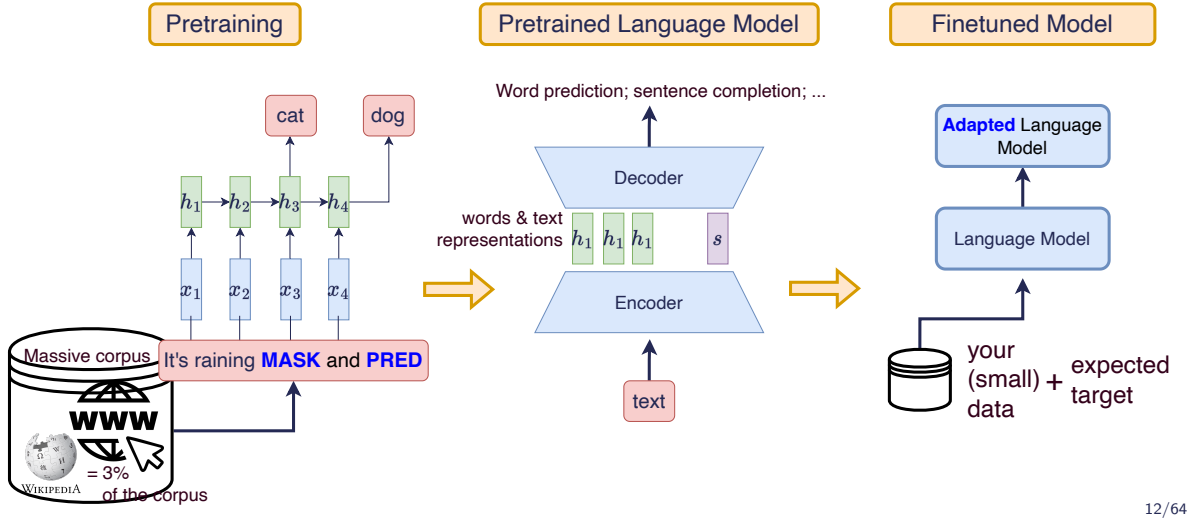


Attention is all you need, Vaswani et al. NeurIPS 2017

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

A new developpement paradigm since 2015

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



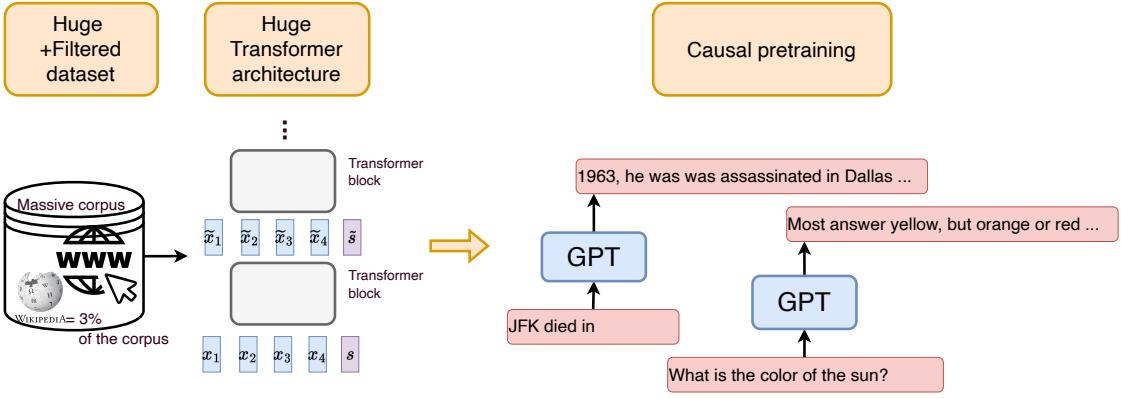
chatGPT

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)



- Grammatical skills: singular/plural agreement, tense concordance
- Knowledges

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

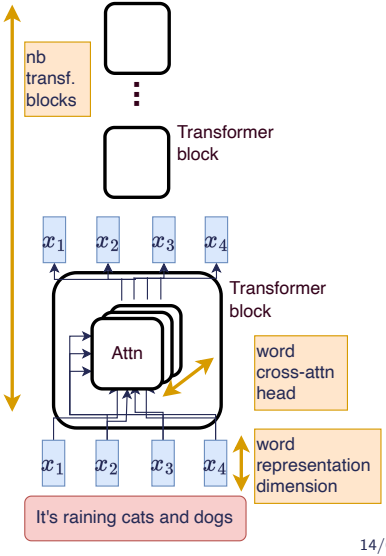
The Ingredients of chatGPT

1. More is better! (GPT)

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

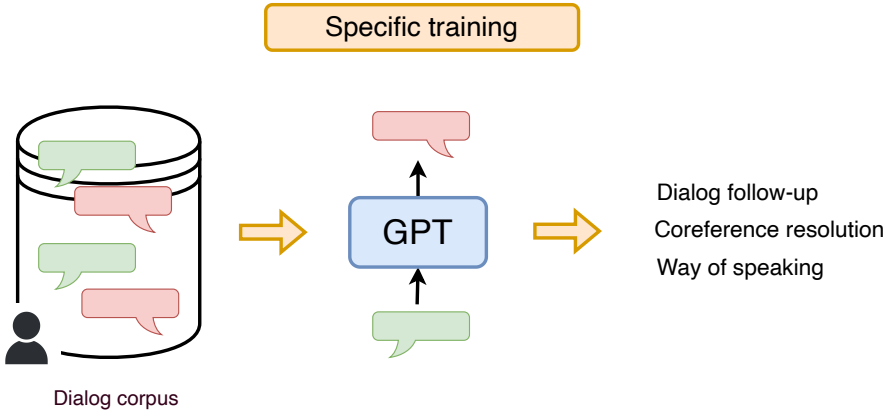
175 Billion parameters... What does it mean?

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



The Ingredients of chatGPT

2. Dialogue Tracking

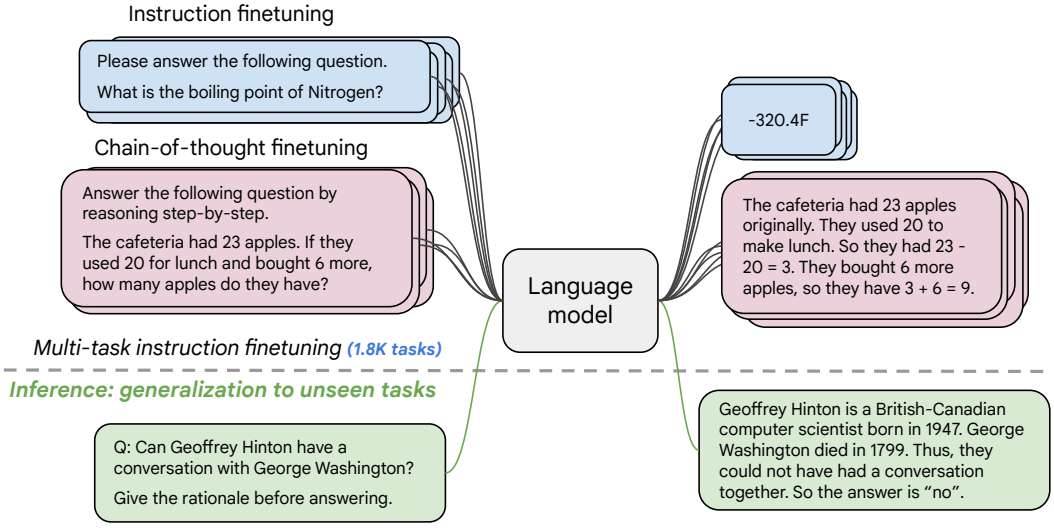


- **Very clean** data

Data generated/validated/ranked by humans

The Ingredients of chatGPT

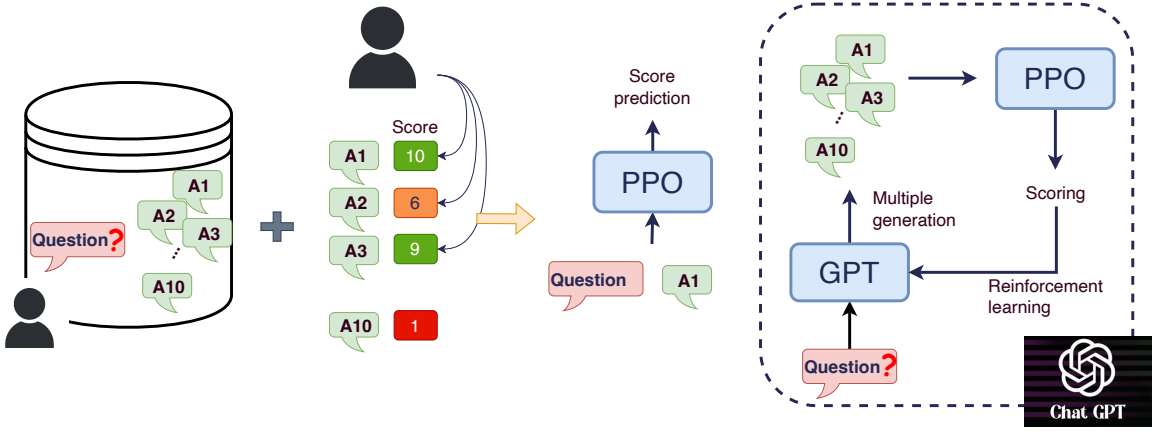
3. Fine-tuning on different (±) complex reasoning tasks



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

The Ingredients of chatGPT

4. Instructions + answer ranking

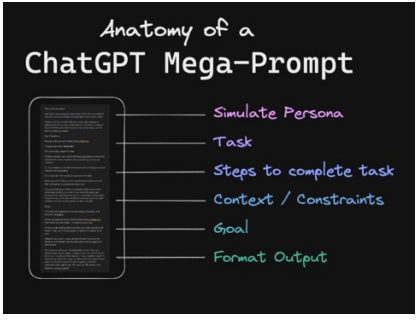


- Database created by humans
- Response improvement / alignment
- ... Also a way to avoid critical topics = censorship

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

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 ...*
  - 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 (!)



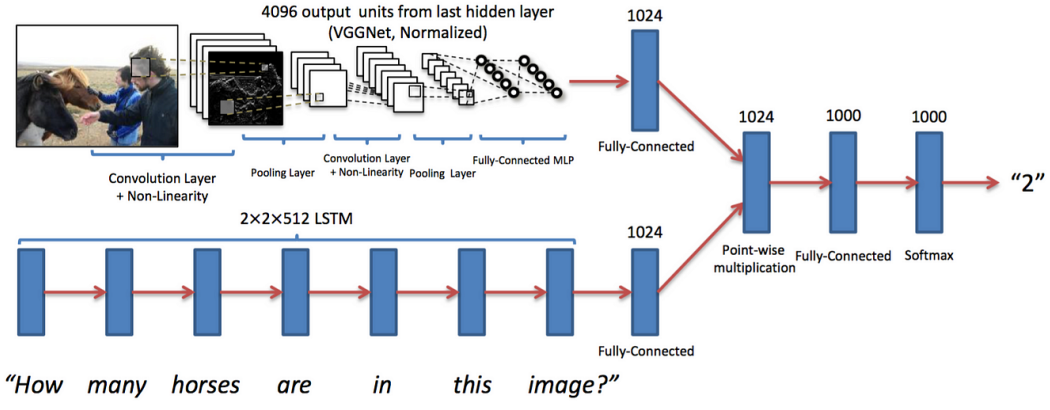
https://chatgptprompts.guru/what-makes-a-good-chatgpt-prompt/

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

GPT4 & Multimodality

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

The example of VQA: visual question answering

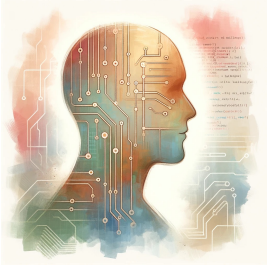


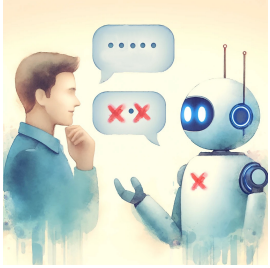
$\Rightarrow$ Backpropagate the error $\Rightarrow$ modify word representations + image analysis

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

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
- Misinterpreted feedback
  - Anthropomorphization of the algorithm and its errors
- Prohibitive cost: what economic, ecological, and societal model?



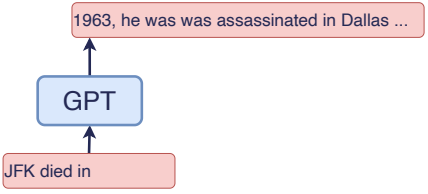


Dall-e generated images

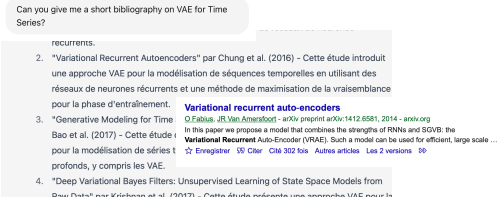
Machine Learning Limits

chatGPT and the relationship with truth

1. Plausibility = 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



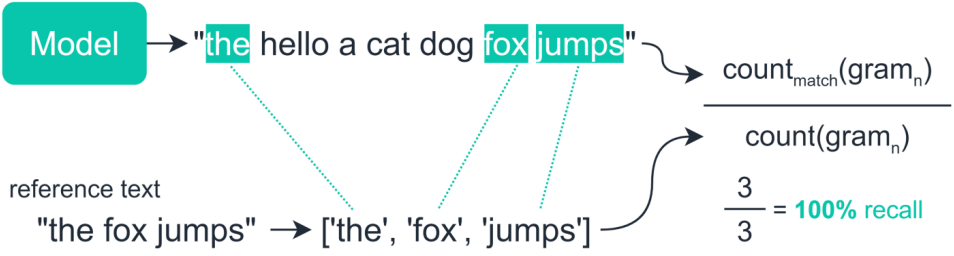
Example: producing a bibliography



Generative AI: how to evaluate performance?

The critical point today

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

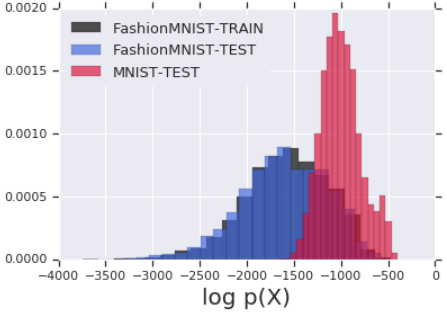


The Ultimate Performance Metric in NLP, J. Briggs, Medium 2021

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

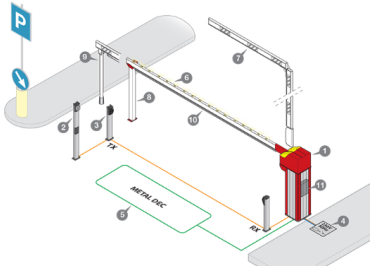
⇒ Little/no use in video games

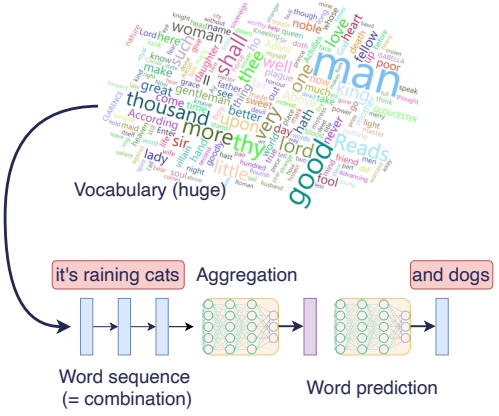
V how old is obama?

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

V and today?

Stability, explainability... And complexity





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

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

Stability, 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

- Model weights (*open-weight*)... ⇒ but not just the weights
- Training data (*BLOOM*) + distribution + instructions
- Learning techniques
- Evaluation

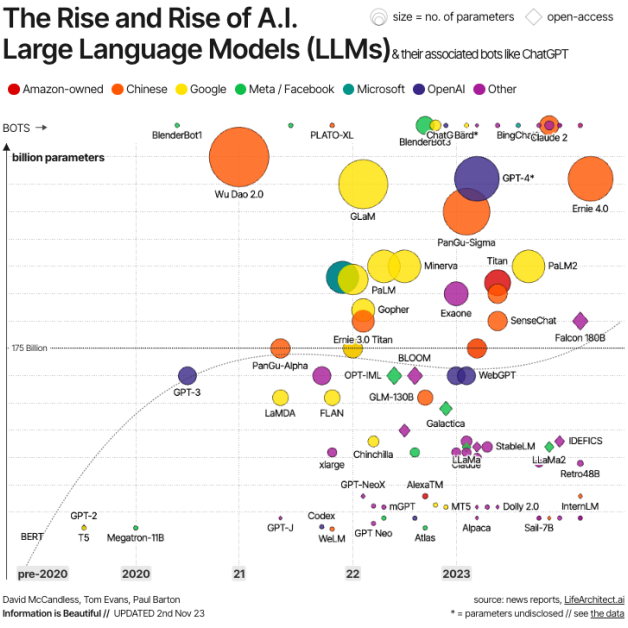
Foundation Model Transparency Index Scores by Major Dimensions of Transparency, 2023

Source: 2023 Foundation Model Transparency Index

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

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

Costs / Frugality

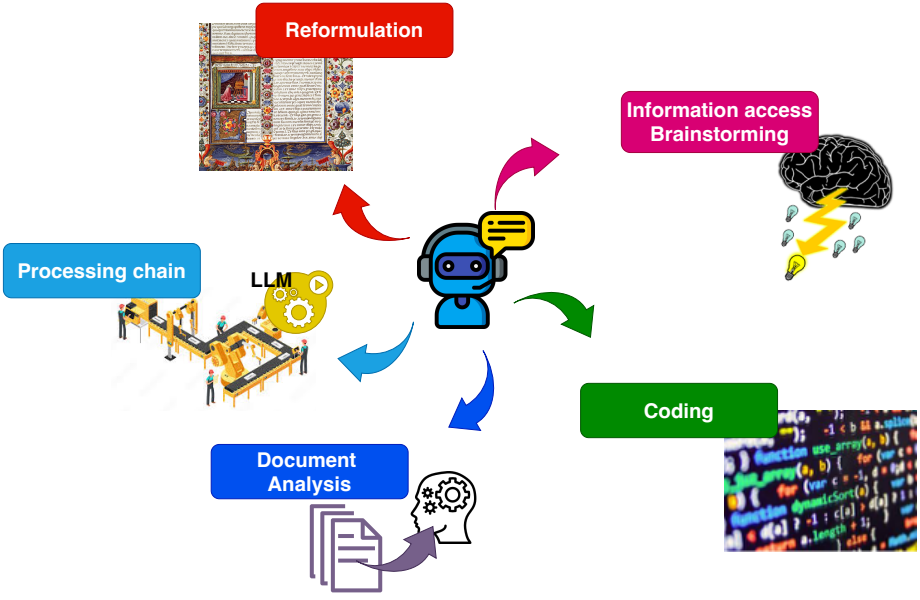


Parameters

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

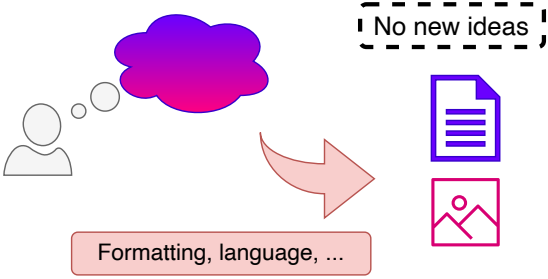
Large Language Models Uses

Key uses in 5 pictures



(1) Formatting information

A fantastic tool for **formatting**



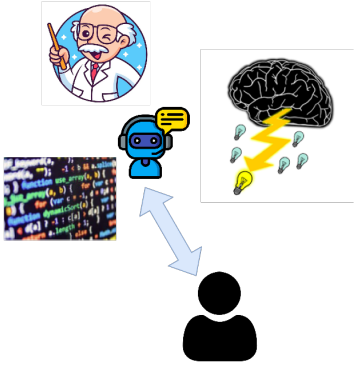
- Personal assistant
  - Standard letters, recommendation letters, cover letters, termination letters
  - Translations
- Meeting reports
  - Formatting notes
- Writing scientific articles
  - Writing ideas, in French, in English
- Document analysis
  - Information extraction, question-answering, ...

⇒ No new information, just writting, 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

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



3 reference articles on the use of transformers in recommendation systems
What is the purpose of the log-normal Poisson law?
Propose 10 sections for a course on Transformers in AI

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

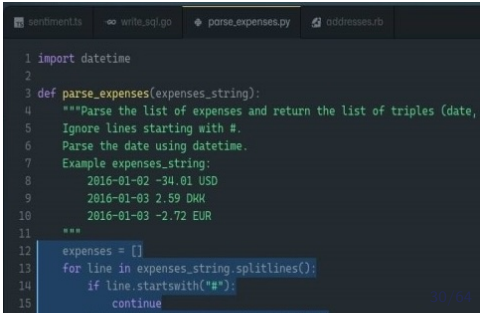
(3) Coding: Different Tools, Different Levels

- 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

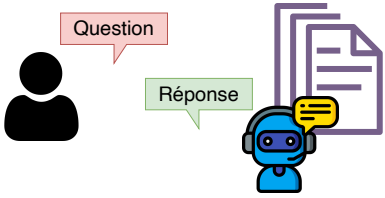
- What about copyrights?
  - What impact on future code processing?
- How to adapt teaching methods?
- How many calls are needed for code completion?

  What about the carbon footprint?
- What is the risk of error propagation?



(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

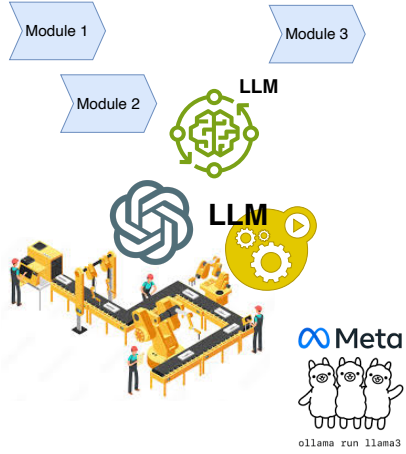


- Will articles still be read in the future?
  - Should we make our articles NotebookLM-proof?
- How to save time while remaining honest and ethical?

(5) LLM in a Production Pipeline / Agentic AI

- Run LLM locally
- Extract knowledge
- Sort documents / generate summaries
- Generate examples to train a model
  [Teacher/student - distillation]
- Generate variants of examples ↗↗ increase dataset size
  [Data augmentation]

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



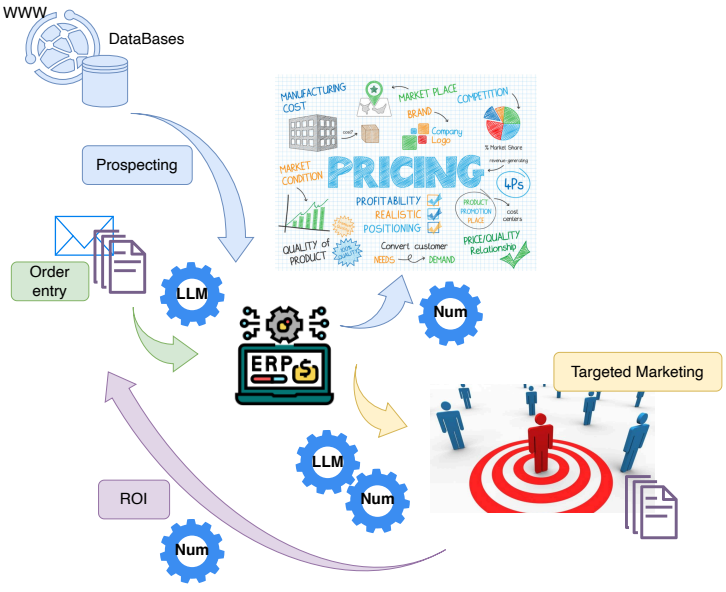
- Can I train models on generated data?
- How much does it cost? ($ + CO_2) Need for GPUs?
- How good are open-weight models?

(5) LLM in a Production Pipeline / Agentic AI

- Run LLM
- Extract kn
- Sort docu
- Generate e
- Generate v dataset siz

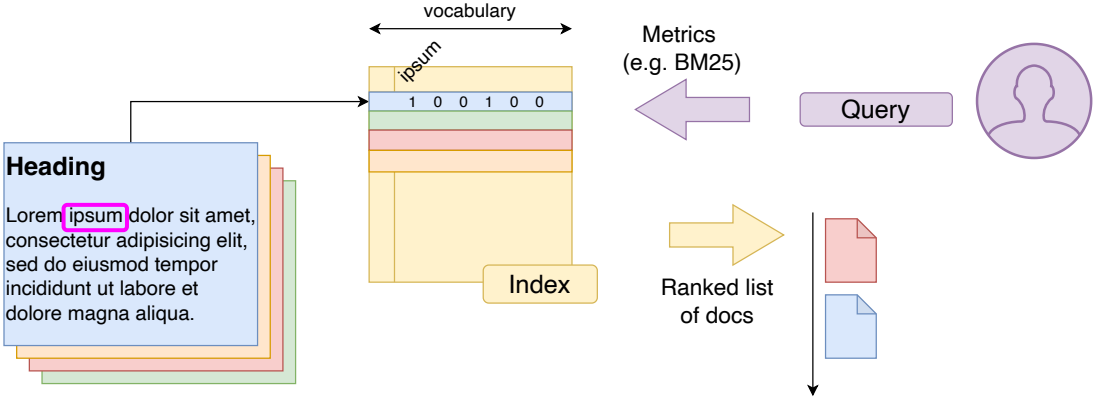
⇒ Integrate t =

- Can I t
- How m
- How go

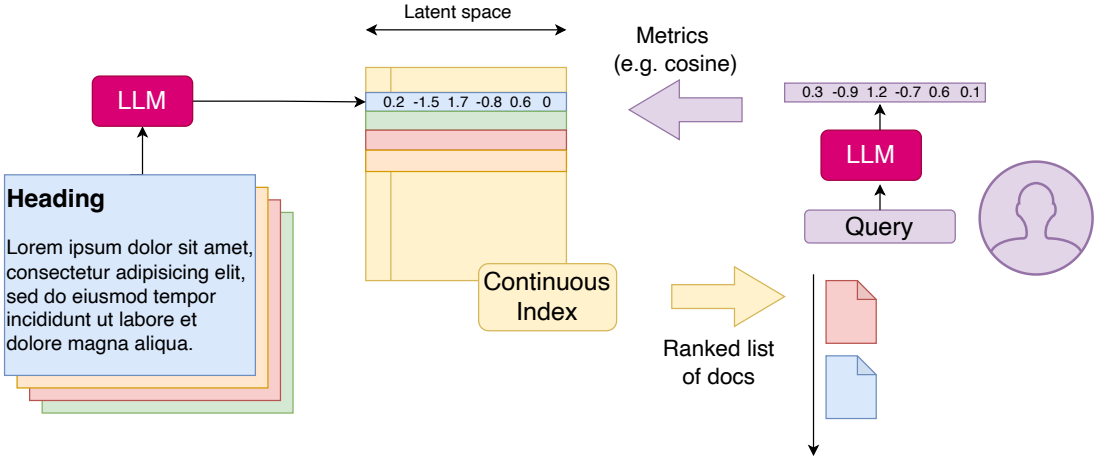


Module 3

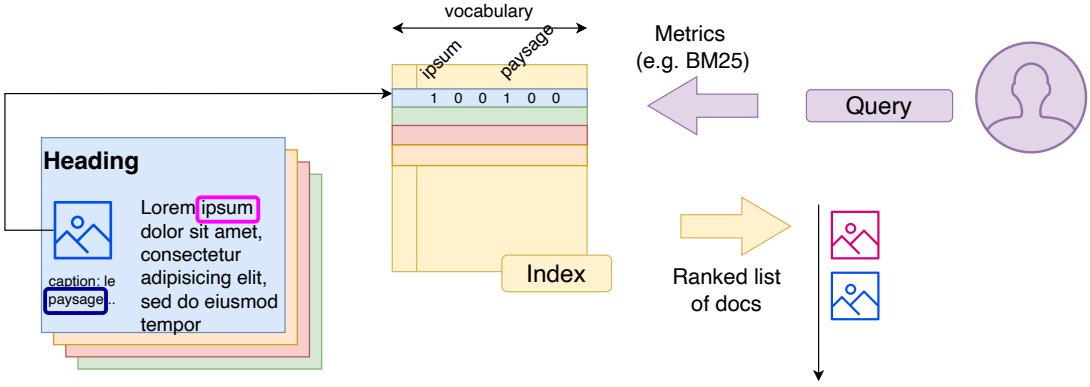
LLM *vs* Information Retrieval



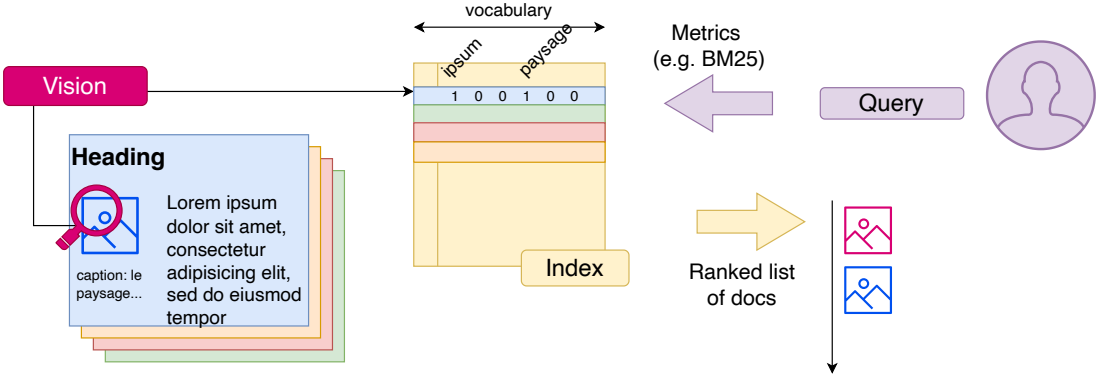
LLM *vs* Information Retrieval



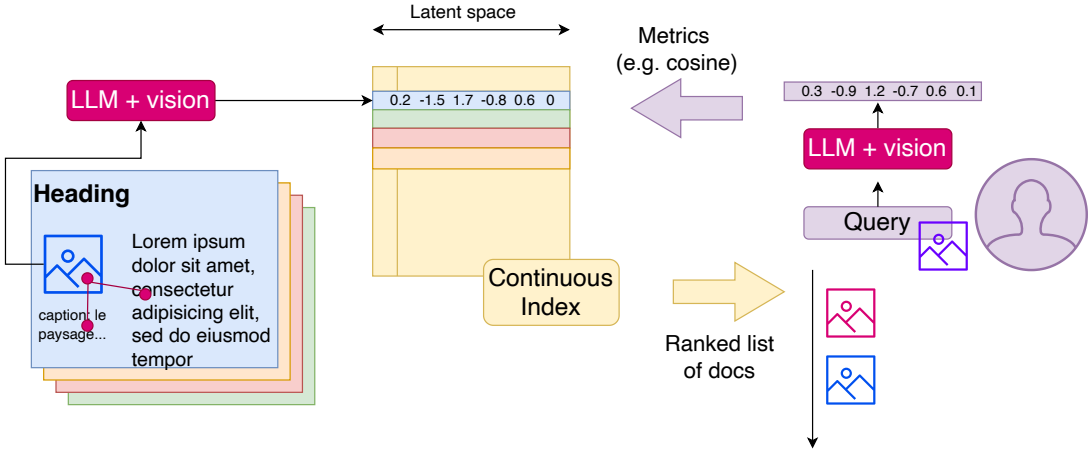
LLM *vs* Information Retrieval



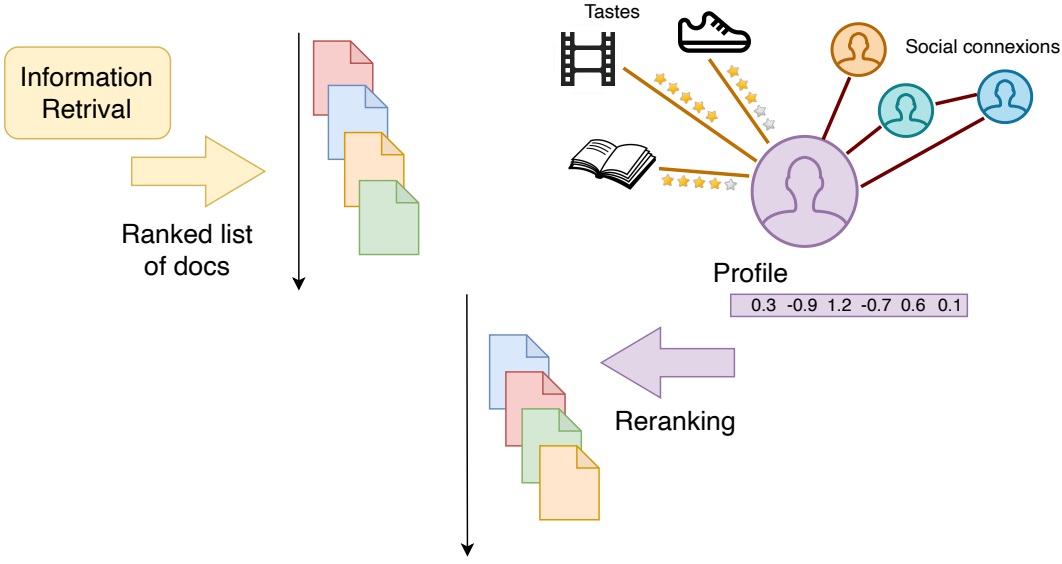
LLM *vs* Information Retrieval



LLM *vs* Information Retrieval

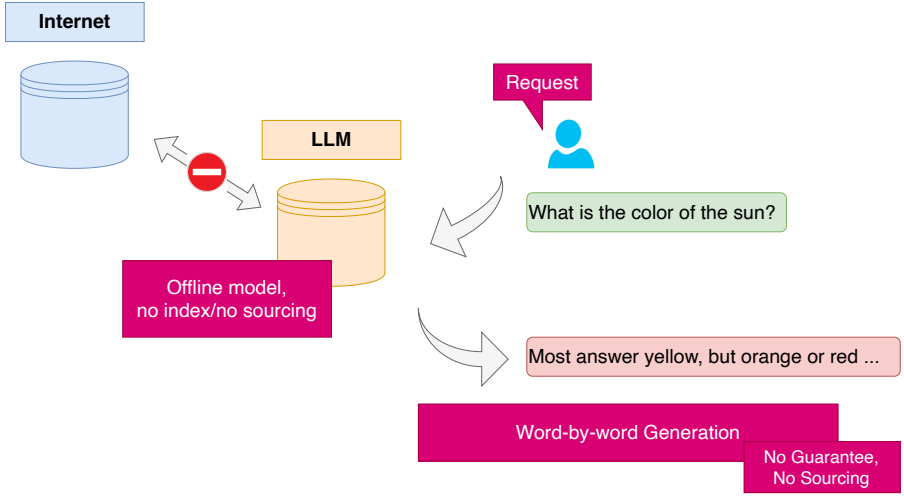


LLM *vs* Information Retrieval

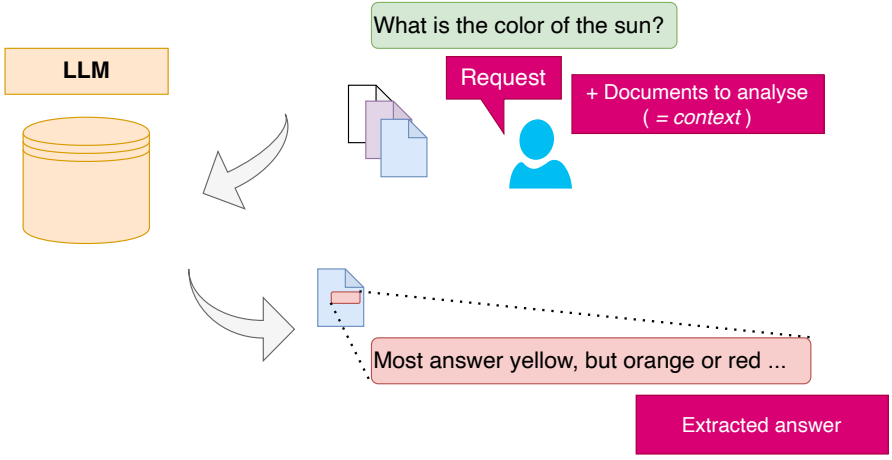


LLMs ⇒ RAG : parametric memory *vs* Info. Extraction

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

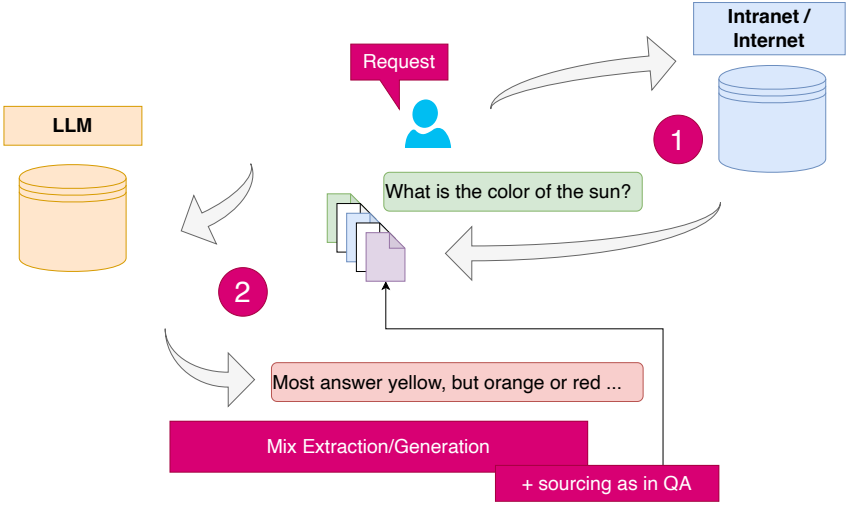


LLMs ⇒ RAG : parametric memory *vs* Info. Extraction



- Web query + analysis, automatic summary, rephrasing, meeting reports...
- (Current) limit on input size (2k then 32k tokens)
- = *pre chatGPT use of LLM for question answering*

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



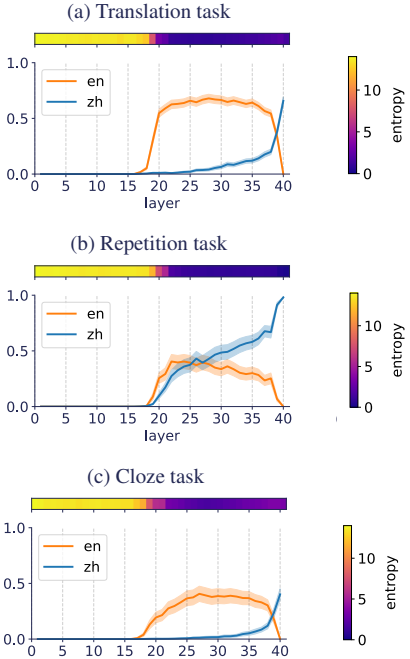
- RAG: Retrieval Augmented Generation
- (Current) limit on input size (2k then 32k tokens)

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



Large Language Models Uses

Two distinct questions

1. Teaching **with** AI
2. Teaching AI

⇒ Anyway, you have to know things about AI!

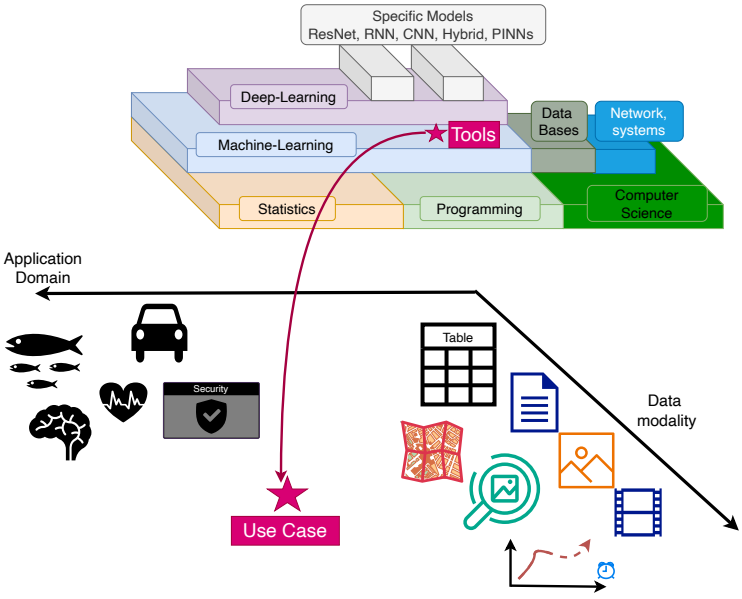
Teaching **of** AI

- 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

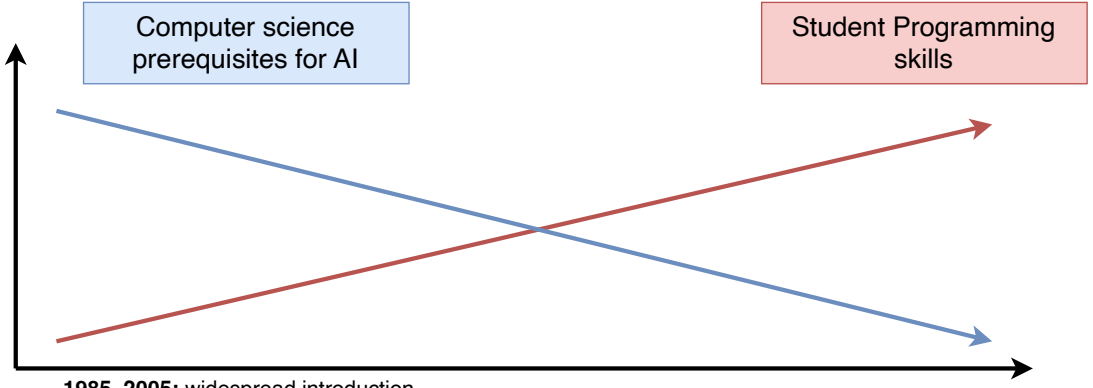
- New opportunities
- New risks
- New constraints

Teaching AI

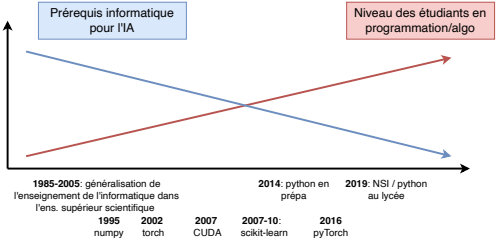


- Different levels of access (awareness, tool usage, development)
- Different data
- Different application domains

Access to AI: At a Crossroads



Access to AI: At a Crossroads



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.

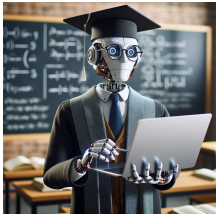
Ladies!

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?

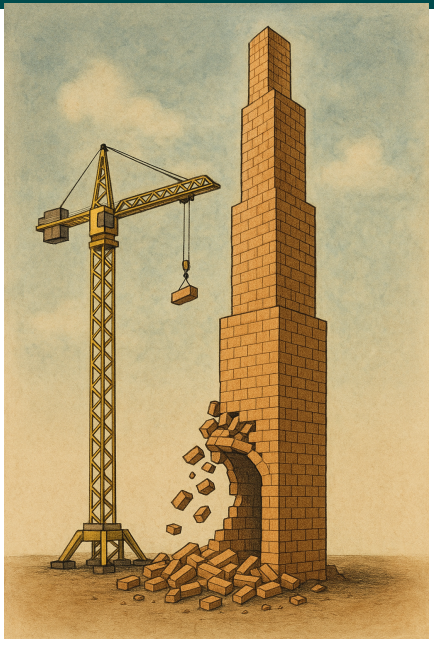
- Writing an essay
- Generate exam answers: code, history, foreign lang., ...
- Tackling a topic found on Wikipedia
- Producing a document analysis

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

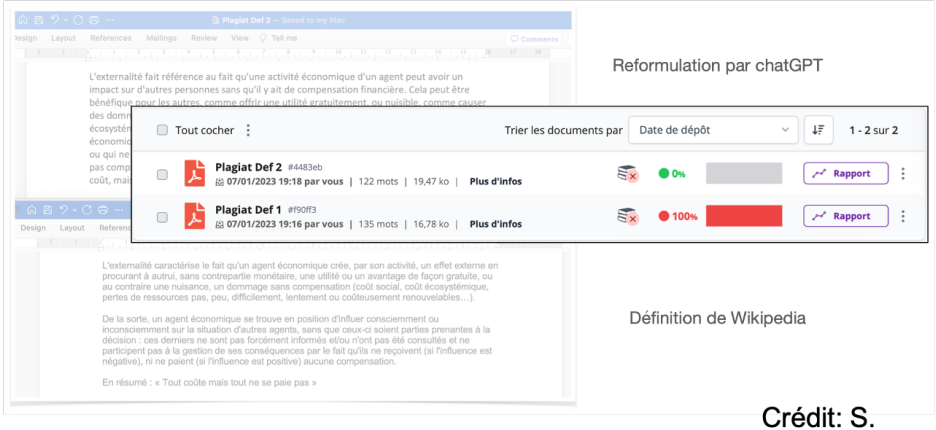
$\Rightarrow$ **LLM = core of the task $\Leftrightarrow$ Personal skills** $\searrow$

What Educational Challenges

- Redefine our educational priorities, subject by subject, as we did with Wikipedia/calculator/...
  - Accept the decline of certain skills if they are fully replaced by the LLM
- Train students in the use of LLMs, while managing to temporarily prohibit their use
  - Paper-based exams, project defenses with individual questions, ...
- Learn to recognize LLM-generated content, use detection tools.

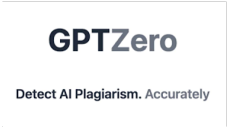


Detection of *chatGPT*-Generated Texts



Crédit: S. Pajak

Detection of *chatGPT*-Generated Texts



ORIGINALITY.AI

- Text **classifier** (as with any author)
  - Detection of biases in word choice / phrasing
- Characterization of text **likelihood** (OpenAI, GPTZero)
  - Overly fluent sentences, excessive logical connectors
  - Language model = statistical $\Rightarrow$ distribution comparison (**perplexity**)
- δ-**likelihood** on perturbed texts (DetectGPT)

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

$+$ confidence score in detection

$-$ depends on text length and edits made

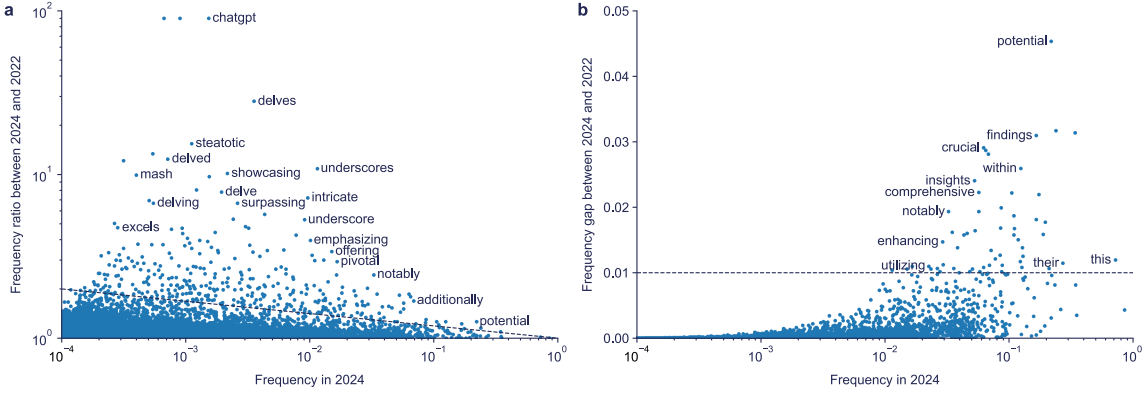
$-$ also detects translations made with LLMs

$\approx$ may flag Wikipedia fragments

What about Watermarking?

Detection of *chatGPT*-Generated Texts

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



L'IA à AgroParisTech: Quelques propositions structurantes

@AgroParisTech

1. Charte préliminaire des usages (assez épurée):

1. Utilisation responsable et éthique des outils d'IAG

- **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'IAG 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'IAG[1].
- **Respect de la vie privée**
  Ne soumettez pas des données sensibles ou personnelles aux outils d'IAG, 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 IAG. Restez conscients des impacts environnementaux et économiques liés à l'utilisation des outils d'IAG.

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

2. Réglement des études:
   - Possibilité de demander des oraux/rattrapages au moindre doute sur l'authenticité d'un devoir
   - Sanction importante : usage inaproprié de l'IA = fraude

@AgroParisTech

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'IAG, 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'IAG. 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

2. Réglement des études:
   - Possibilité de demander des oraux/rattrapages au moindre doute sur l'authenticité d'un devoir
   - Sanction importante : usage inaproprié de l'IA = fraude

@AgroParisTech

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`

2. Réglement des études:
   - Possibilité de demander des oraux/rattrapages au moindre doute sur l'authenticité d'un devoir
   - Sanction importante : usage inaproprié de l'IA = fraude

@AgroParisTech

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

2. Réglement des études:
   - Possibilité de demander des oraux/rattrapages au moindre doute sur l'authenticité d'un devoir
   - Sanction importante : usage inaproprié de l'IA = fraude

La proposition de l'éducation nationale

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 (générative)

- Module léger AgroParisTech / Hercule 4.0

  [en cours de déploiement sur ecampus]
- Module avancé BrevetIA / Paris-Saclay
  - Formats 10h/20h : IA + IA générative

$\Rightarrow$ 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

(Main) Risks derived from ML & LLM

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

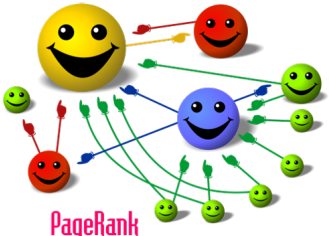
Harms that arise from users overly trusting the language model, or treating it as human-like.

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 digital oblivion
- 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

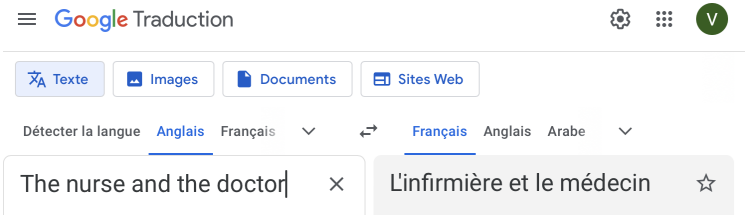
Bias in the data ⇒ bias in the responses

Machine learning is based on extracting statistical biases...

⇒ Fighting bias = manually adjusting the algorithm

Machine Learning & Bias

Sterreotypes from *Pleated Jeans*



- 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?

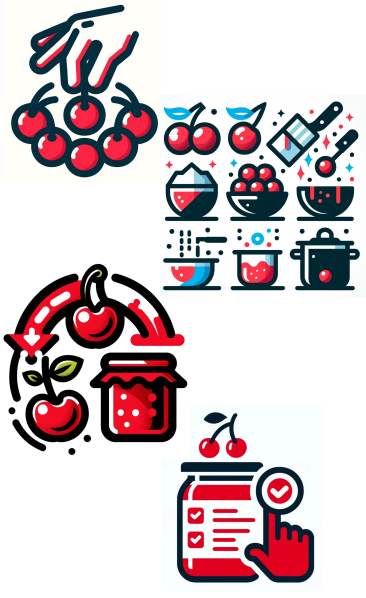
- Domain experts / specifications
- Engineers, during algorithm design
- Ethics group, during result validation
- Communication group / user response

⇒ What legitimacy? What transparency? What effectiveness?

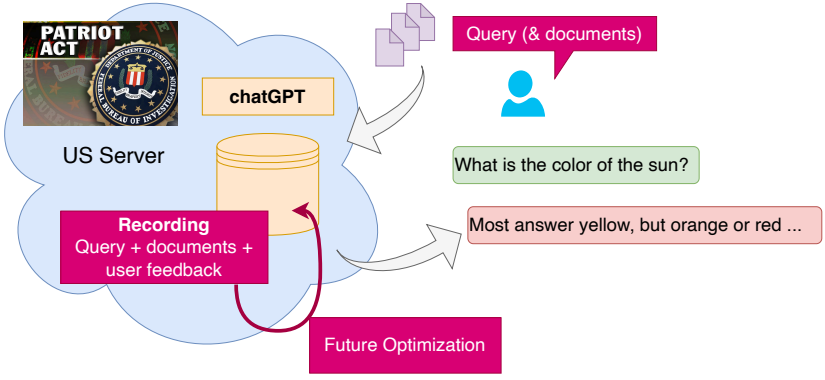
Machine learning is never neutral

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

$\Rightarrow$ Choices that influence algorithm results



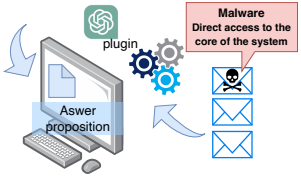
Data Leak(s)

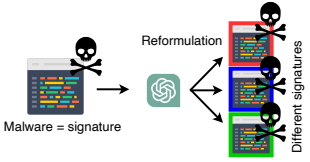


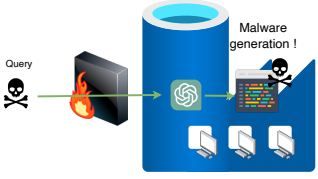
- Transfer of sensitive data
- Exploitation of data by OpenAI (or others)
- Data leakage in future models

Security Issues

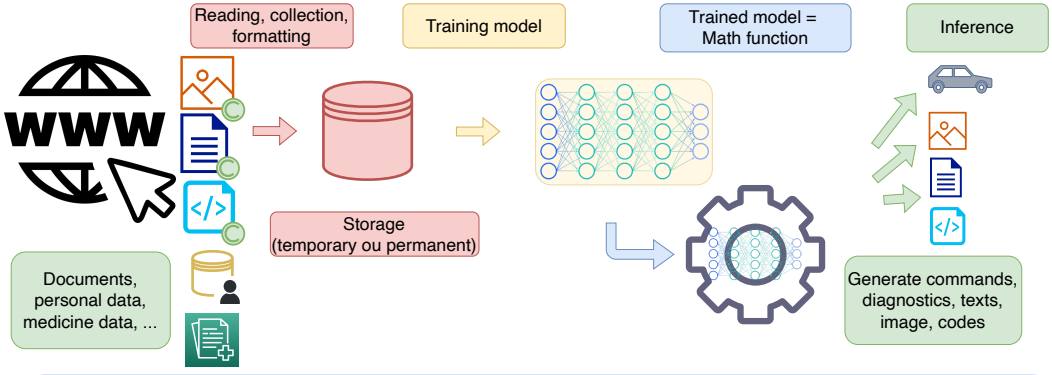
- Plug-ins ⇒ Often significant security vulnerabilities for users
  - Email access / transfer of sensitive information etc...
- Management issues for companies
  - Securing (very) large files
- Increased opportunities for malware signatures
  - ≈ software rephrasing
- New problems!
  - Direct malware generation







Legal Risks/Questions



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

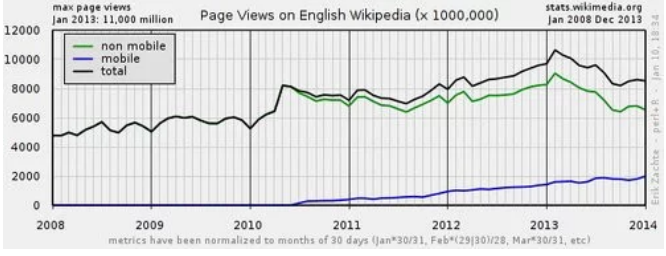
Economic Questions

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

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

Google's Knowledge Graph Boxes: killing Wikipedia?

by Gregory Kohs

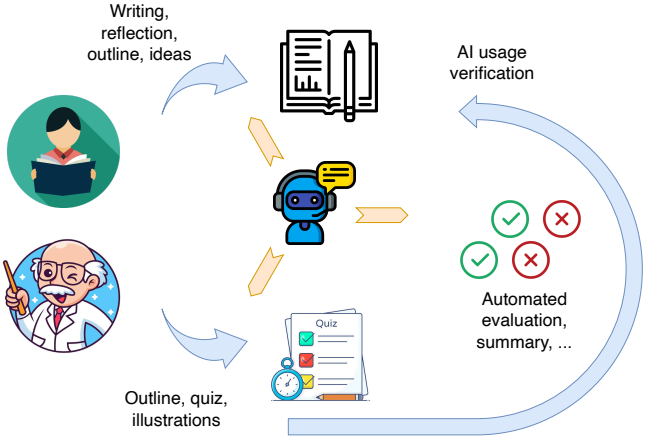


$\Rightarrow$ Who does **benefit from the feedback**? [StackOverFlow]

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. **Justice:** 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:** including the environment?
3. **Non-maleficence:** Humans + environment / sustainability / malicious uses
4. **Justice:** 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:**

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. **Justice:** 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:** including the environment?
3. **Non-maleficence:** Humans + environment / sustainability / malicious uses
4. **Justice:** 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:**

Conclusion

Tools and Questions

New tools:

- New ways to handle existing problems
- Address new problems
- ... But obviously, it doesn't always work!
- AI often makes mistakes (assistant *vs* replacement)

Learning to use an AI system

- AI not suited for many problems
- AI = part of the problem (+interface, usage, acceptance...)

Maturity of Tools & Environments

(More) mature tools

- **Environments**: Jupyter, Visual Studio Code, ...
- **Machine Learning** `Scikit-Learn`: blocks to assemble
  - Training: 1 week
  - Project completion: few hours to few days
- **Deep Learning** `pytorch, tensorflow`: building blocks... but more complex
  - Training: 2-5 weeks
  - Project completion: few days to few months
  - Mandatory for text and image

- A data project = 10 or 100 times less time / 2005
- Developing a project is **accessible to non-computer scientists**

Levels of Access to Artificial Intelligence

1. User via an interface: *chatGPT*
   - WARNING: 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

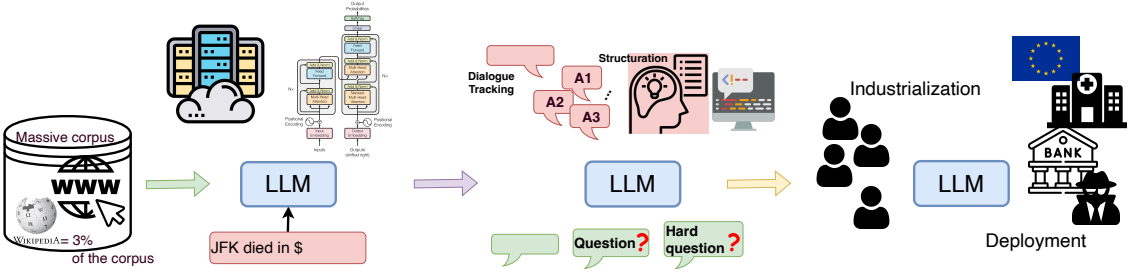
Training
- Computation power (x1000 GPU)
- Architecture ML

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

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

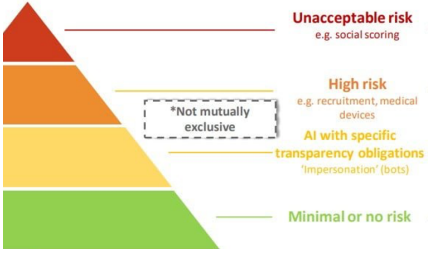
+ **DPO** :
Data Protection Officer

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)



chatGPT: A Simple Step

- Training & Tuning Costs

  4-5 Million Euros / training $\Rightarrow$ chatGPT is **poorly trained**!

- Data Efficiency

  chatGPT $>$ 1000x a human's lifetime reading

- Identify Entities, Cite Sources

  Anchoring responses in knowledge bases
  Anchoring responses in sources

- Multiplication of initiatives: GPT, LaMBDA, PaLM, BARD, BLOOM, Gopher, Megatron, OPT, Ernie, Galactica...
- Public involvement,
  impact on information access