

DE L'INTELLIGENCE ARTIFICIELLE AUX MODÈLES DE FONDATION

Jeudi 5 décembre 2024
GDR RADIA

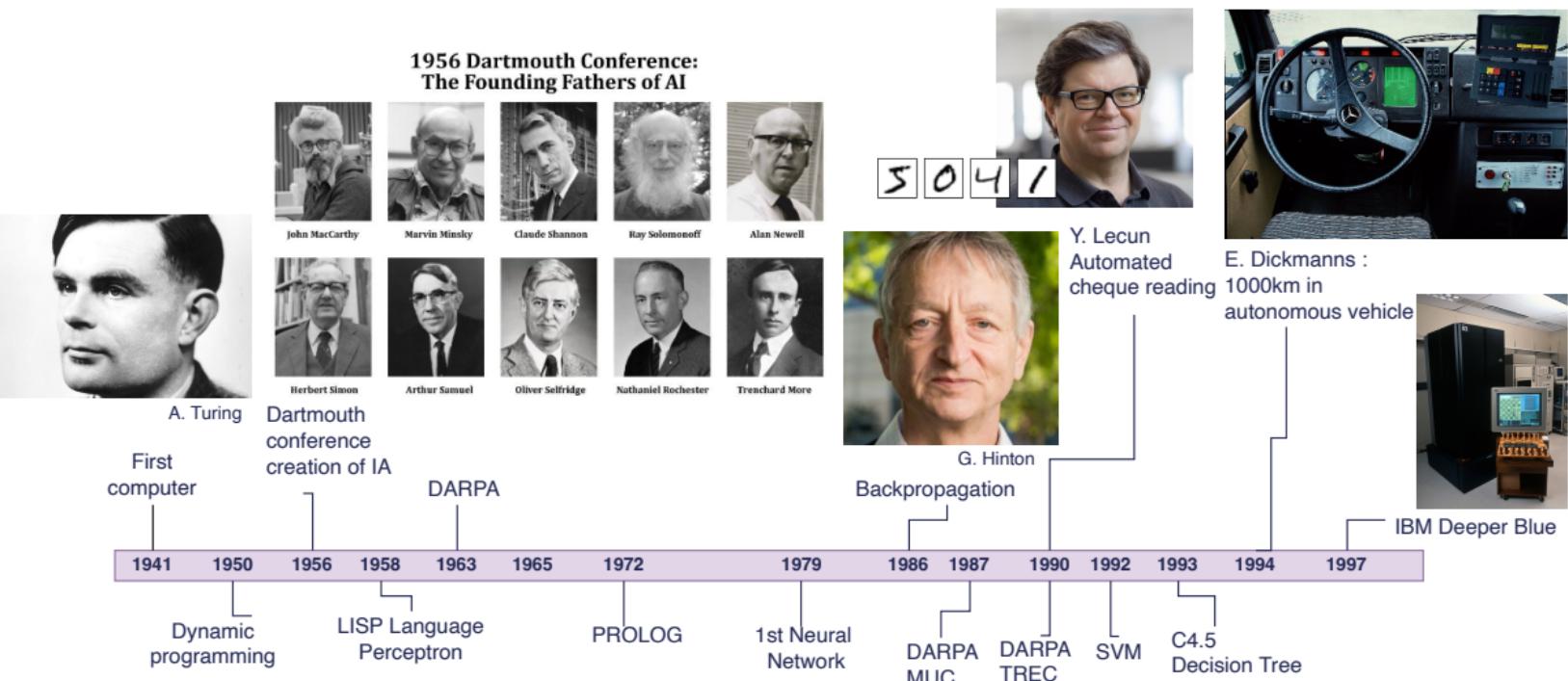
Vincent Guigue
vincent.guigue@agroparistech.fr
<https://vguigue.github.io>

FROM AI TO MACHINE-LEARNING



Historique rapide de l'IA

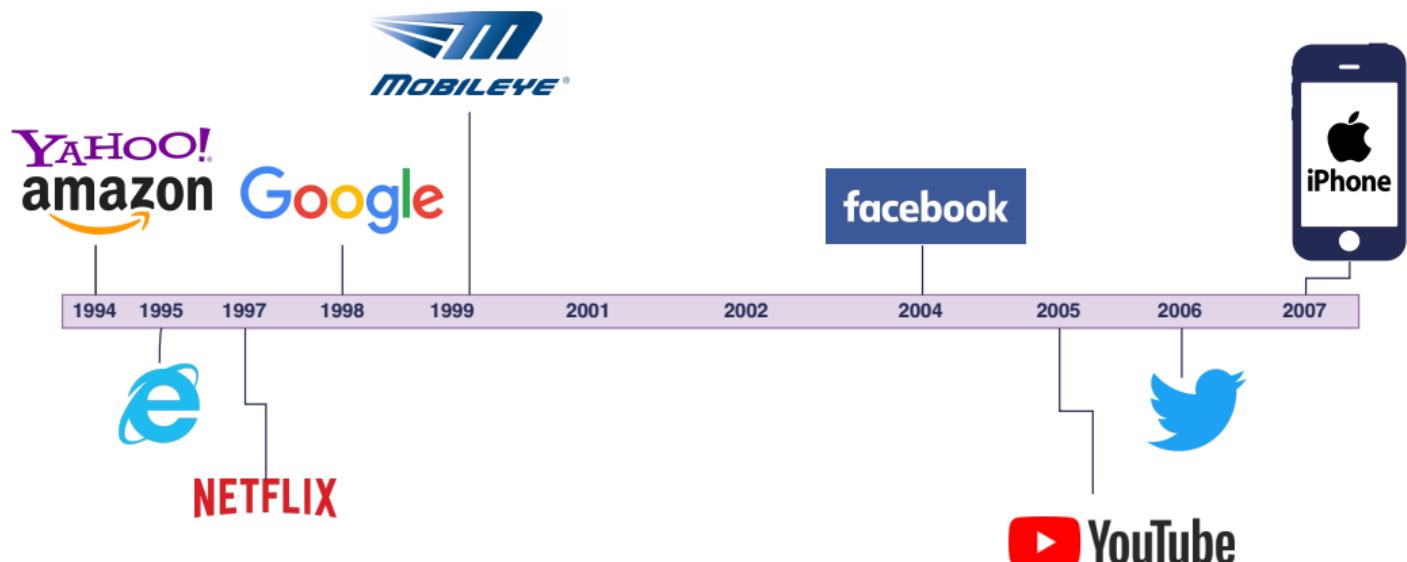
Naissance de l'informatique... Et de l'Intelligence Artificielle





Historique rapide de l'IA

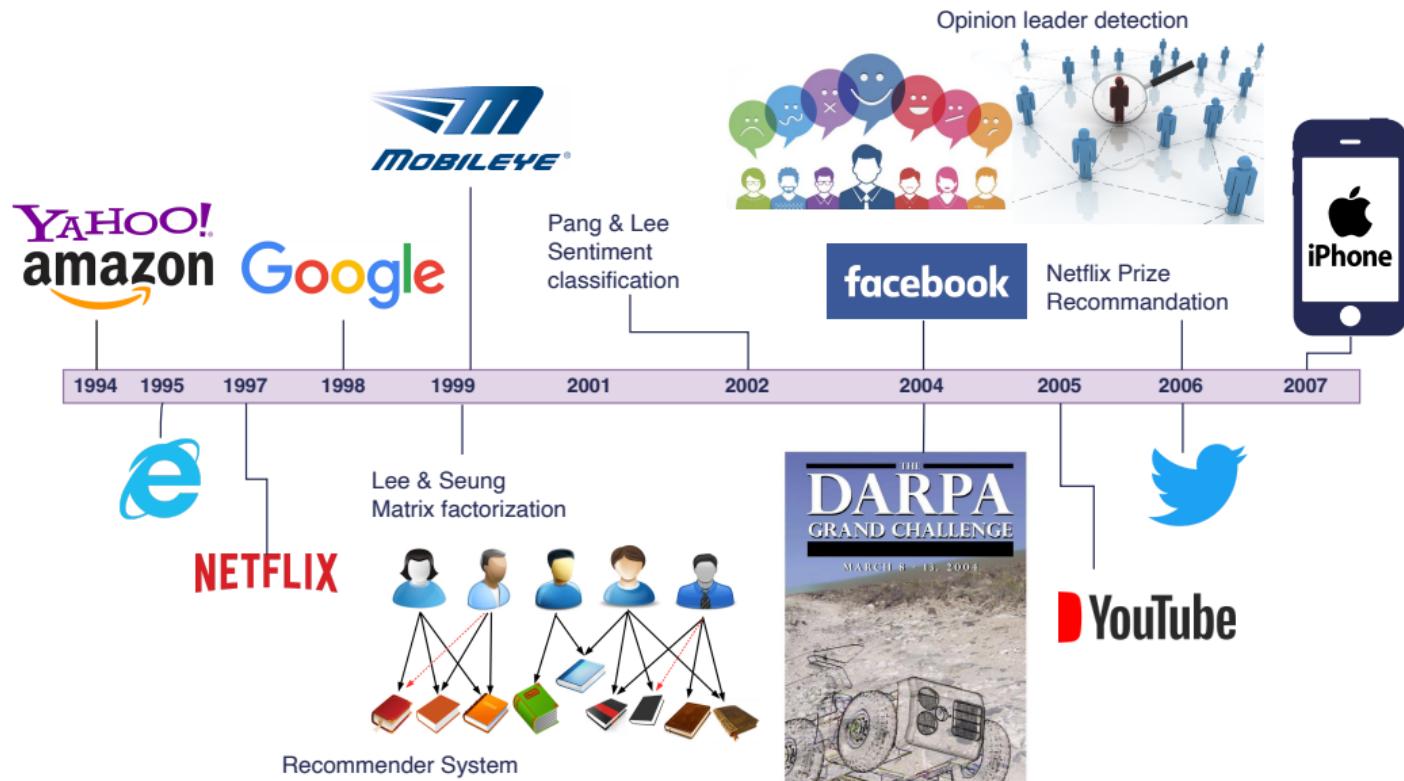
Emergence (ou refondation) des GAFAM/GAMMA





Historique rapide de l'IA

Emergence (ou refondation) des GAFAM/GAMMA



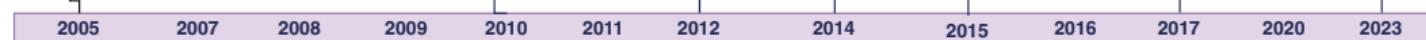


Historique rapide de l'IA

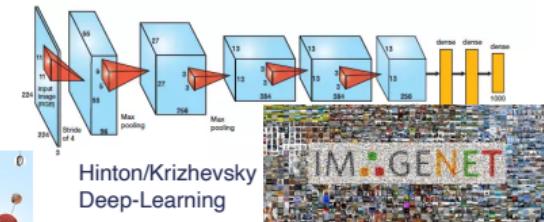
Formation d'une vague de l'Intelligence Artificielle



Sebastian Thrun:
DARPA Grand Challenge
victory



IBM Jeopardy win



Hinton/Krizhevsky Deep-Learning

K. Cho Traduction auto.
Translate (v2)

amazon alexa

Google DeepMind
Acquisition : \$400M



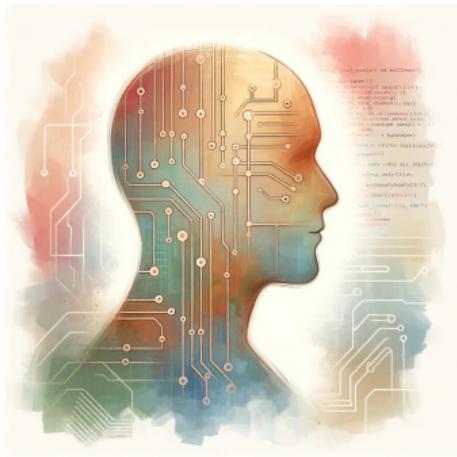
An Intel company
Acquisition : \$15B



OpenAI
DALL·E 2



Intelligence Artificielle & Machine Learning



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

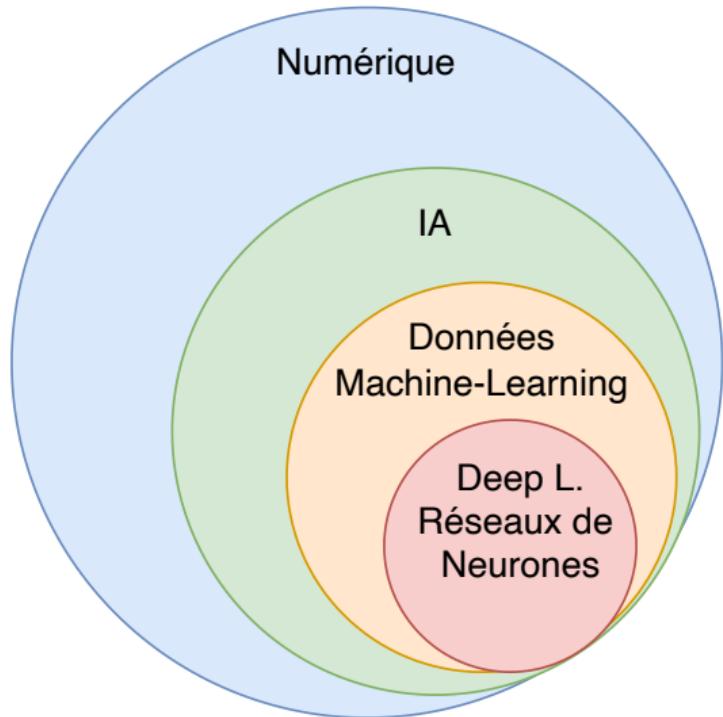
IA : programmes informatiques qui s'adonnent à des tâches qui sont, pour l'instant, accomplies de façon plus satisfaisante par des êtres humains car elles demandent des processus mentaux de haut niveau.

Marvin Lee Minsky, 1956

N-AI (Narrow Artificial Intelligence), dédiée à une tâche
≠ G-AI (General AI) qui remplace l'humain dans des systèmes complexes.
Andrew Ng, 2015

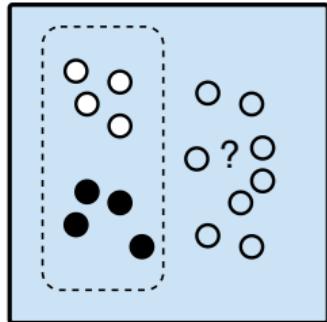


Place de l'IA dans le numérique

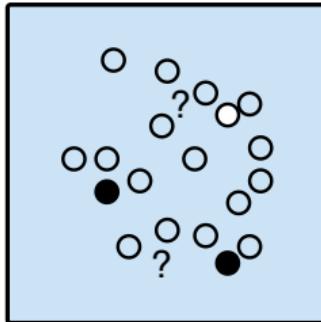


- Caisse automatique du supermarché
- Google Maps
- Système prédictif (e.g. marché immobilier), recommandation
- chatGPT

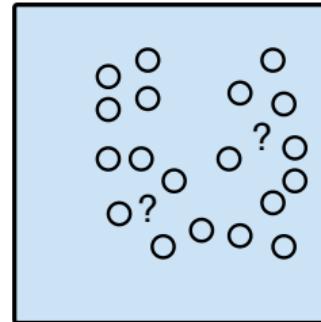
Cadres en machine-learning



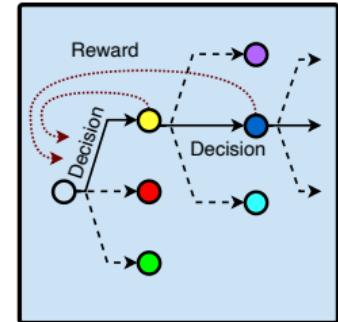
Apprentissage supervisé



Apprentissage semi-supervisé



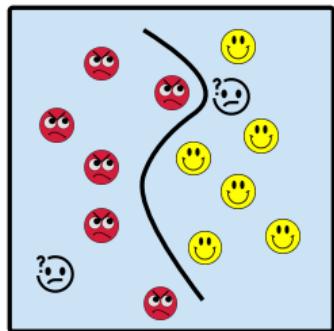
Apprentissage non-supervisé



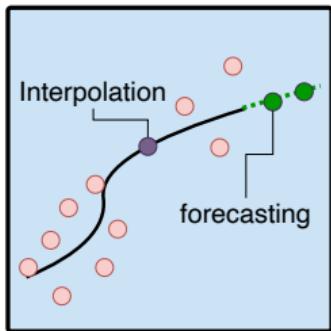
Apprentissage par renforcement

- Différentes **modalités** de données (images, textes, données numériques...)
- Différents **étiquetages**

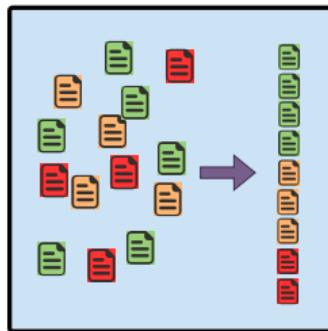
Cadres en machine-learning



Classification



Regression

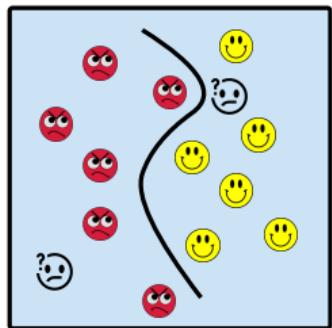


Ranking

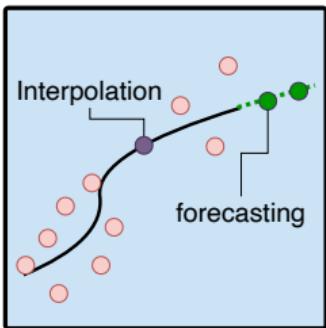
- Différentes **modalités** de données (images, textes, données numériques...)
- Différents **étiquetages**
- Différentes types de **prédictions**



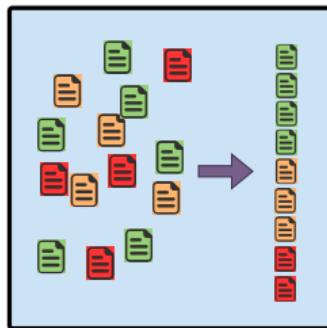
Cadres en machine-learning



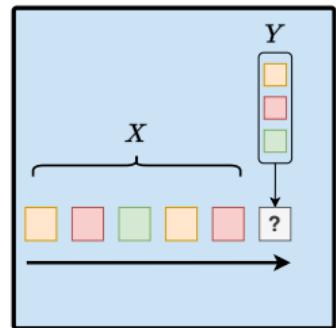
Classification



Regression



Ranking

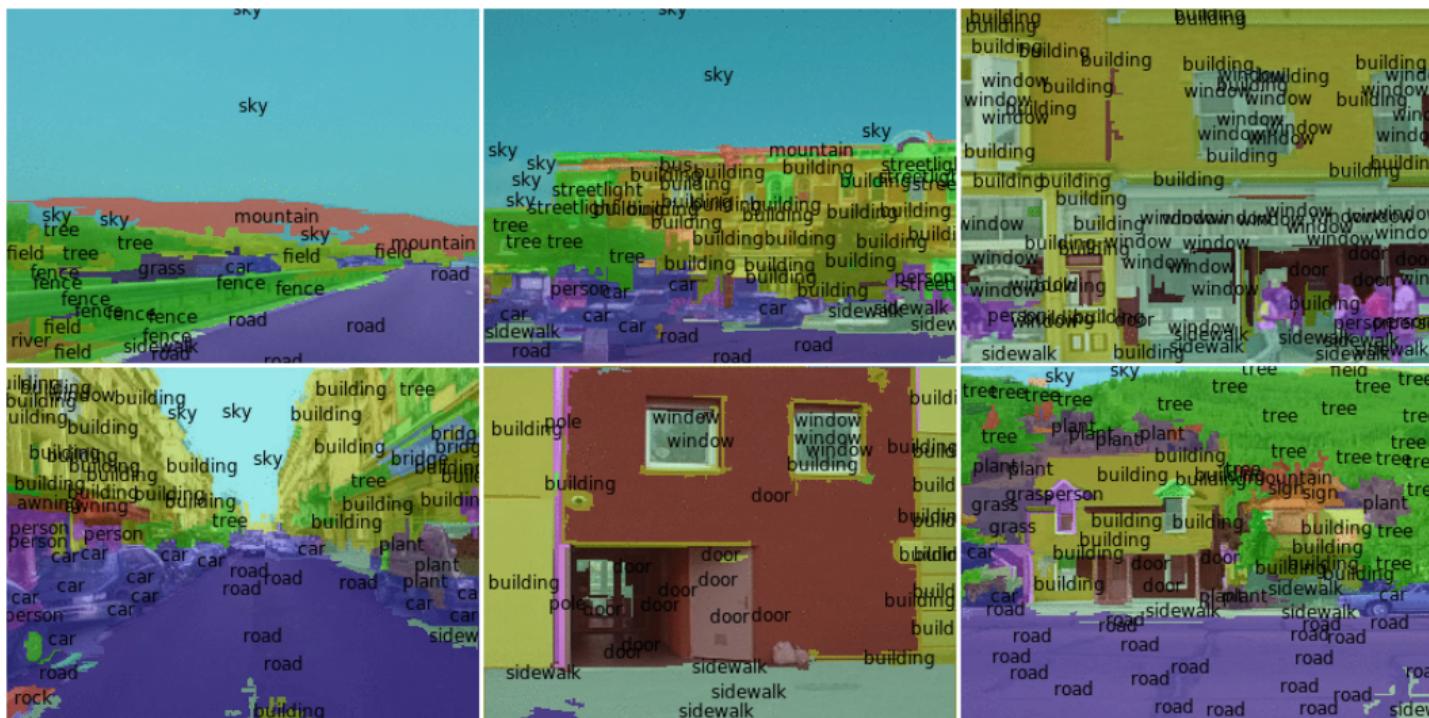


Generative AI

- Différentes **modalités** de données (images, textes, données numériques...)
- Différents **étiquetages**
- Différentes types de **prédictions**

Exemples de problèmes

Imaginer des solutions pour:



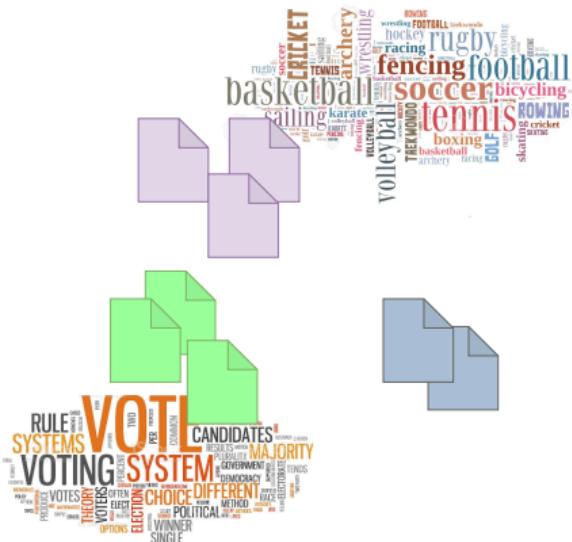
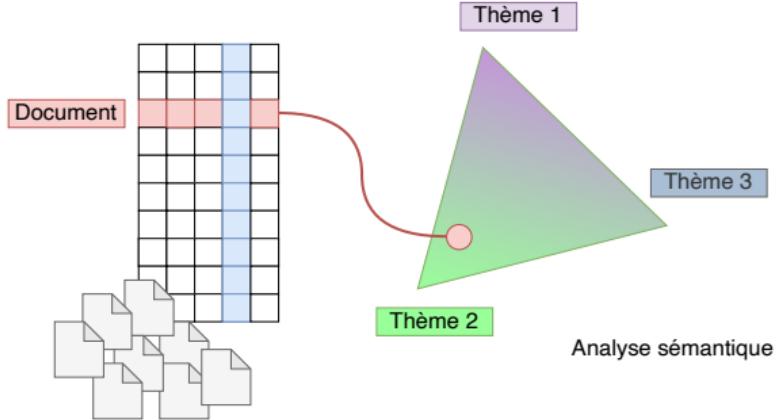
Exemples de problèmes

Imaginer des solutions pour:



Exemples de problèmes

Imaginer des solutions pour:

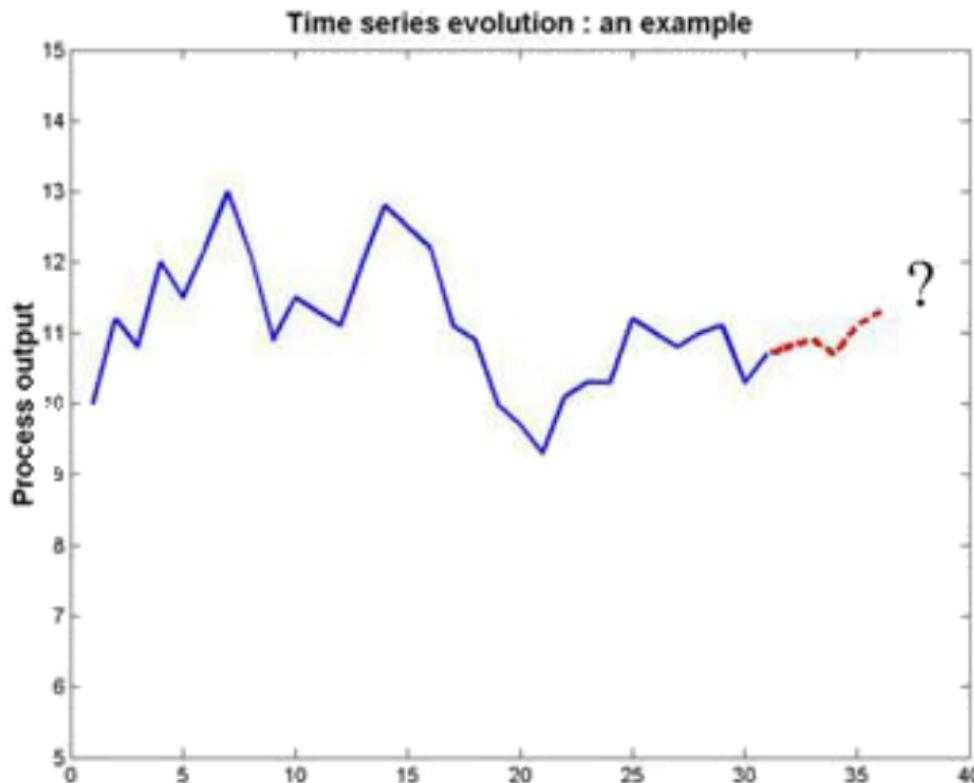


Analyse sémantique latente: LSA, pLSA, LDA



Exemples de problèmes

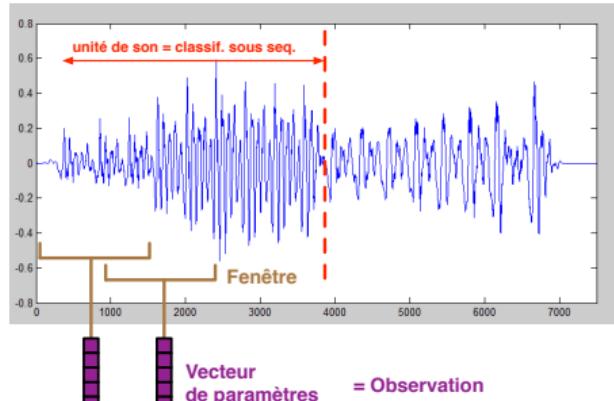
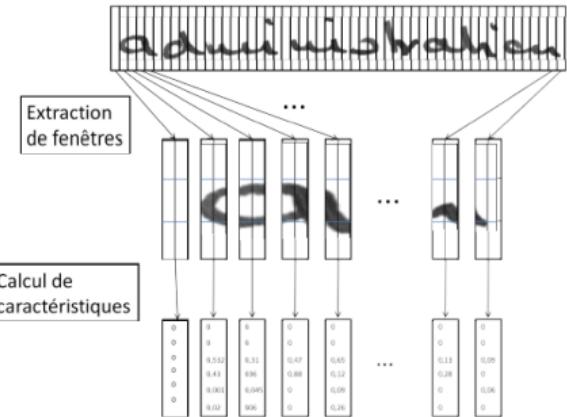
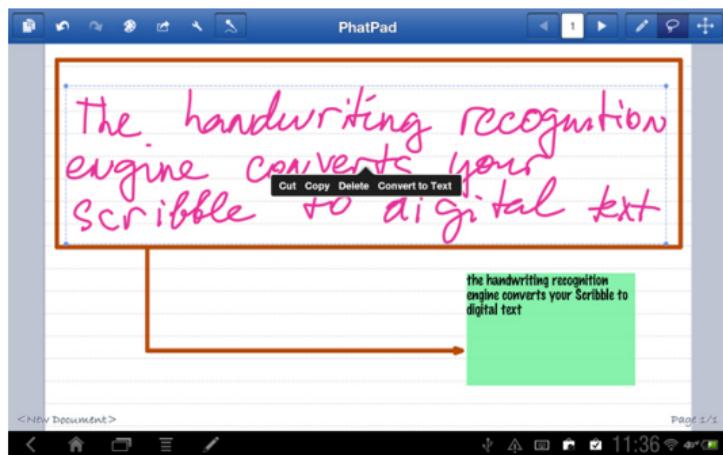
Imaginer des solutions pour:



A

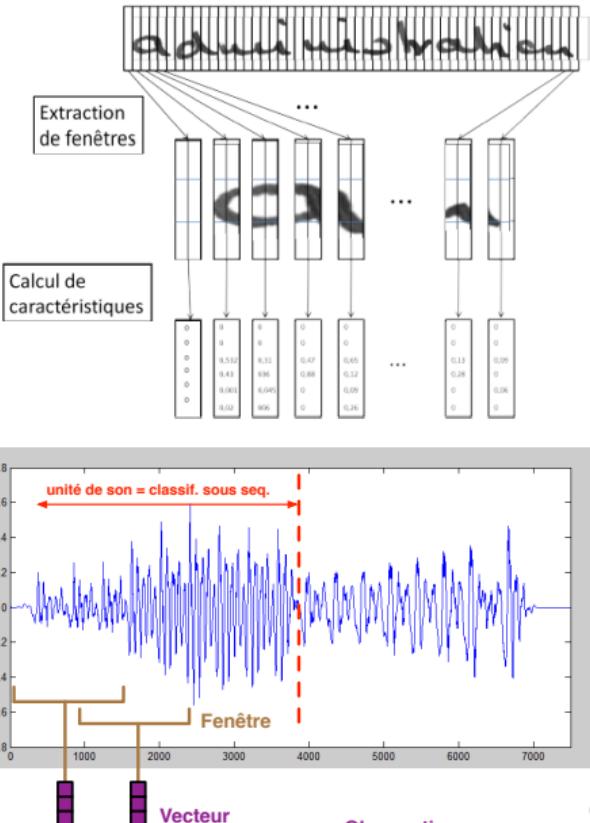
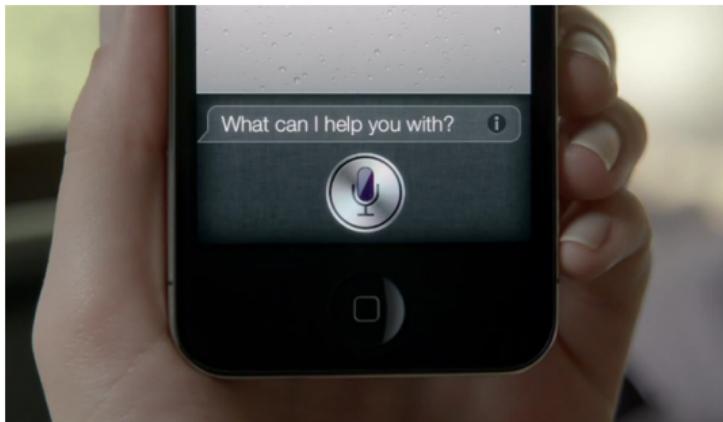
Exemples de problèmes

Imaginer des solutions pour:



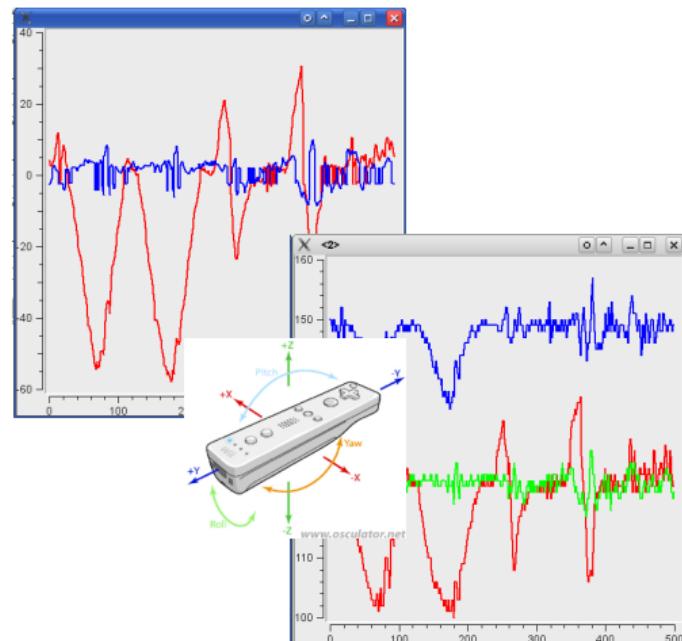
Exemples de problèmes

Imaginer des solutions pour:



Exemples de problèmes

Imaginer des solutions pour:



Exemples de problèmes

Imaginer des solutions pour:



Exemples de problèmes

Imaginer des solutions pour:

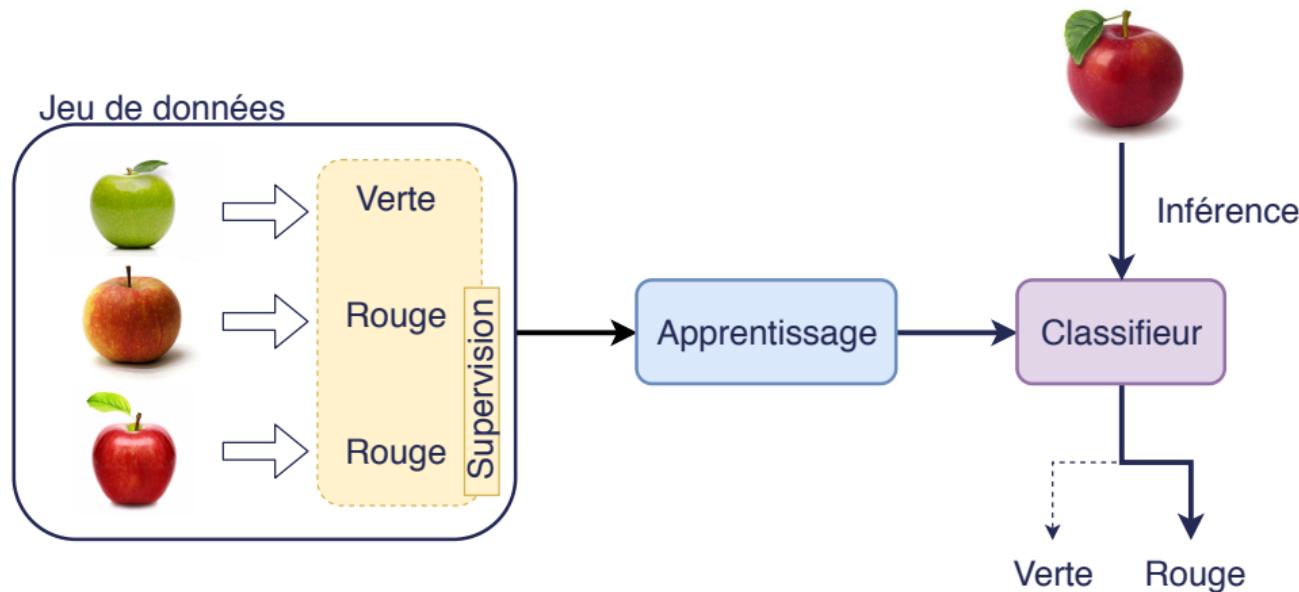


Exemples de problèmes

Imaginer des solutions pour:

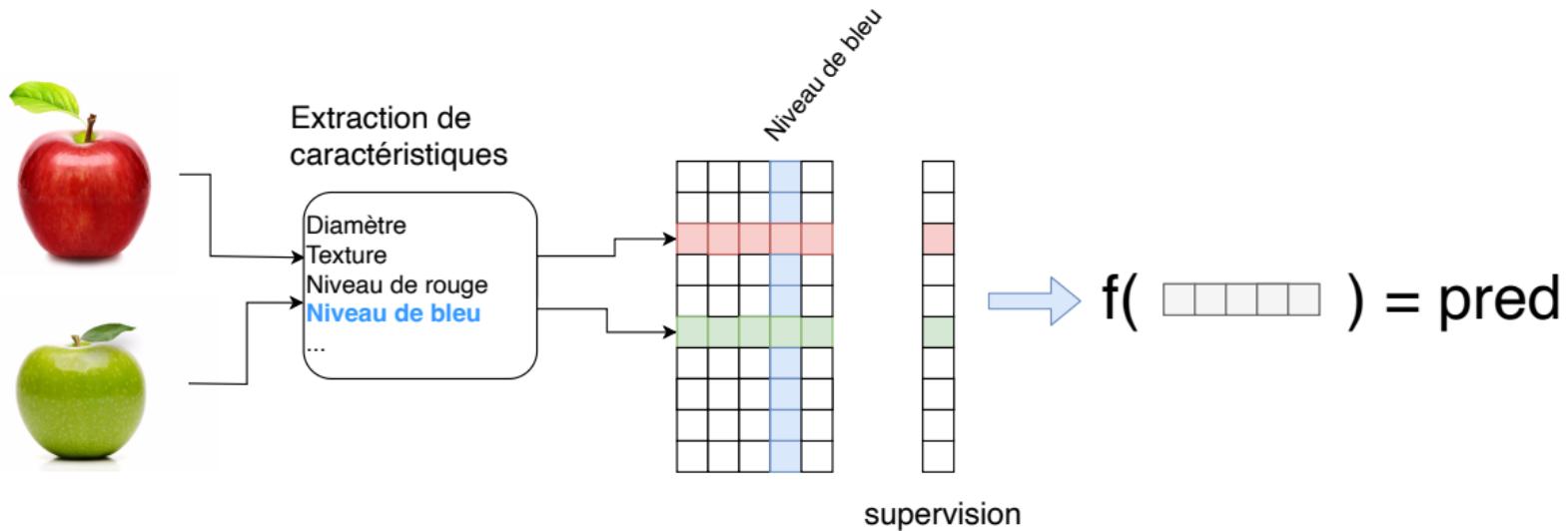


Chaine de traitements supervisée & modèles

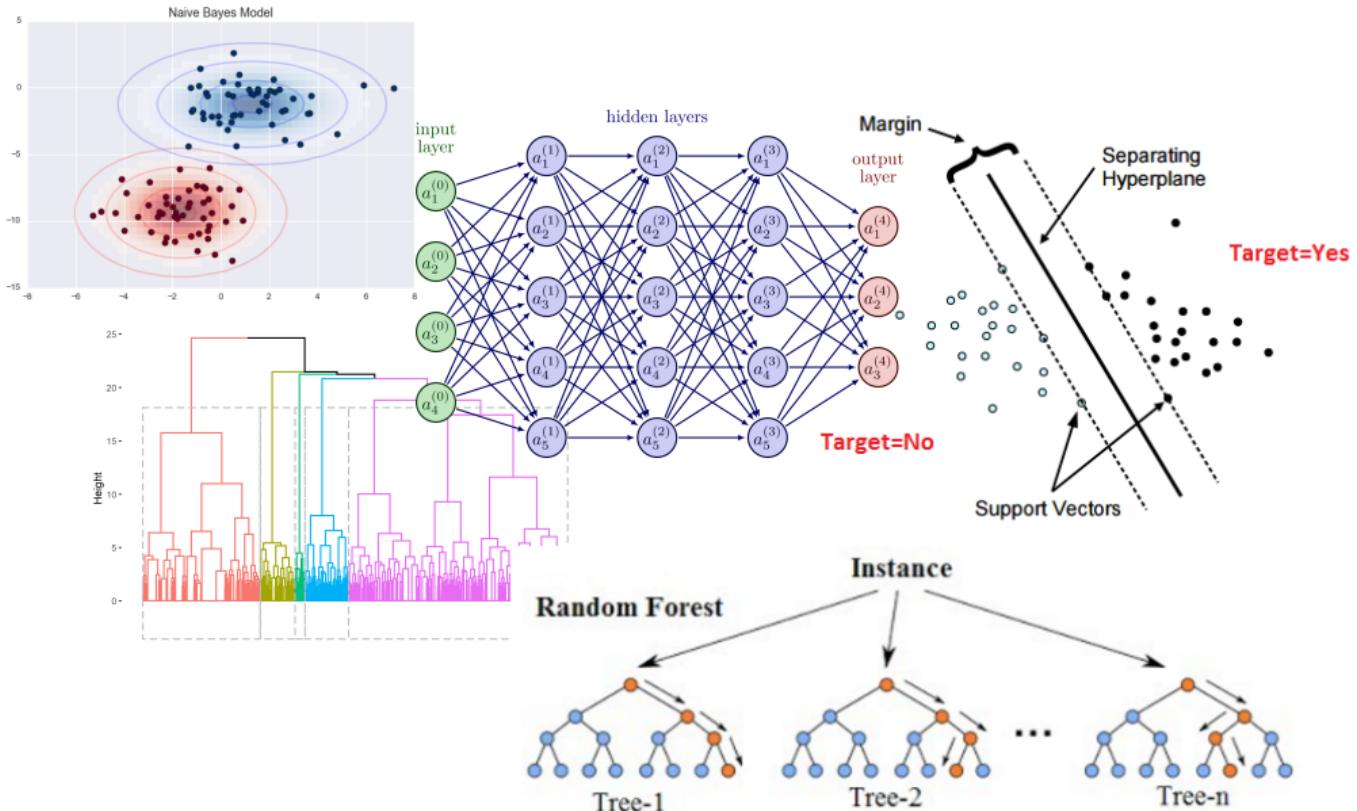


- Promesse = construire un modèle *uniquement* à partir des observations

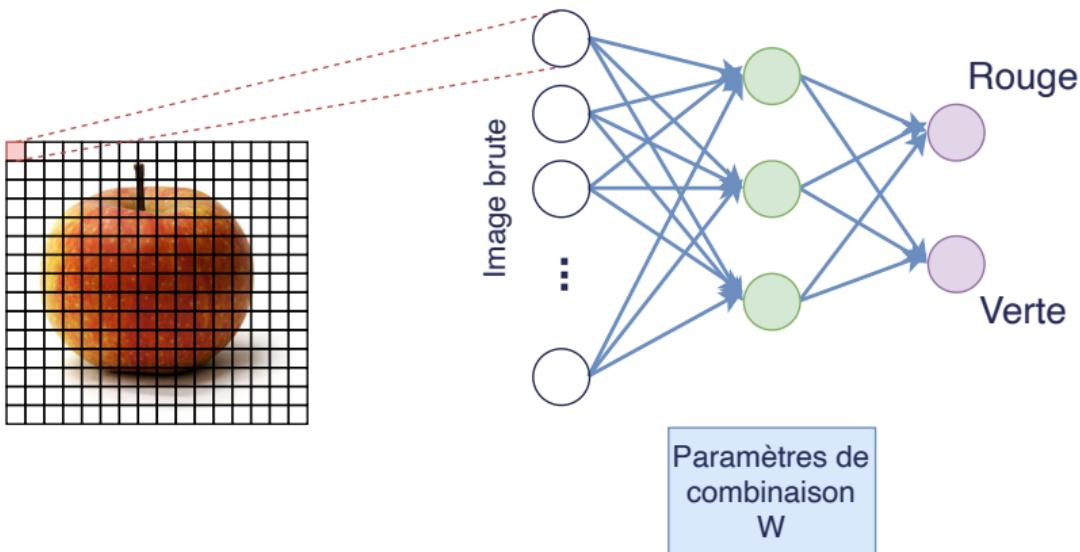
Chaine de traitements supervisée & modèles



Chaine de traitements supervisée & modèles



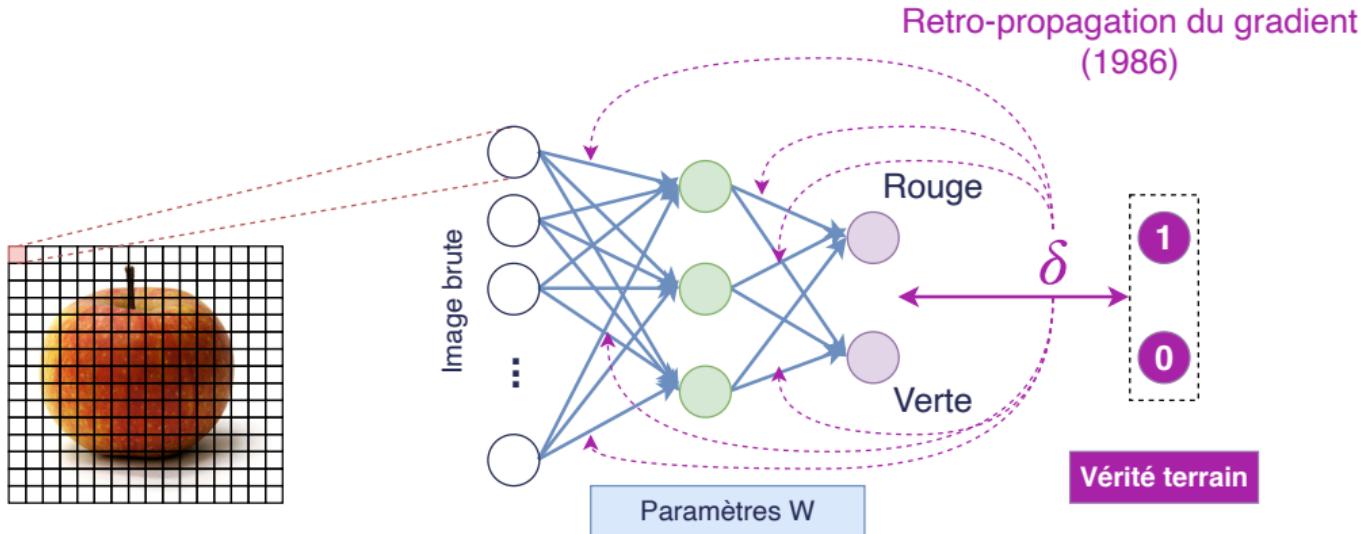
Chaine de traitements supervisée & modèles



■ Initialisation aléatoire...

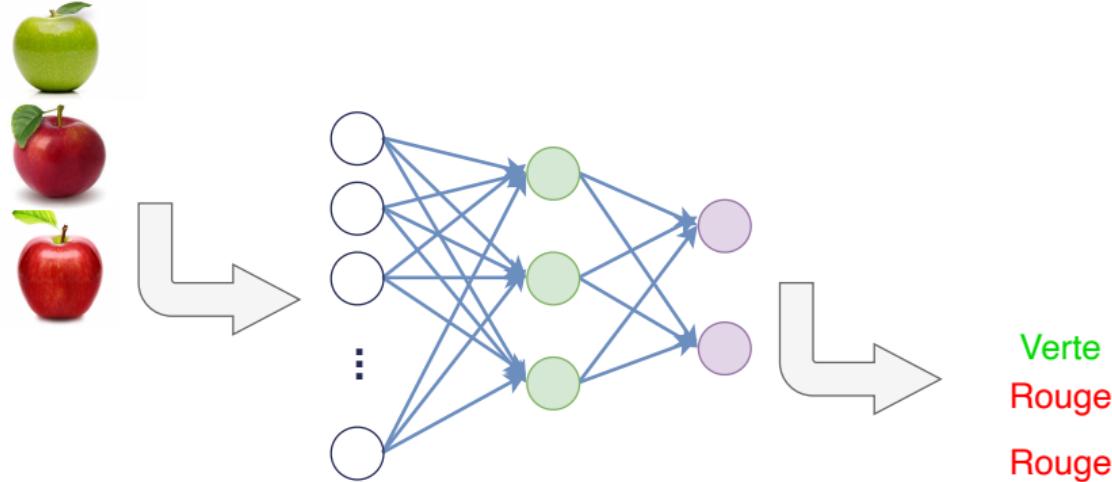
Et décision aléatoire (au début!)

Chaine de traitements supervisée & modèles



- Mise à jour des poids
- Pas à pas epsiloniques, nombreuses itérations sur les données

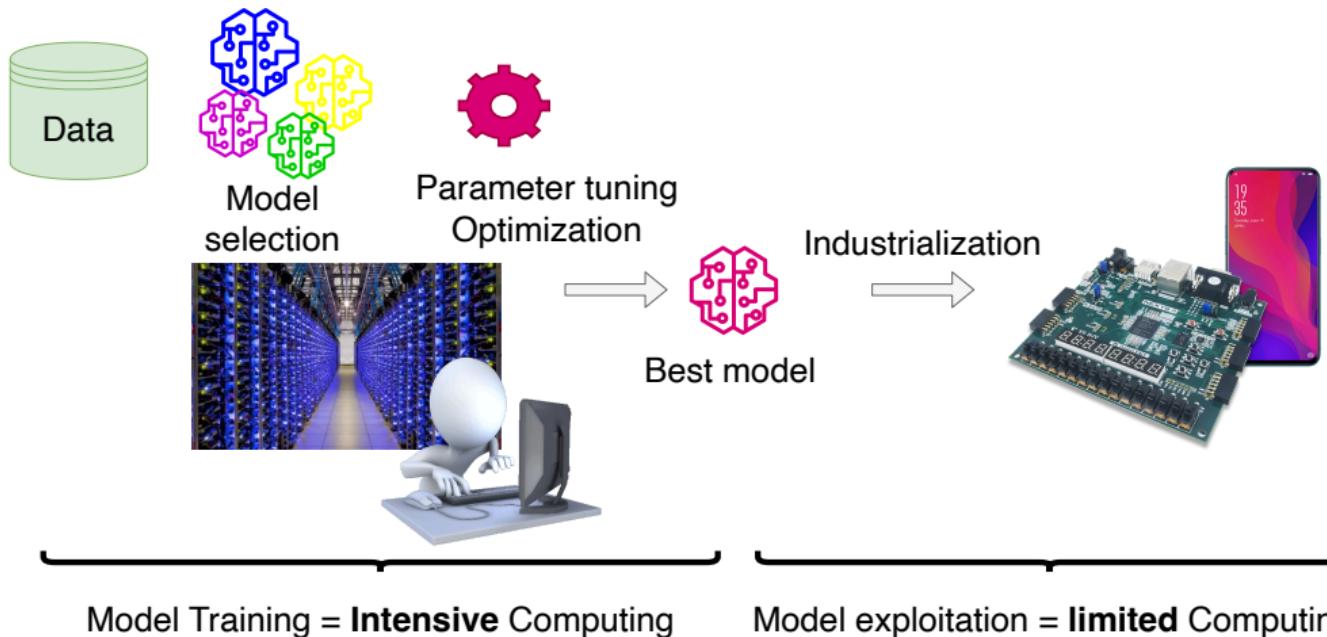
Chaine de traitements supervisée & modèles



- Apprentissage lent et couteux
- Inférence (beaucoup plus) rapide

Chaine de traitements supervisée & modèles

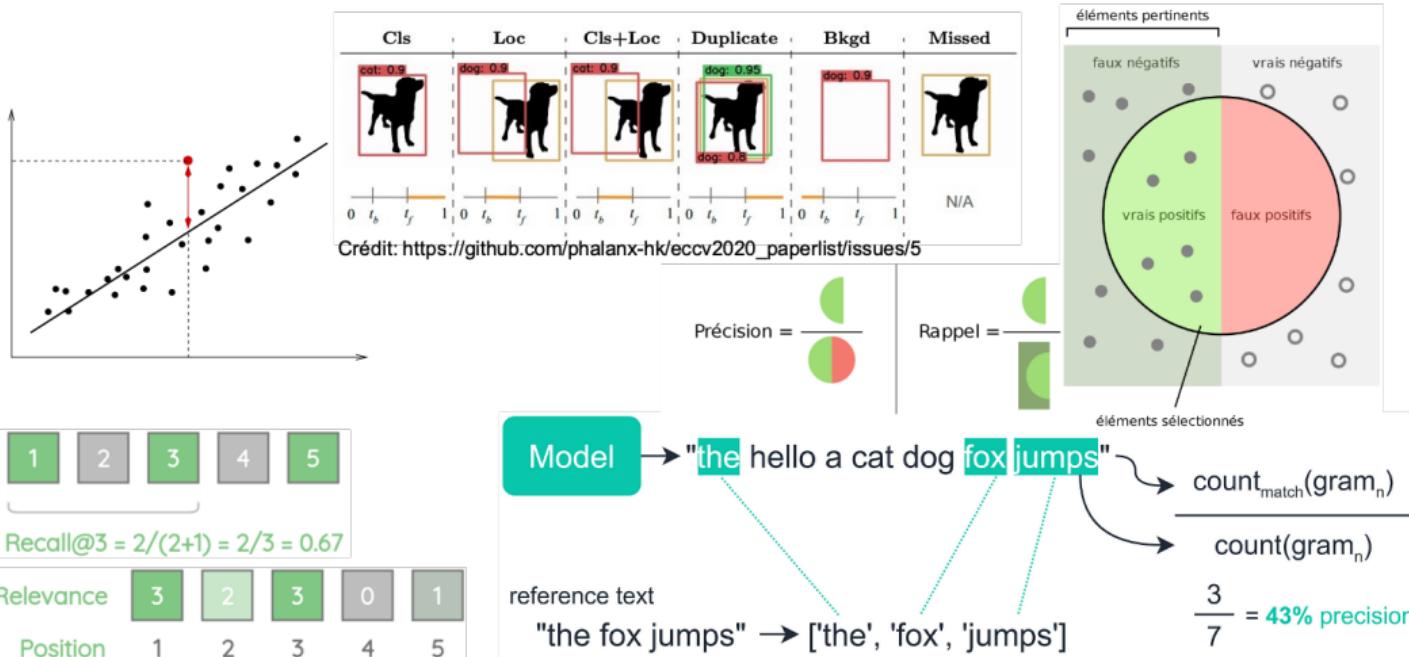
Différentes étapes en machine-learning



Mesurer les performances

Estimer les performances (en généralisation)...

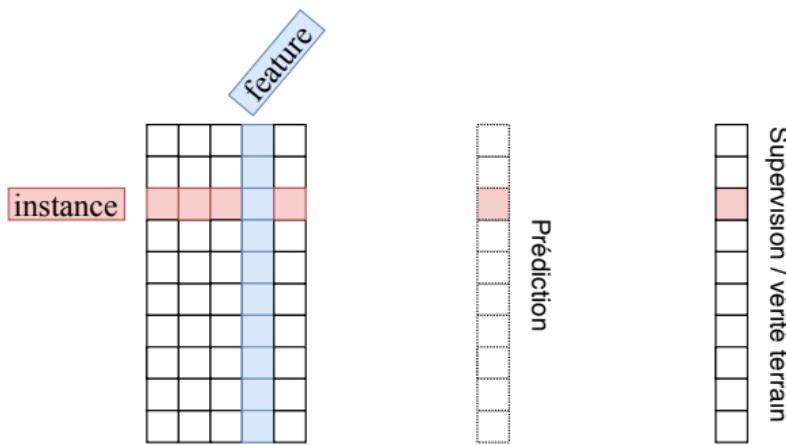
Est aussi important que l'apprentissage du modèle lui-même!



Mesurer les performances

Estimer les performances (en généralisation)...

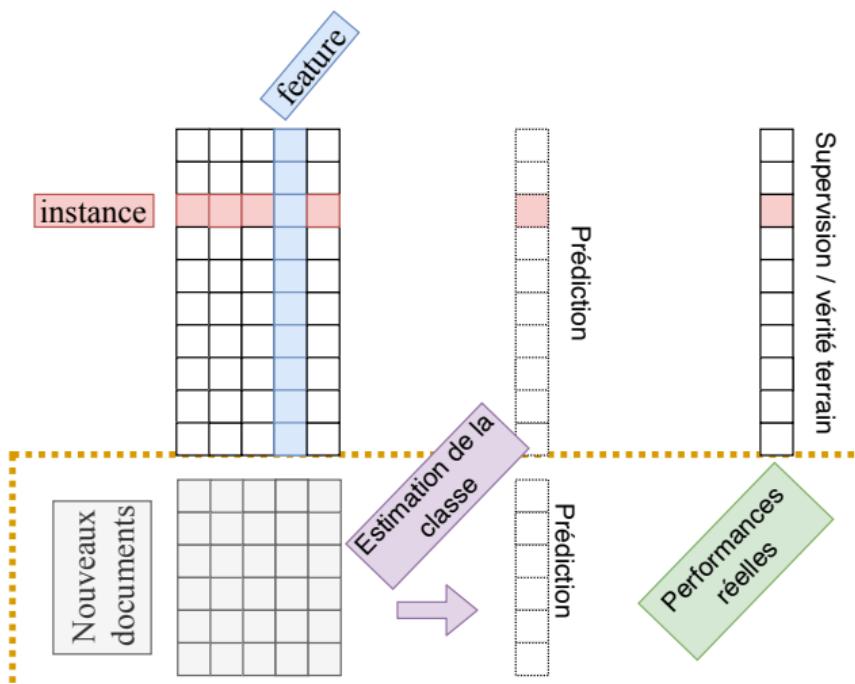
Est aussi important que l'apprentissage du modèle lui-même!



Mesurer les performances

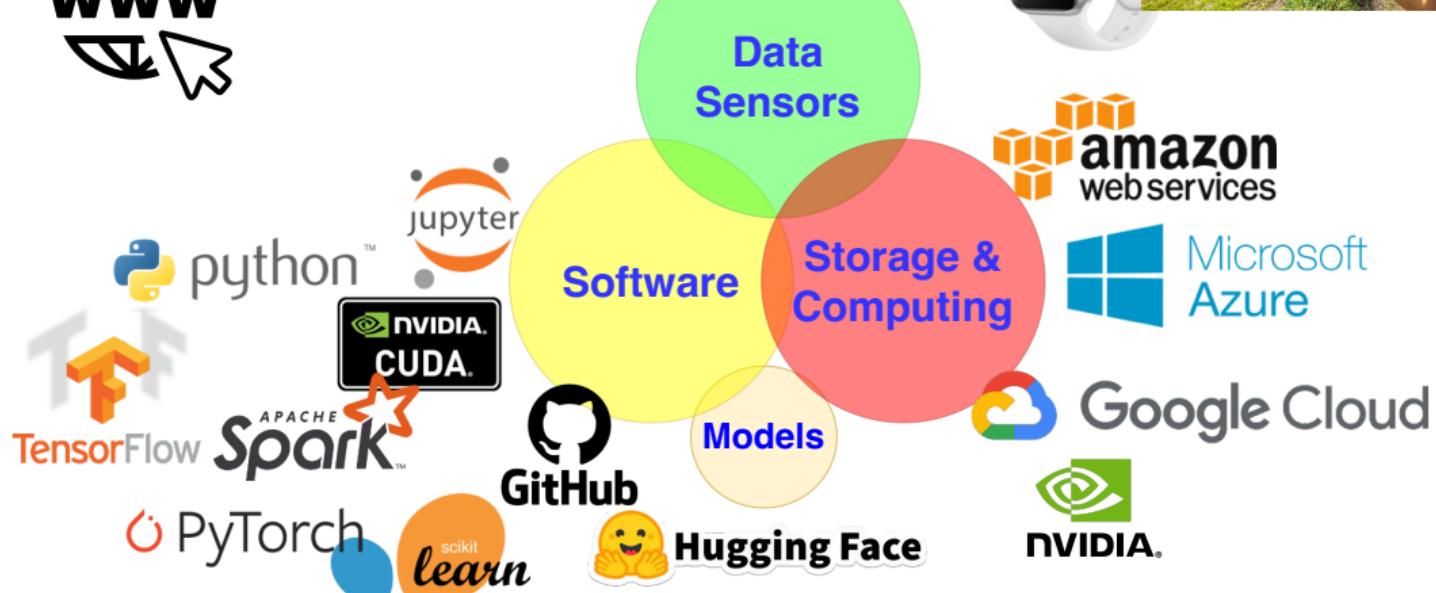
Estimer les performances (en généralisation)...

Est aussi important que l'apprentissage du modèle lui-même!





Ingrédients de l'Intelligence Artificielle



Accès aux données

- VOC 2007, caltech256
 - milliers d'images, dizaines de catégories
- ImageNet 2009:
 - 3.5M d'images annotées, 1000 catégories (moins de 0.3% d'erreur)
Construction sur plusieurs années : objectif 50M images
Localisation pour 1.2M d'objets (2012)
- Google car
 - 2018: *5 million miles on public roads and more than 5 billion miles via simulation*
- Mobileye
 - 500 personnes au Sri-Lanka: étiquetage scènes/objets
 - Réduire les modèles pour les embarquer dans une caméra
- (Google) Captcha



L'augmentation de la puissance de calcul et des tailles de bases de données expliquent les succès actuels

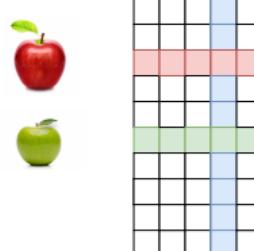
DEEP-LEARNING & NLP^{*}

[^{*} TRAITEMENT AUTOMATIQUE DE LA LANGUE
NATURELLE]



From tabular data to text

- Tabular data
 - Fixed dimension
 - Continuous values

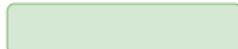


$$\rightarrow f(\boxed{\quad \quad \quad}) = \text{pred}$$

- Textual data
 - Variable length
 - Discrete values

this new iPhone, what a marvel

An iPhone? What a scam!



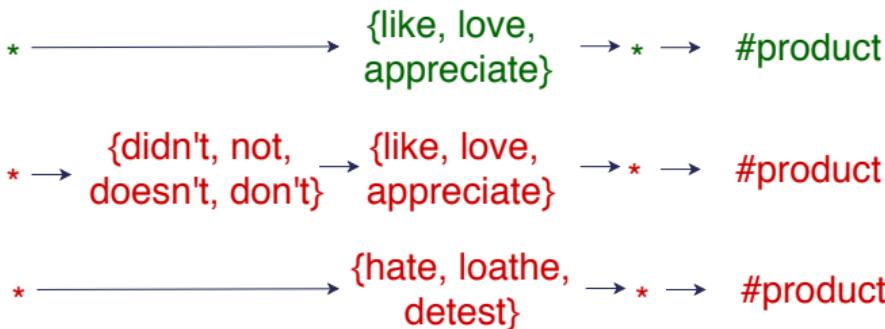


AI + Textual Data: Natural Language Processing (NLP)

NLP = largest scientific community in AI

Linguistics [1960-2010]

Rule-based Systems:



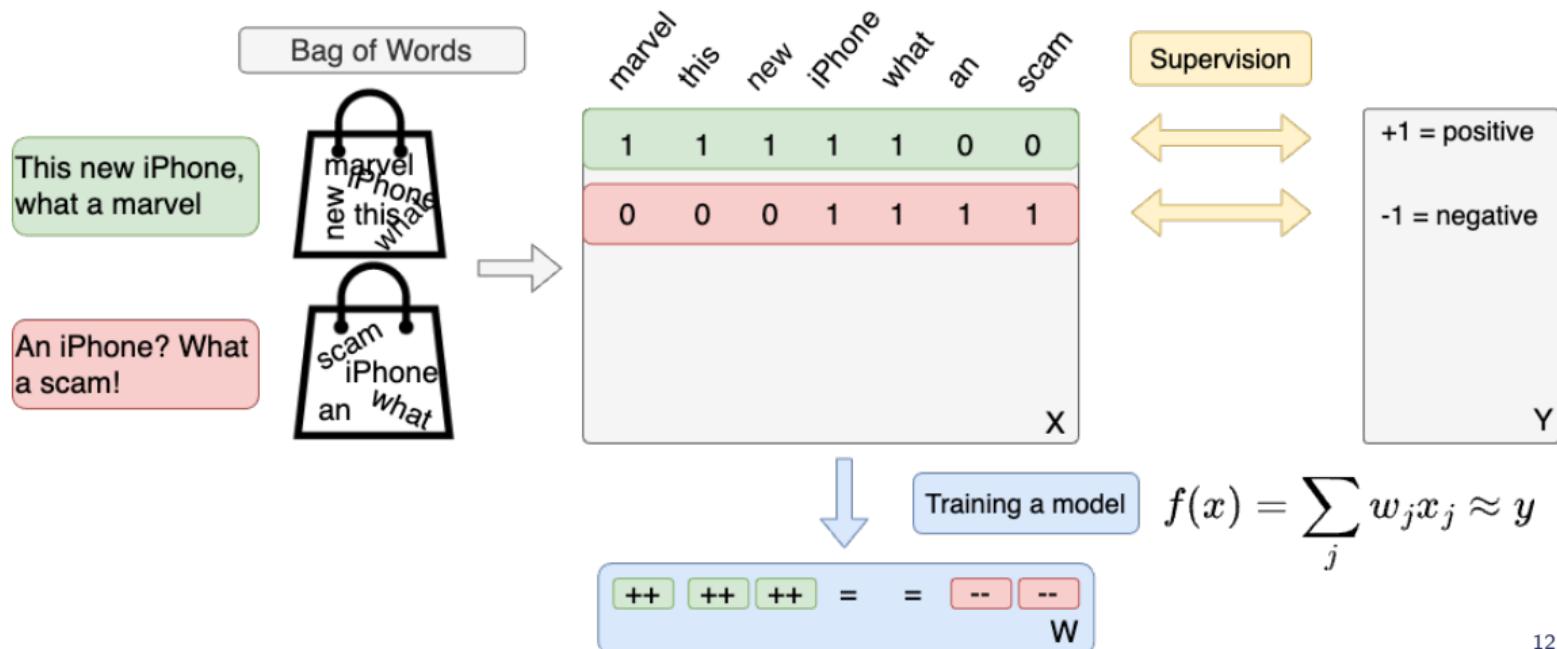
- Requires expert knowledge
- Rule extraction ⇔ very clean data
- Very high precision
- Low recall
- Interpretable system



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 ⇔
very clean data
- + Interpretable system
- + Very high precision
- Low recall

Machine Learning [1990-2015]

- Little expert knowledge needed
- Statistical extraction ⇔
robust to noisy data
- ≈ Less interpretable system
- Lower precision
- + Better recall

Precision = criterion for acceptance by industry

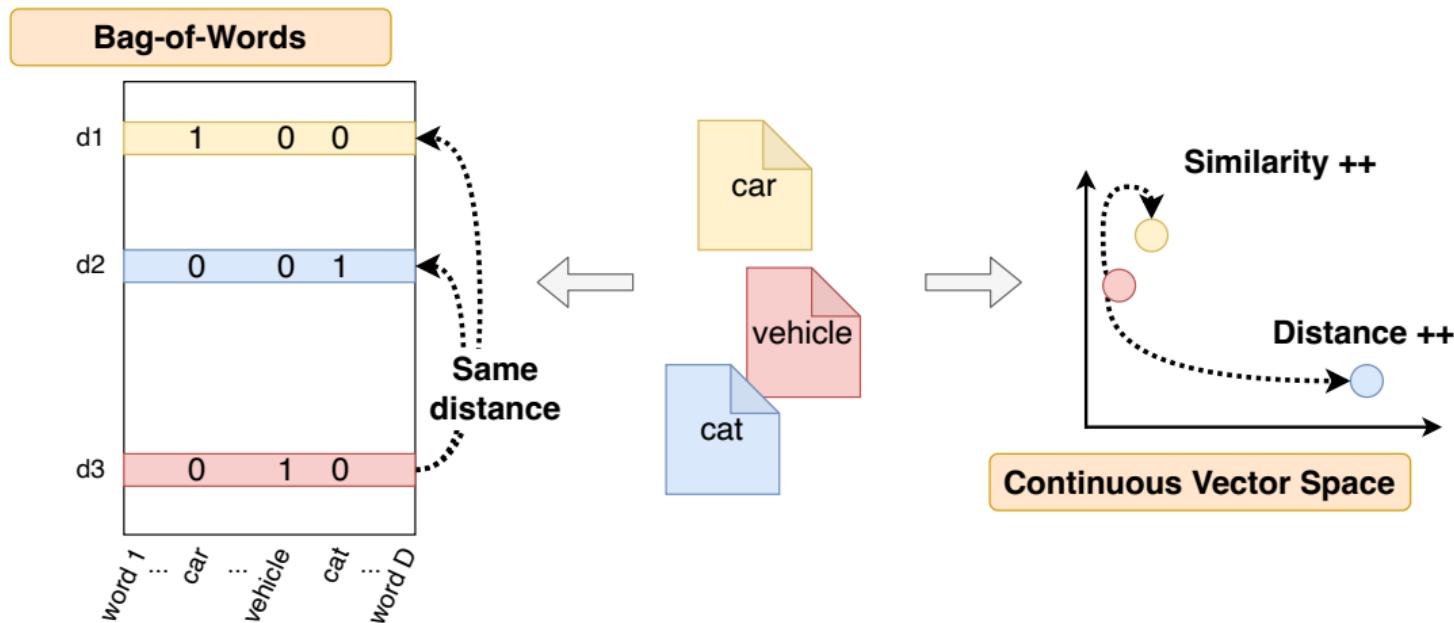
→ Link to metrics



Deep/Representation Learning for Text Data

From Bag of Words to Vector Representations

[2008, 2013, 2016]

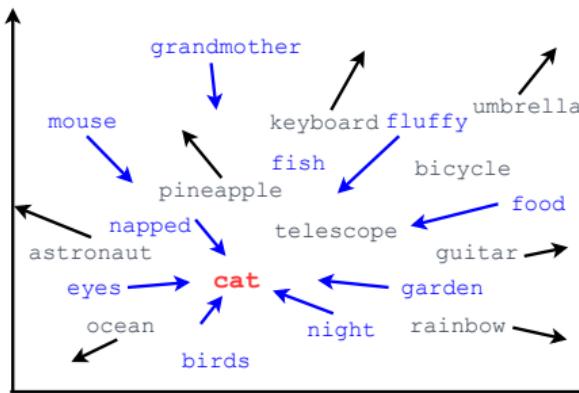
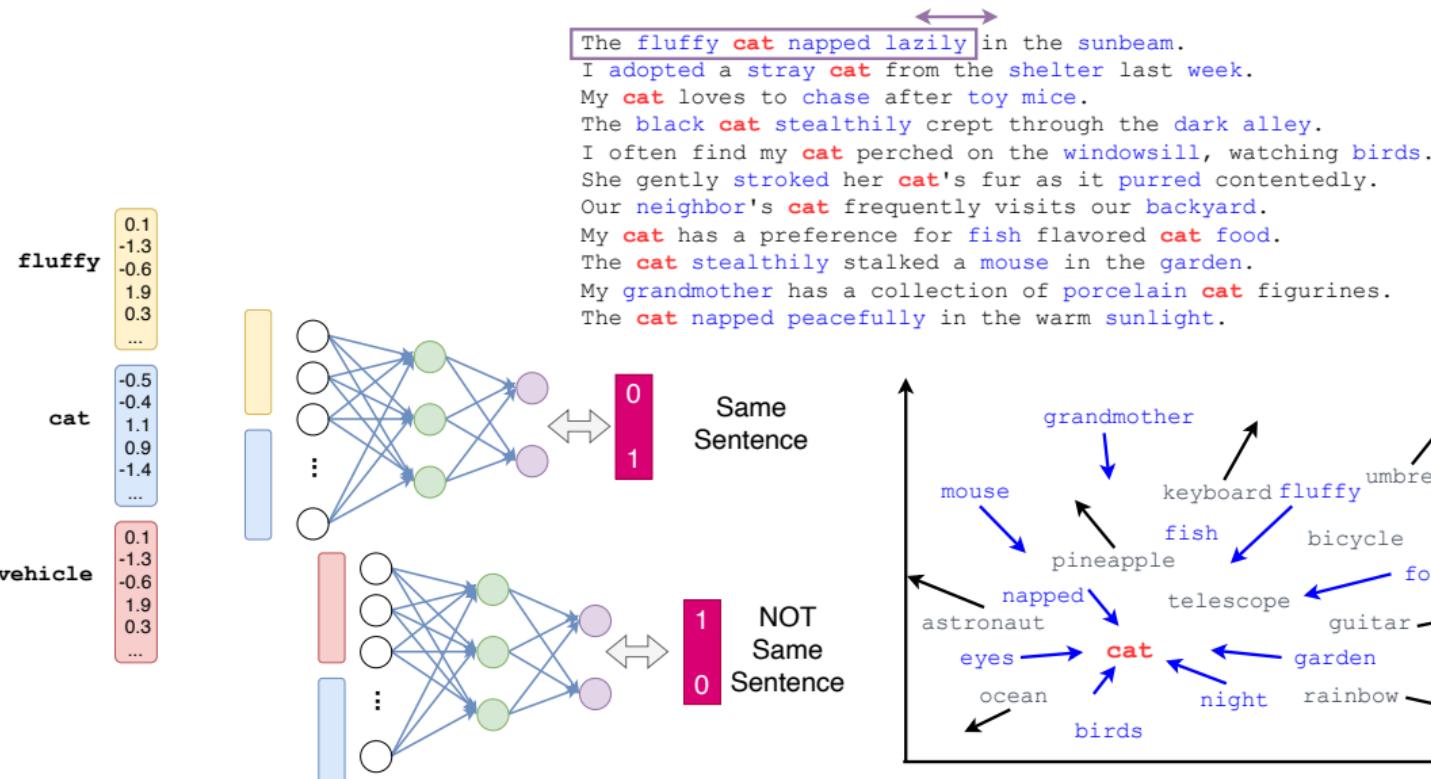


A

Deep/Representation Learning for Text Data

From Bag of Words to Vector Representations

[2008, 2013, 2016]

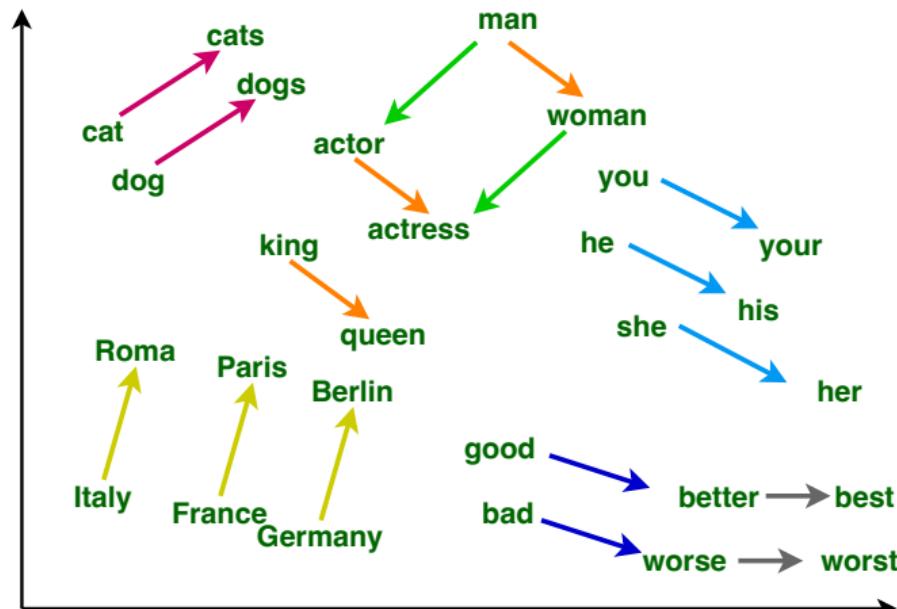




Deep/Representation Learning for Text Data

From Bag of Words to Vector Representations

[2008, 2013, 2016]



- Semantic Space:
similar meanings
↔
close positions
- Structured Space:
grammatical regularities,
basic knowledge, ...



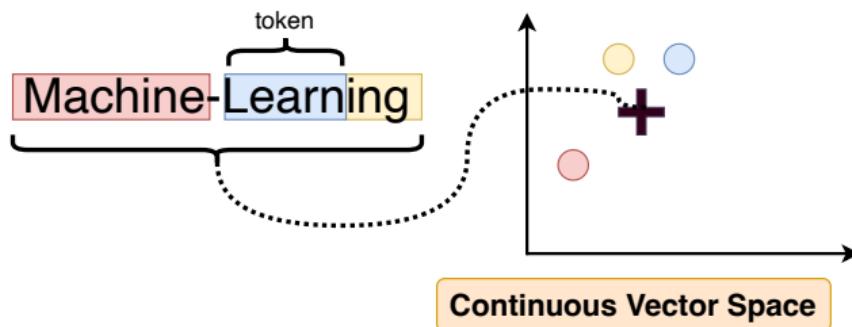
Deep/Representation Learning for Text Data

From Bag of Words to Vector Representations

[2008, 2013, 2016]

From Words to Tokens

Word Piece statistical split



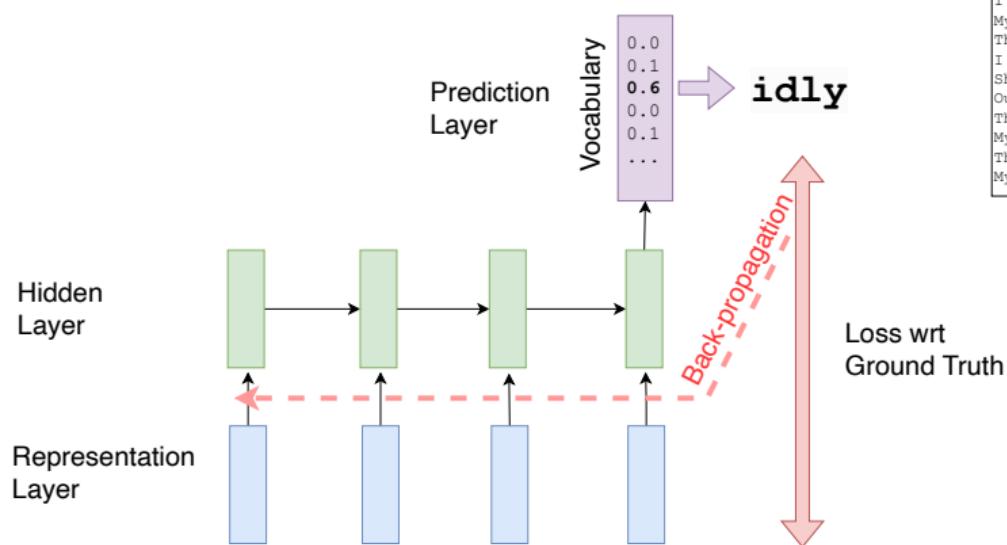
- Representation of unknown words
- Adaptation to technical domains
- Resistance to spelling errors

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.

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



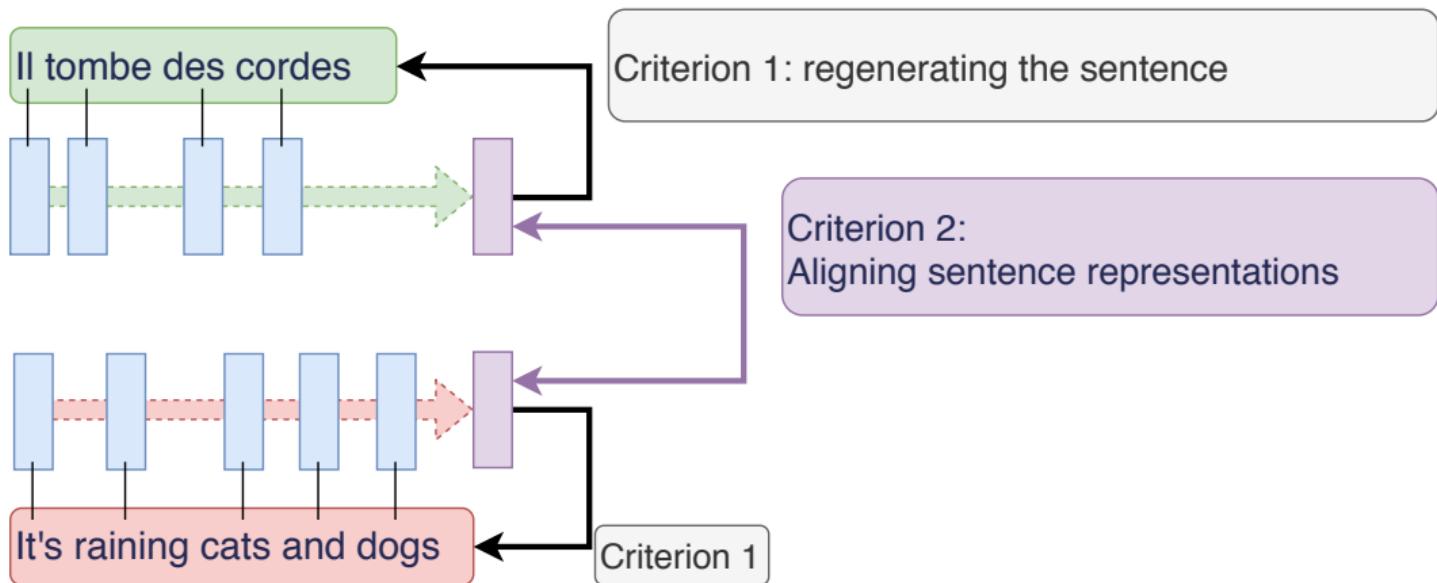
Use-Case: Machine Translation



Beyond word-for-word translation, multilingual representation of sentences



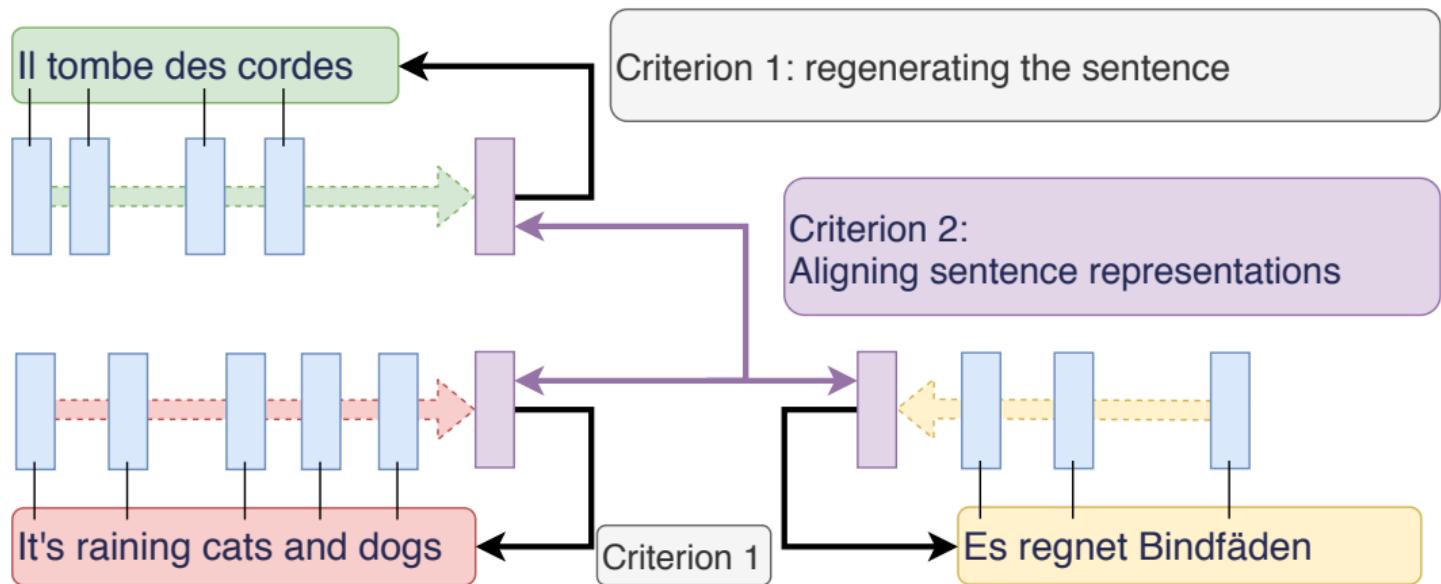
Use-Case: Machine Translation



Beyond word-for-word translation, multilingual representation of sentences



Use-Case: Machine Translation



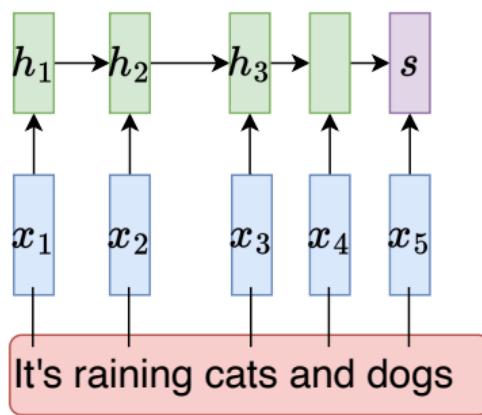
Beyond word-for-word translation, multilingual representation of sentences



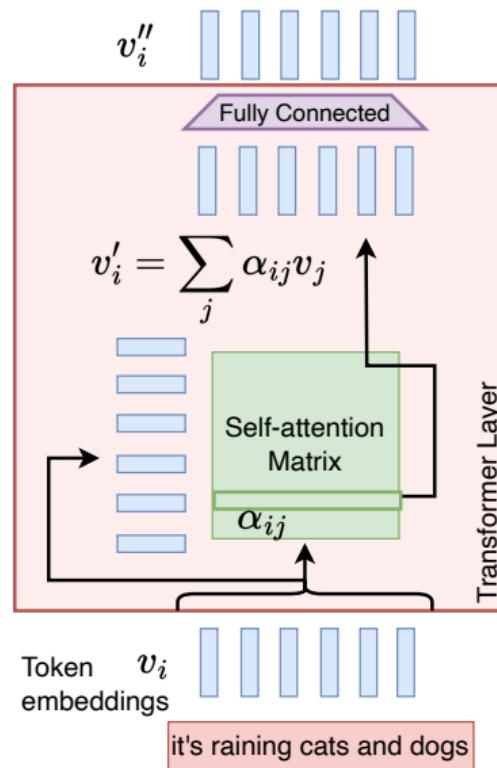
Transformer architecture: state-of-the-art aggregation

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](#)

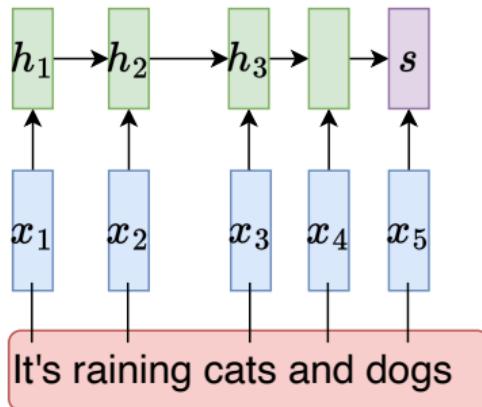
Sequence to Sequence Learning with Neural Networks, [Sutskever et al. NeurIPS 2014](#)



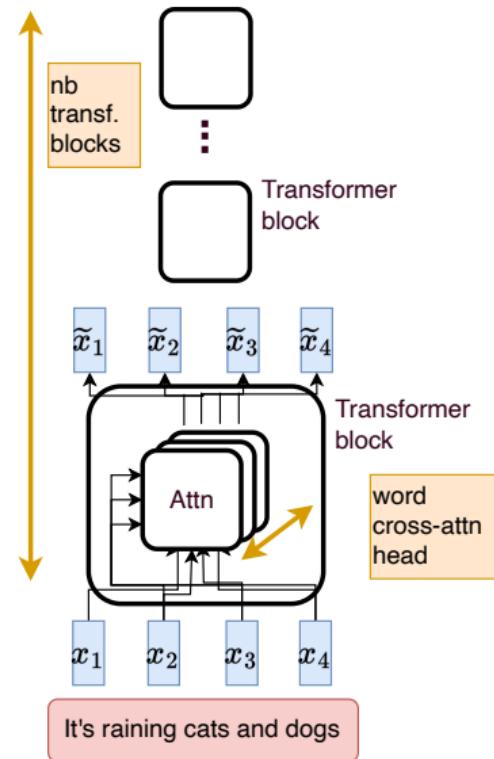
Transformer architecture: state-of-the-art aggregation

Recurrent Neural Network:

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



Transformer:

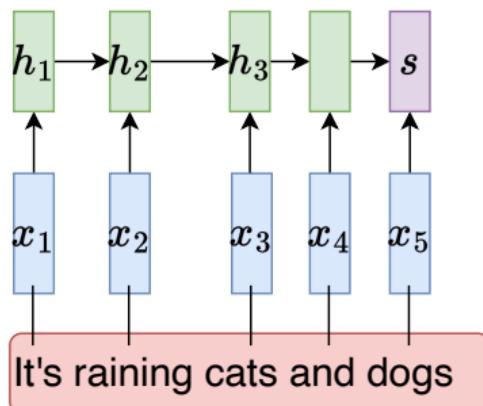




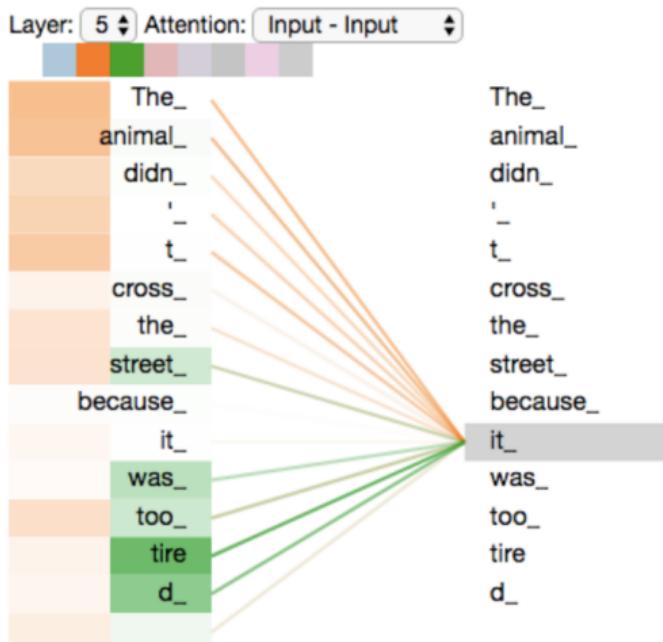
Transformer architecture: state-of-the-art aggregation

Recurrent Neural Network:

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



Transformer:



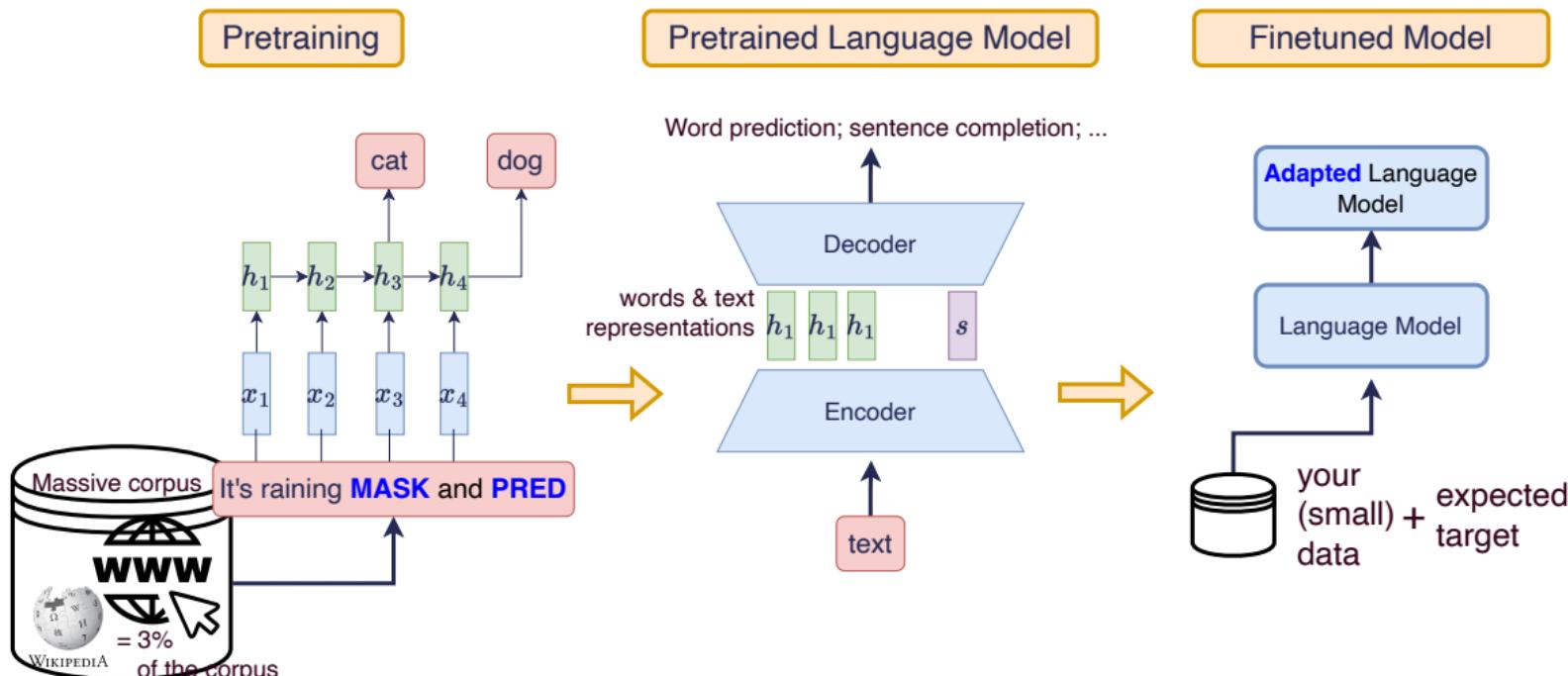
Attention is all you need, [Vaswani et al. NeurIPS 2017](#)

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

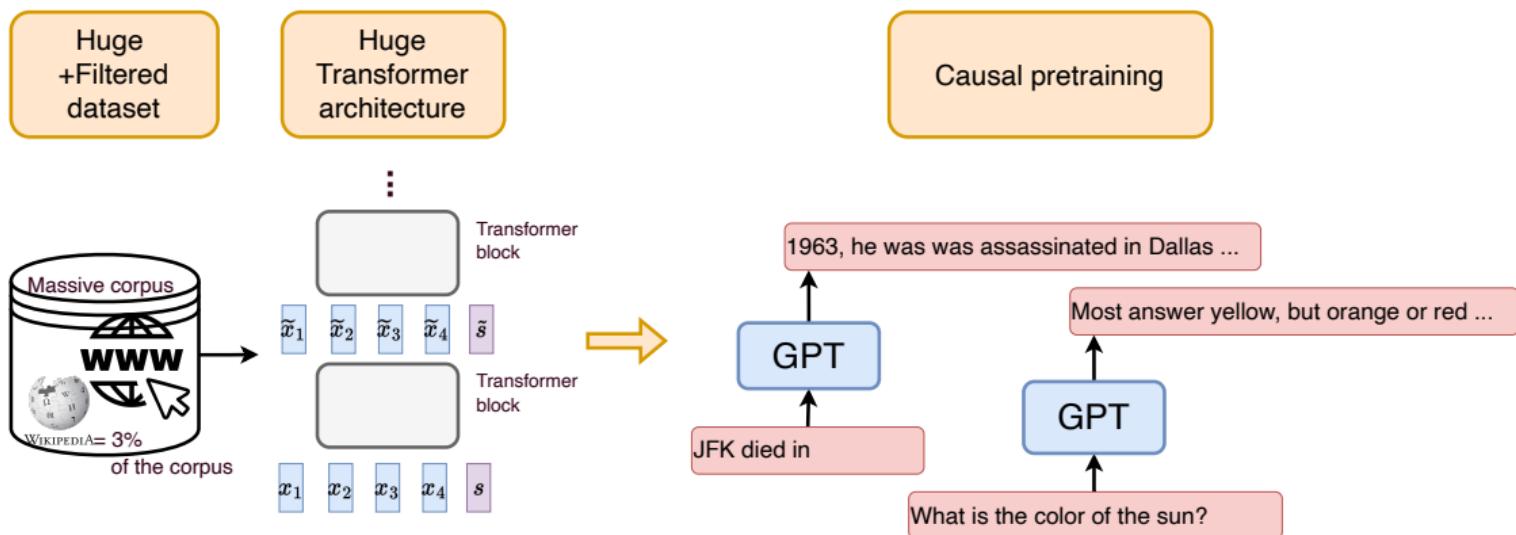
1 MILLION USERS IN 5 DAYS

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: entities, names, dates, places



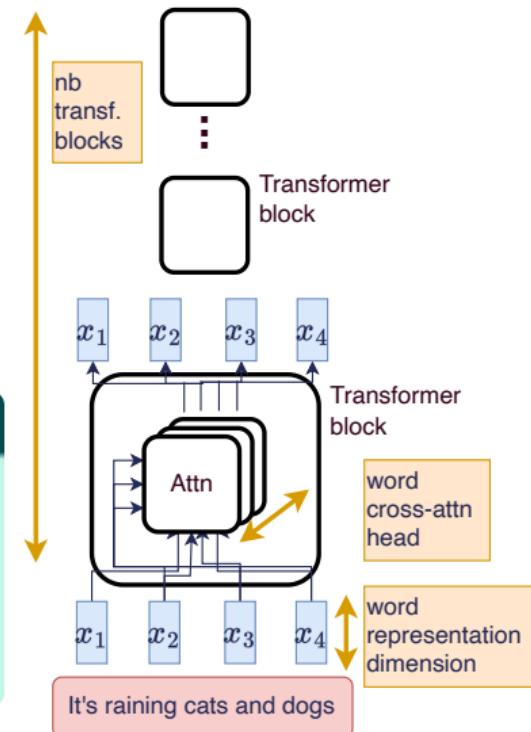
The Ingredients of chatGPT

1. More is better! (GPT)

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

175 Billion parameters... What does it mean?

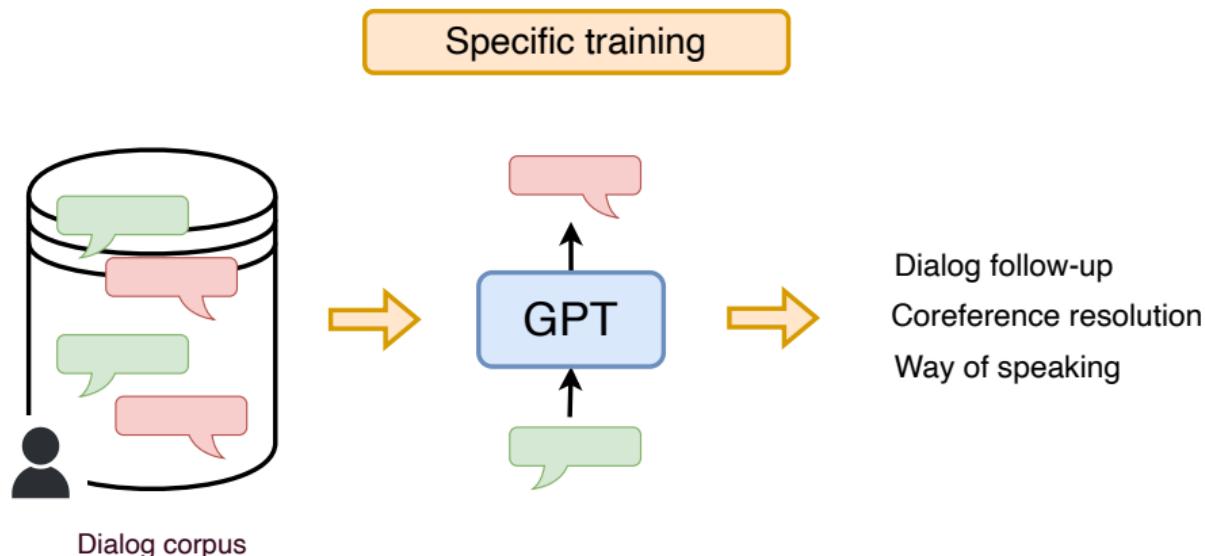
- $1.75 \cdot 10^{11} \Rightarrow 300 \text{ GB} + 100 \text{ GB}$ (data storage for inference) $\approx 400\text{GB}$
- NVidia A100 GPU = 80GB of memory (=20k€)
- Cost for (1) training: 4.6 Million €





The Ingredients of chatGPT

2. Dialogue Tracking



■ **Very clean** data

Data generated/validated/ranked by humans



The Ingredients of chatGPT

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

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?

-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 $3 + 6 = 9$.

Language model

Multi-task instruction finetuning (1.8K tasks)

Inference: generalization to unseen tasks

Q: Can Geoffrey Hinton have a conversation with George Washington?

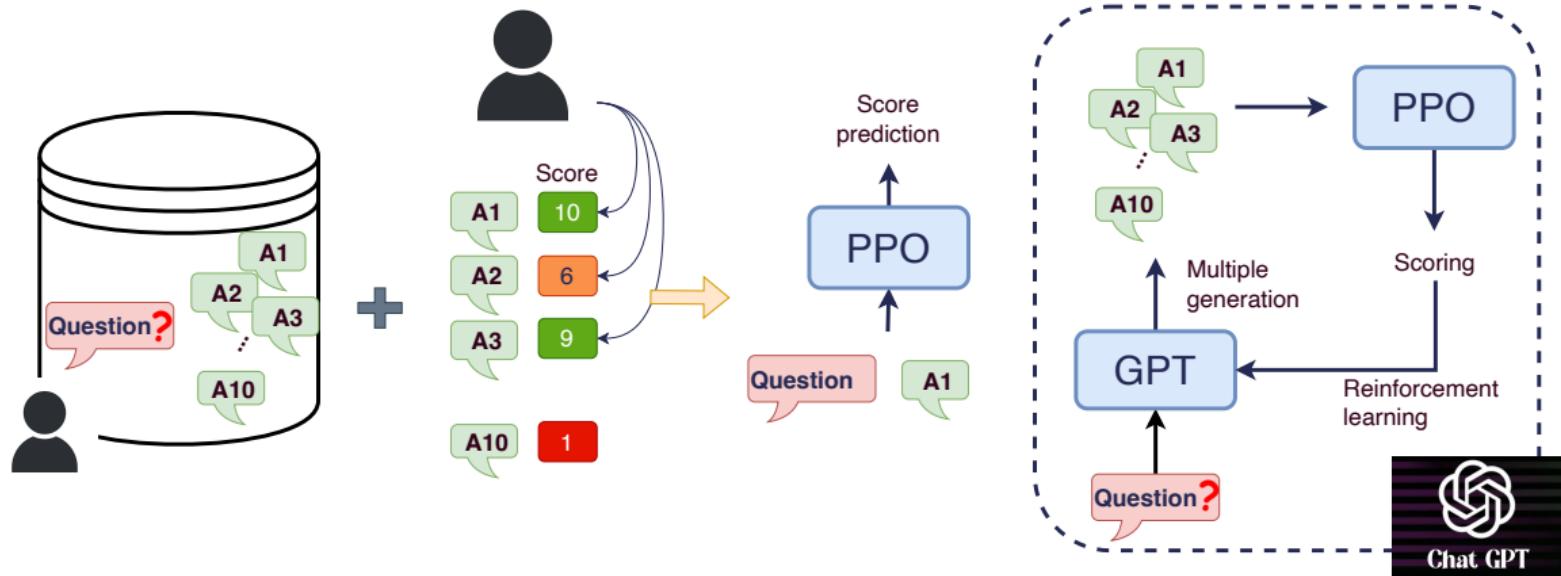
Give the rationale before answering.

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".



The Ingredients of chatGPT

4. Instructions + answer ranking



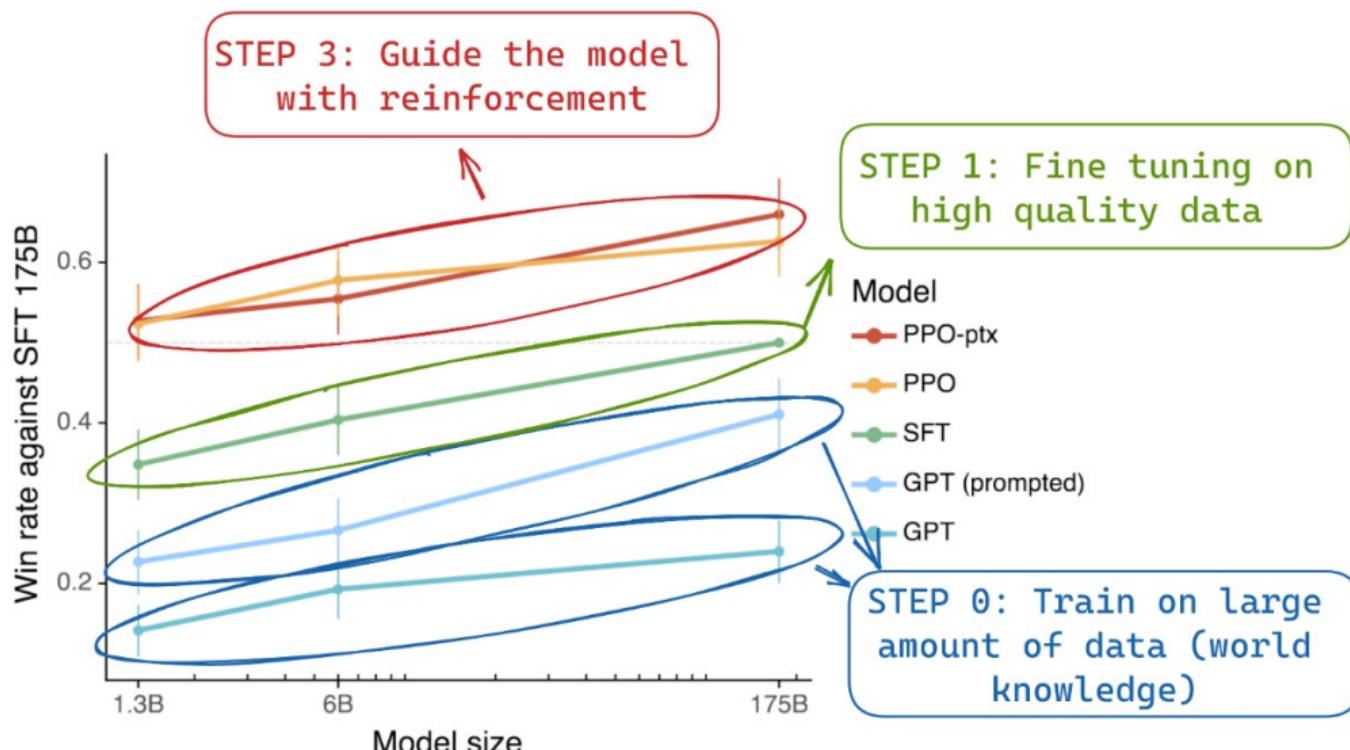
- Database created by humans
- Response improvement

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



Steps & Performance

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





Usage of chatGPT & Prompting

- Asking chatGPT = skill to acquire ⇒ *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

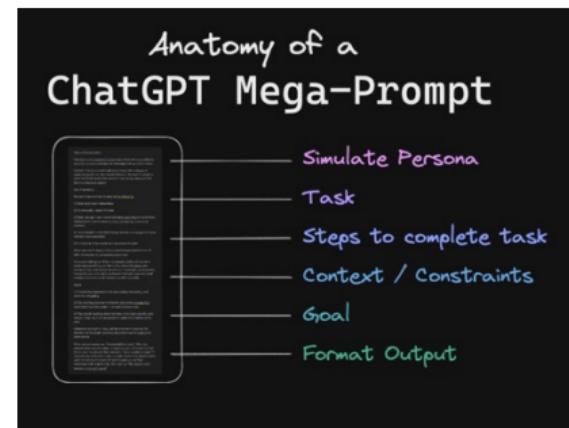
■ Don't stop at the first question

- Detail specific points
- Redirect the research
- Dialogue

■ Rephrasing

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

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

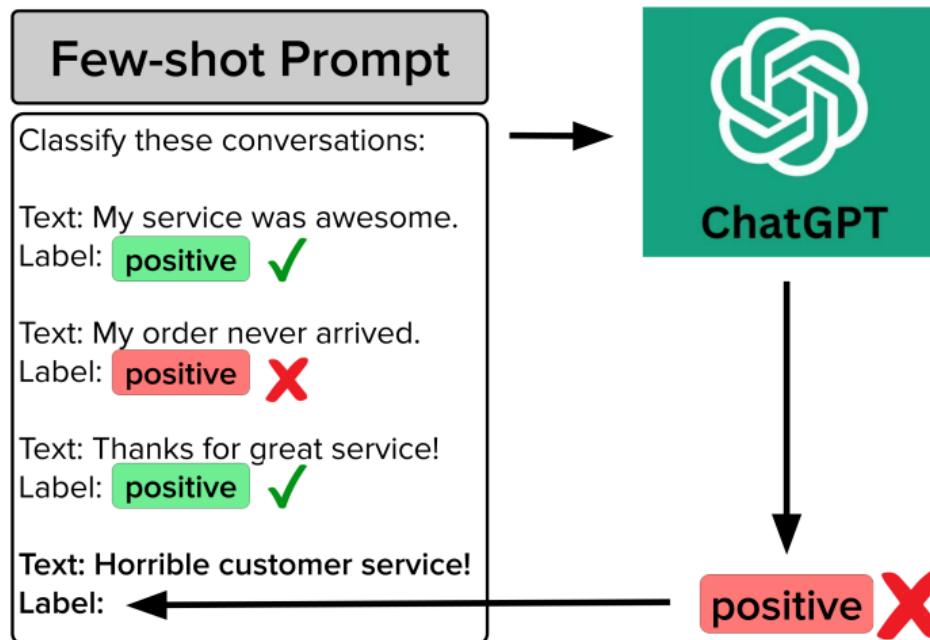


<https://chatgptprompts.guru/what-makes-a-good-chatgpt-prompt/>



Towards few-shot learning

- Learning without modifying the model = examples in the prompt

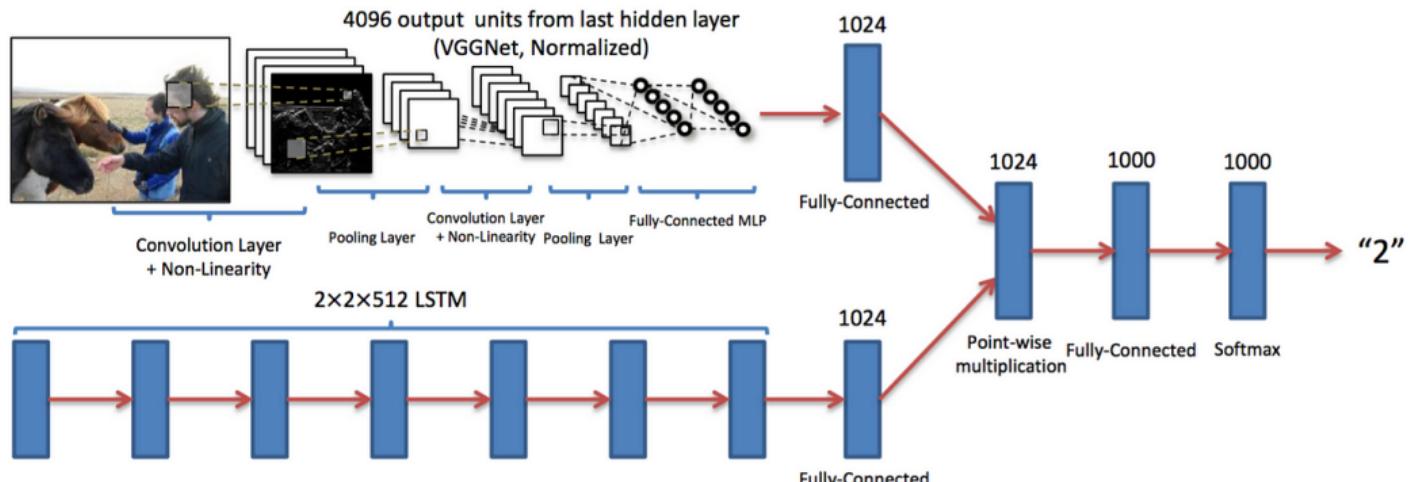




GPT4 & Multimodality

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

The example of VQA: visual question answering



"How many horses are in this image?"

⇒ Backpropagate the error ⇒ modify word representations + image analysis

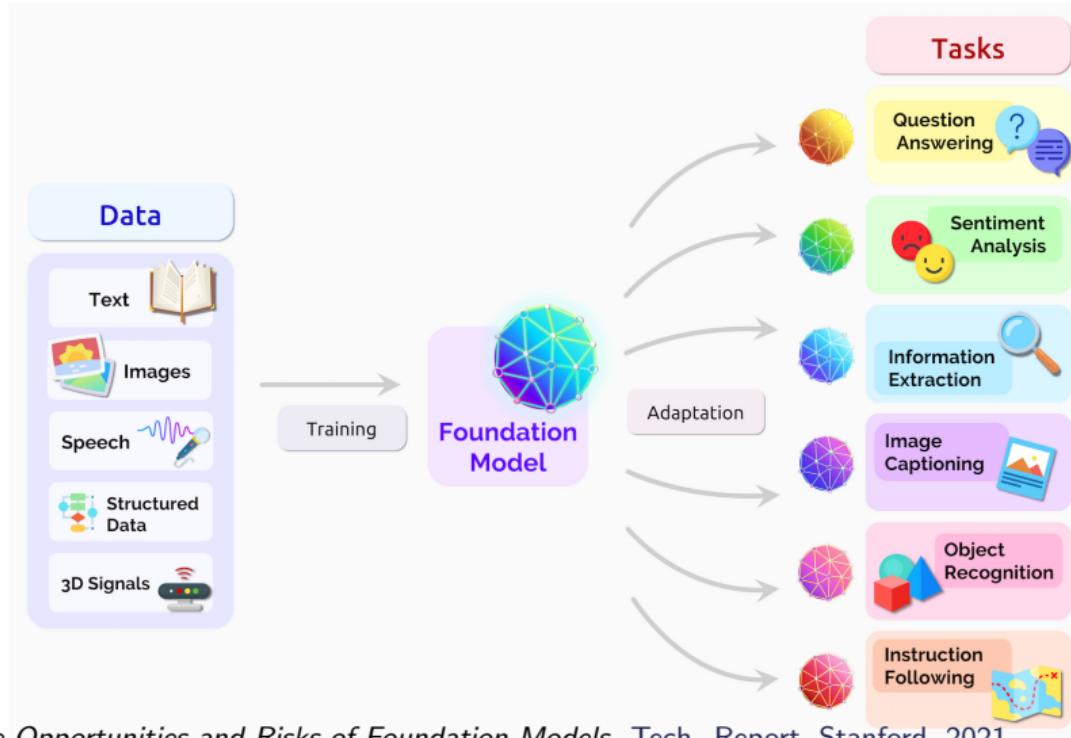


VQA: Visual Question Answering, arXiv, 2016 , A. Agrawal et al.



Towards Larger Foundation Models?

- Let the modalities enrich each other

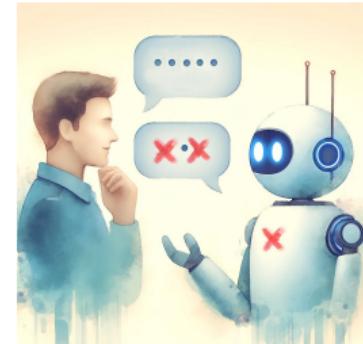


On the Opportunities and Risks of Foundation Models, Tech. Report, Stanford, 2021
Bommasani et al.



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





At the end of the day

Statistical Modeling of Texts

Texts splitting = tokens

Large Language Models (LLMs), such as GPT-3 and GPT-4, utilize a process called tokenization. Tokenization involves breaking down text into smaller units, known as tokens, which the model can process and understand. These tokens can range 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

Iterative Process

Dictionary	Large entire For units ... can may ...	0.02 0.01 0.00 0.00 0.00 0.09 ...
------------	---	---

0.02 0.01 0.00 0.00 0.00 0.09 ...

Starting text

Language Model

Token forecasting

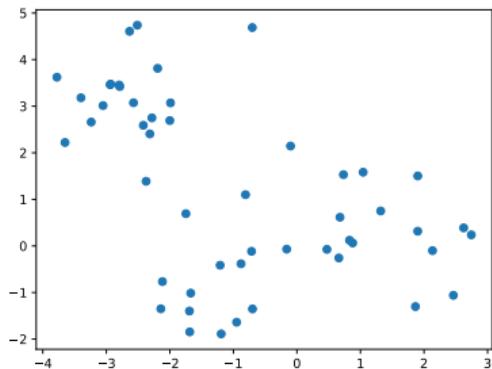
FROM GENERATIVE AI TO FOUNDATION MODELS



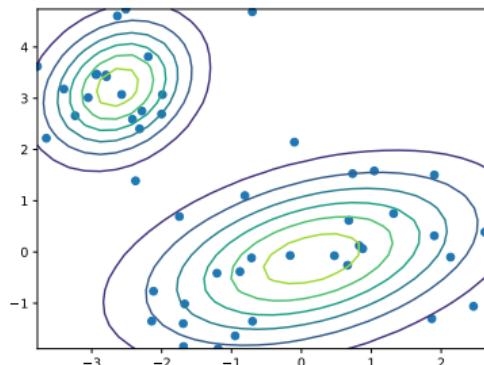
At the origin of statistical modeling

- 1 **Observing** data (and context)
- 2 **Modeling** = Choosing probabilistic model / bayesian network
- 3 **Optimize** parameters (Max. Likelihood, EM, BFGS, ...)
- 4 **Sampling** / Inference + Evaluate distances : existing vs sampled

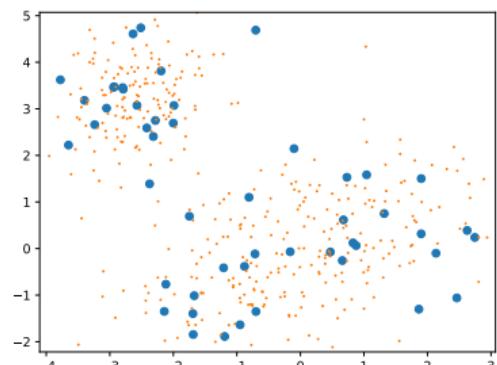
Observations



Modeling: choice+optim.



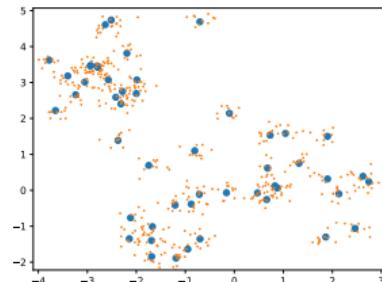
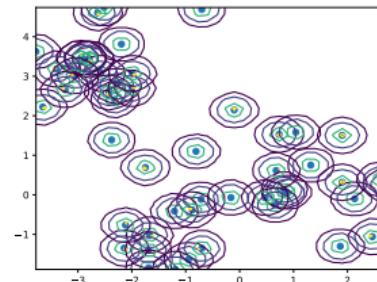
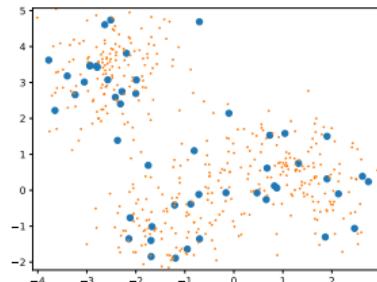
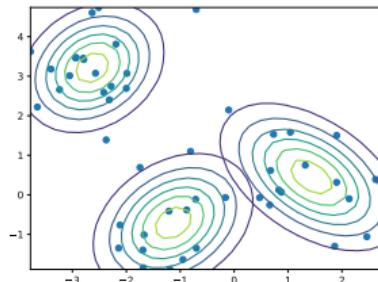
Sampling / eval.



At the origin of statistical modeling

- 1 **Observing** data (and context)
- 2 **Modeling** = Choosing probabilistic model / bayesian network
- 3 **Optimize** parameters (Max. Likelihood, EM, BFGS, ...)
- 4 **Sampling** / Inference + Evaluate distances : existing vs sampled

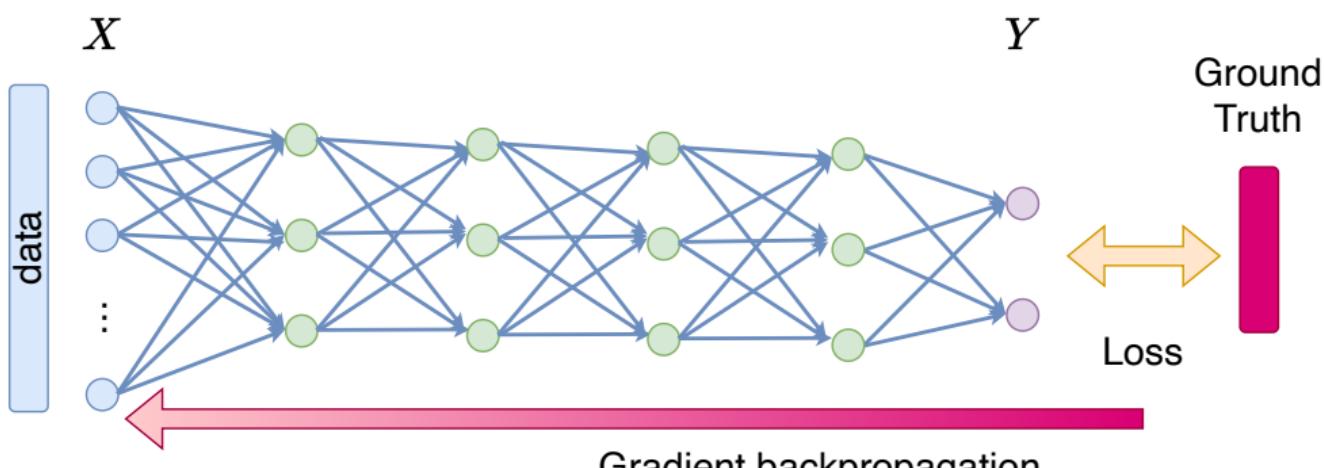
Different modeling options / different traps





At the origin of deep learning

- Gradient vanishing issue in deep architecture

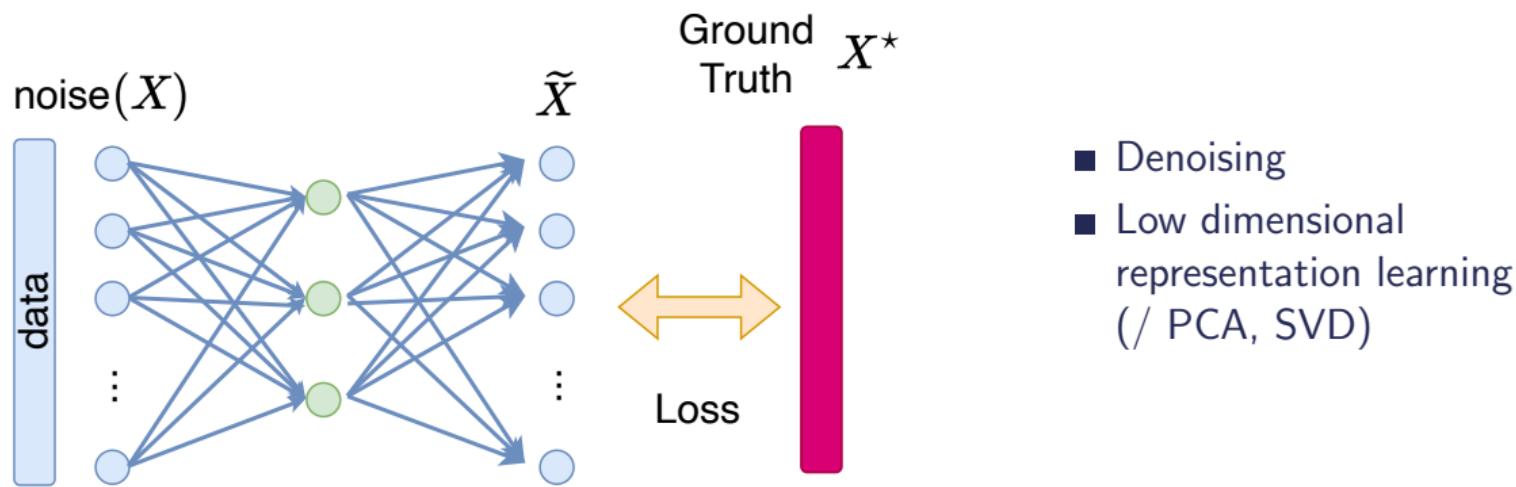


Gradient **weakening => vanishing**



At the origin of deep learning

- Gradient vanishing issue in deep architecture
- Auto-Encoder architecture / facing unsupervised dataset with NN

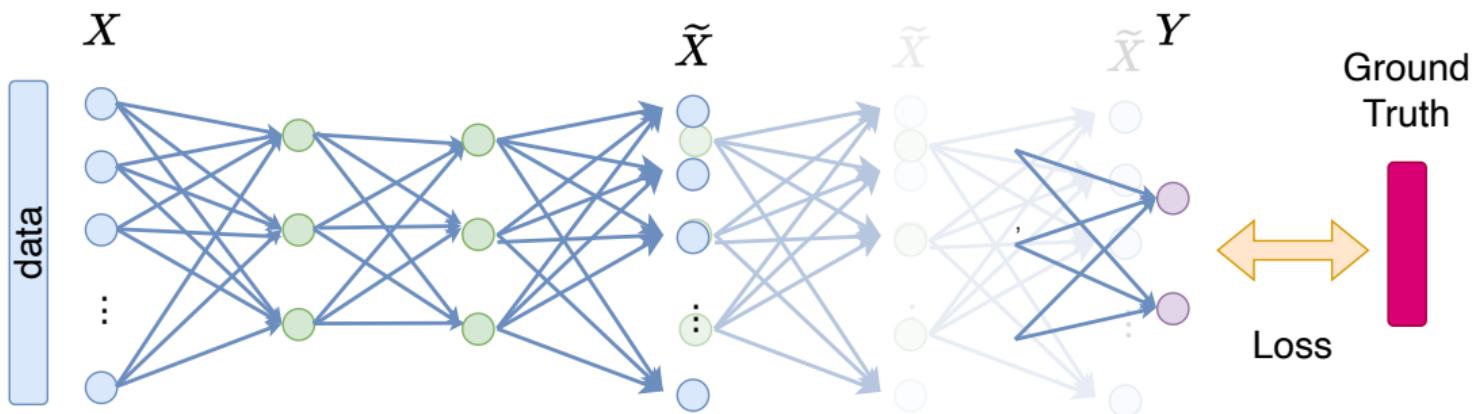


Auto-association by multilayer perceptrons and singular value decomposition, Biological Cybernetics, 1988
H. Bourlard & Y. Kamp



At the origin of deep learning

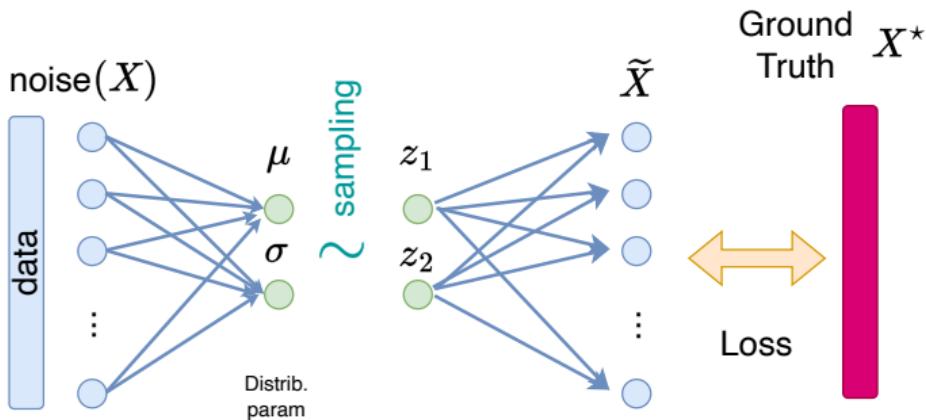
- Gradient vanishing issue in deep architecture
- Auto-Encoder architecture / facing unsupervised dataset with NN
- Stacked Denoising Auto-Encoder : iterative training / **pretraining**



The difficulty of training deep architectures and the effect of unsupervised pre-training, AIS, PMLR 2009
Erhan, D., Manzagol, P. A., Bengio, Y., Bengio, S., & Vincent, P.



Variational Auto-Encoder



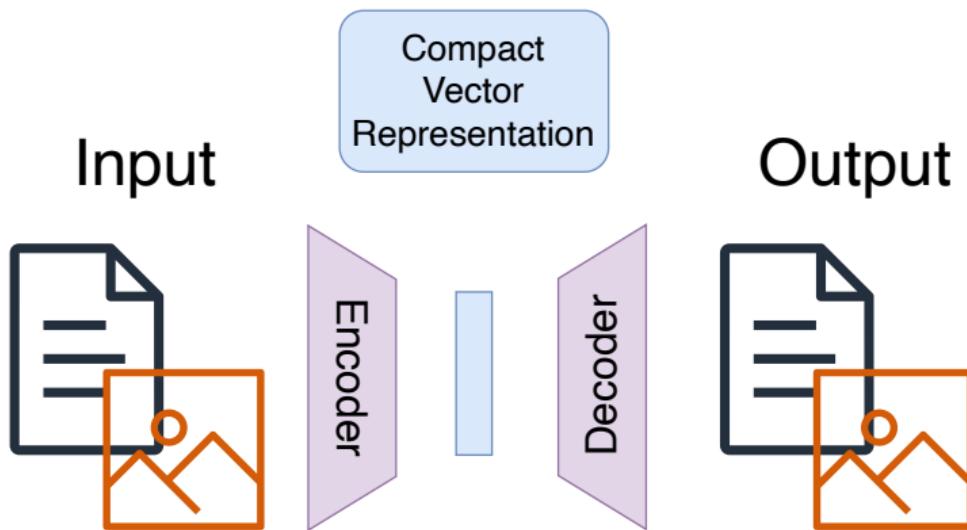
- a priori on the distribution
- Structuring of the latent space

Generative AI (for statisticians)



Auto-Encoding Variational Bayes, 2013
DP Kingma

Different Forms of Generative AI

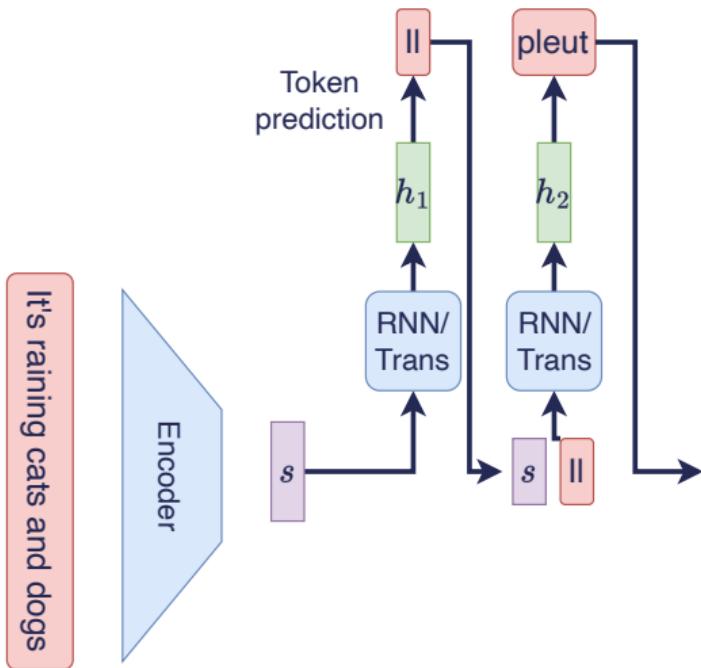


- 1 Encode an input = construct a vector
- 2 Decode a vector = *generate* an output



Different Media / Different Architectures

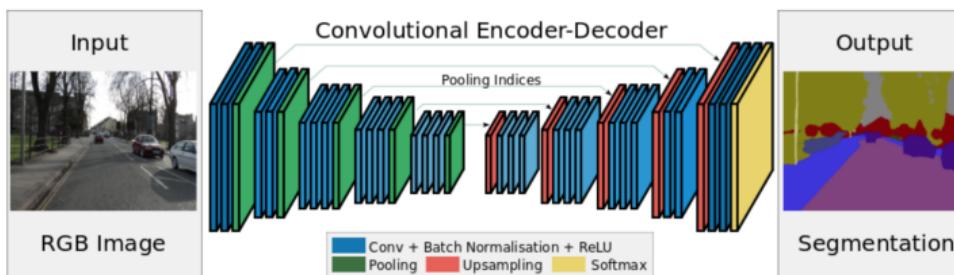
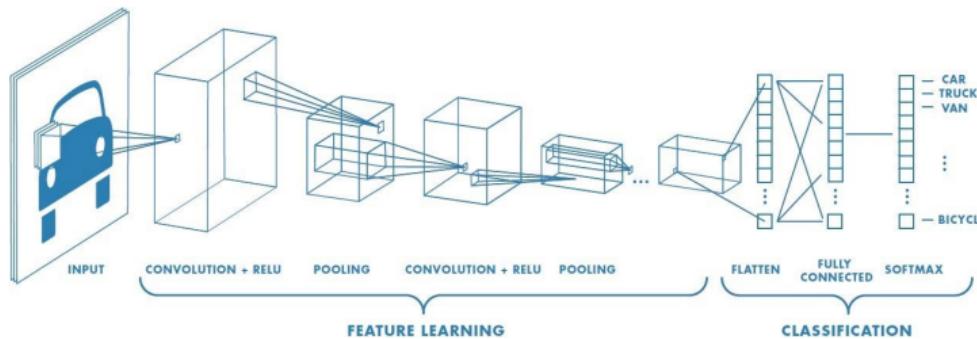
- Texts: classification problem





Different Media / Different Architectures

- Texts: classification problem
- Images: multivariate regression problem



*U-Net: Convolutional Networks for Biomedical Image Segmentation, MICCAI, 2015
Ronneberger et al.*

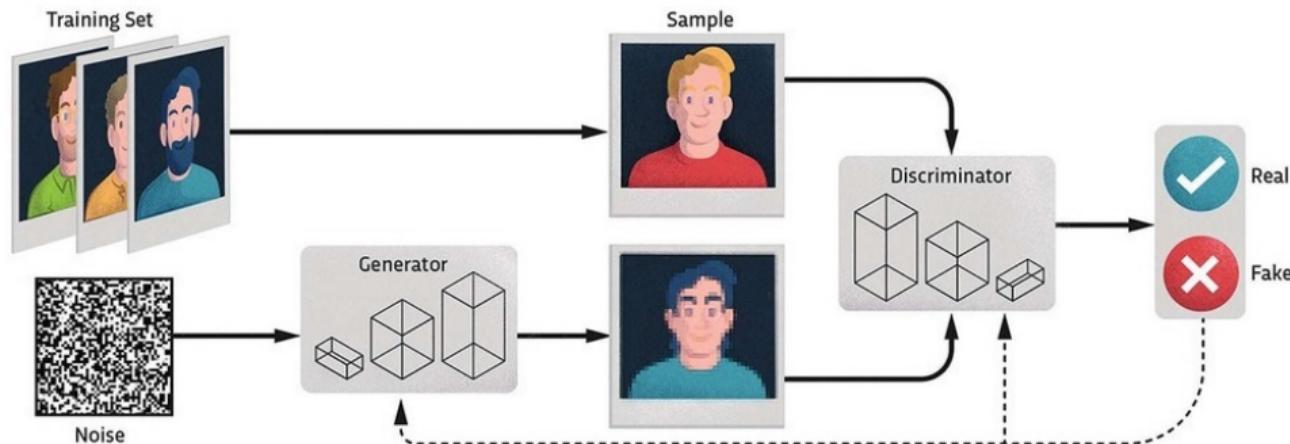
NVidia Lab.



Different Media / Different Architectures

- Texts: classification problem
- Images: multivariate regression problem

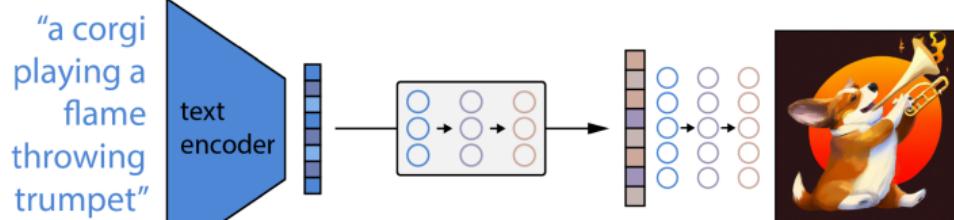
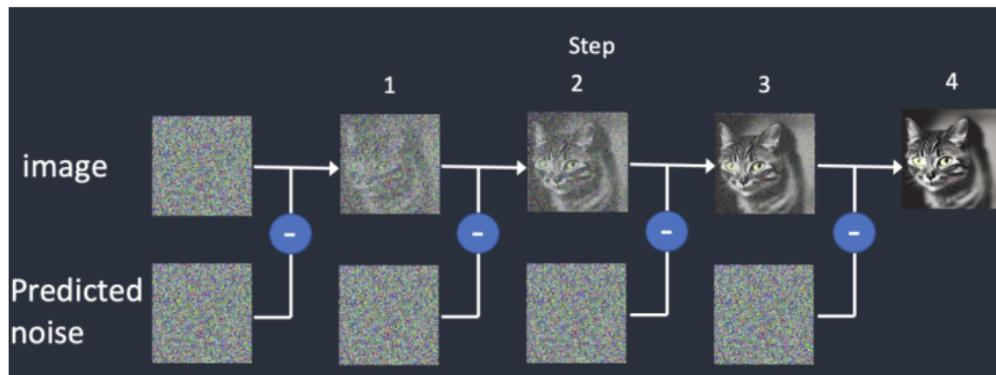
Generative Adversarial Networks (GAN): detecting generated samples



Generative Adversarial Nets, NeurIPS 2014
Goodfellow et al.

Different Media / Different Architectures

- Texts: classification problem
- Images: multivariate regression problem
- Physical processes



Denoising Diffusion Probabilistic Models, NeurIPS, 2020
Ho, J., Jain, A., & Abbeel, P.



Hierarchical Text-Conditional Image Generation with CLIP Latents, arXiv, 2022
Ramesh et al.



Different Media / Different Architectures

- Texts: classification problem
- Images: multivariate regression problem
- Physical processes
- Complex structures / 3D / graphs: sequential problem
- Mix mechanistic and *data-driven* approaches

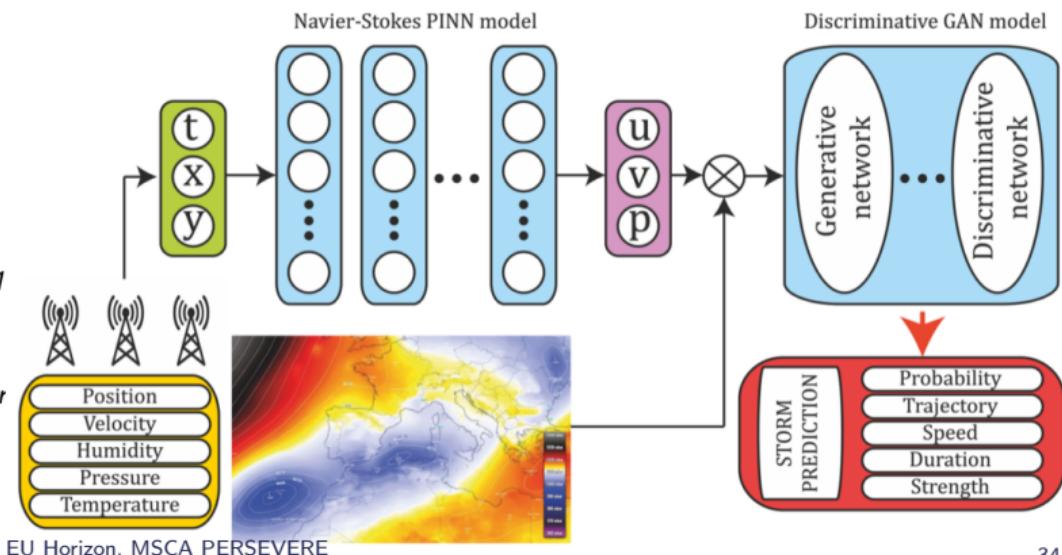
e.g. Model differential equations in a neural network



Neural ordinary differential equations, NeurIPS, 2018
Chen et al.



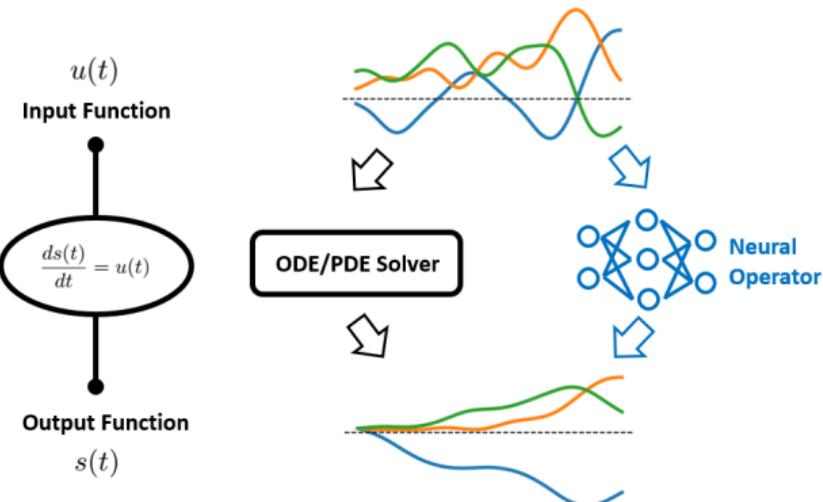
Physics-informed neural networks, J. Comp. Physics, 2019
Raissi et al.





Different Media / Different Architectures

- Texts: classification problem
- Images: multivariate regression problem
- Physical processes
- Complex structures / 3D / graphs: sequential problem



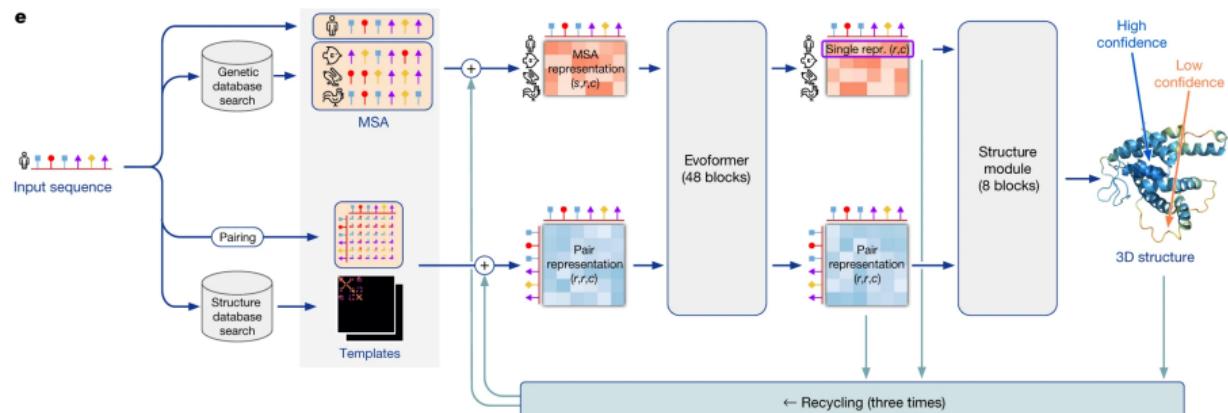
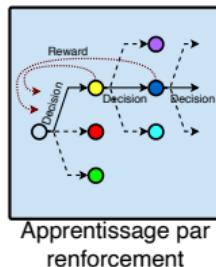
Data + Models :

- PDE, neural ODE
- Simulation approximations
- Residual Models
- Hybrid Complex Systems



Different Media / Different Architectures

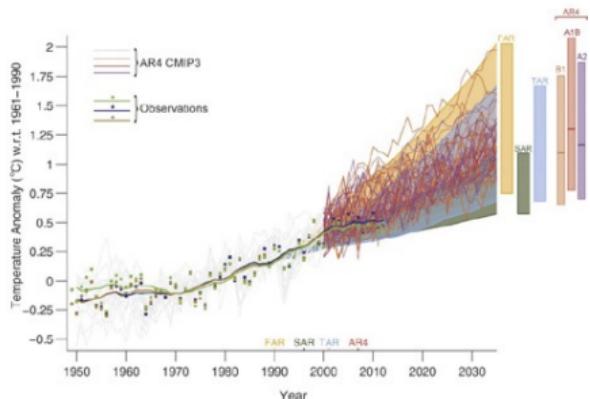
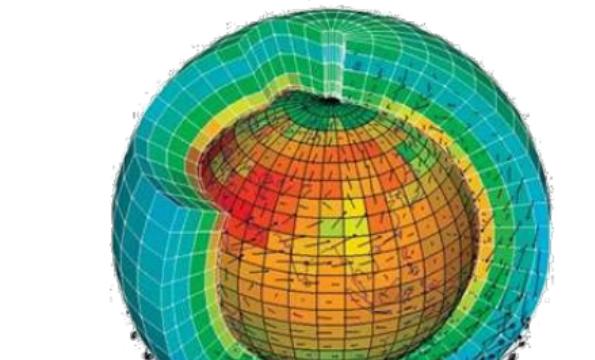
- Texts: classification problem
 - Images: multivariate regression problem
 - Physical processes
 - Complex structures / 3D / graphs: sequential problem
- Reinforcement learning: action/reward



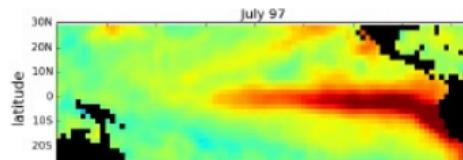
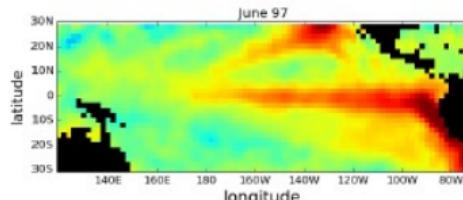
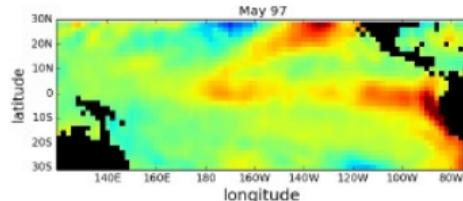
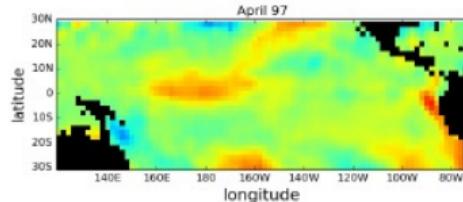
Highly accurate protein structure prediction with AlphaFold, Nature, 2021
Jumper et al.



Data-driven vs Modeling



Ground Truth





Data-driven vs Modeling

Chapitre T8 - Conduction thermique

4. Résistance thermique (ECU)

2. Géométrie radiale cylindrique

$R_{\text{m,eff}} = \frac{(T_c - T_f)}{\dot{Q}_{\text{in}}}$

$\forall x \in [0, R + \epsilon]$ $\dot{Q}_{\text{in}} = \dot{Q}_c = \text{const}$

au EPS, même condition, mais pas de valeur

soit n quelque chose $\in [0, R + \epsilon]$

$\dot{Q}_{\text{in}}(x) = f(x) \times 2\pi x L$

Loi de Fourier

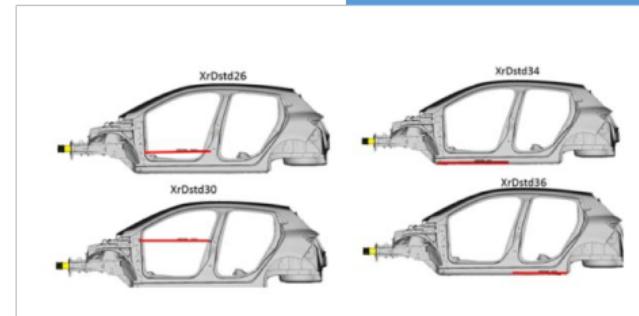
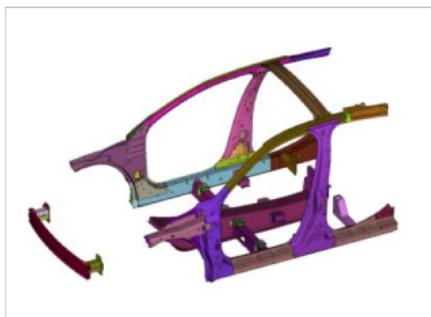
$\Rightarrow \dot{Q}_{\text{in}}(x) = -\lambda \frac{dT}{dx} 2\pi x L$

comme

$\Rightarrow \dot{Q}_{\text{in}} = -2\pi A L \frac{dT}{dr}$

Frontal crash model

6 Million Finite Elements
n HPC





Data-driven vs Modeling

Mecanistic model / simulation

$\forall x \in [r; R+r] \quad \Phi_{k(x)} = \Phi_k = \omega_k$

en EPS, nous calculons, ni perte de chaleur

soit si quelque $x \in [r, r+\epsilon]$

$$\Phi_{k(x)} = \delta_{k(x)} \times 2\pi r \epsilon$$

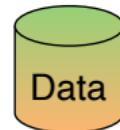
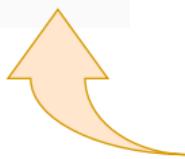
loi de Fourier

$$\Rightarrow \Phi_{k(n)} = -\lambda \cdot \frac{dT}{dr} \cdot 2\pi r \epsilon$$

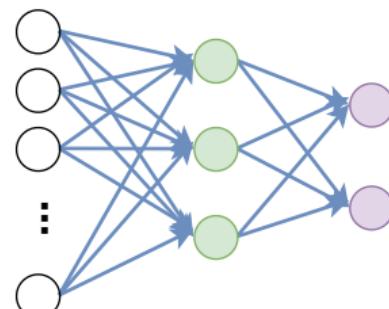
cas

$$\Rightarrow \Phi_{k_n} = -2\pi \lambda \cdot r_n \cdot \frac{dT}{dr}$$

Boundary conditions
Calibration



Data driven



Model training

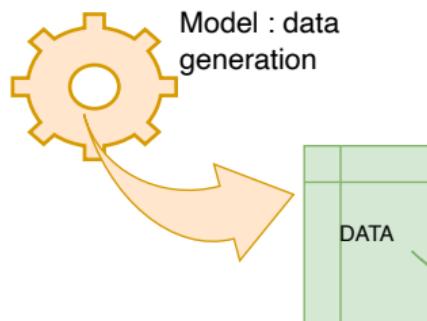




Data-driven vs Modeling

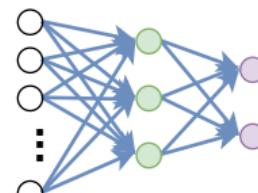
Mecanistic model / simulation

Slow / costly
Accurate



Data driven

Fast
Approximation

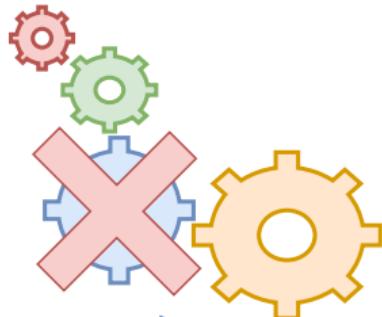




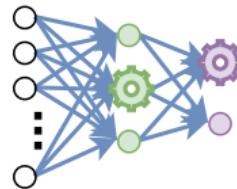
Data-driven vs Modeling

Huge composite mechanistic model

Mechanistic model / simulation



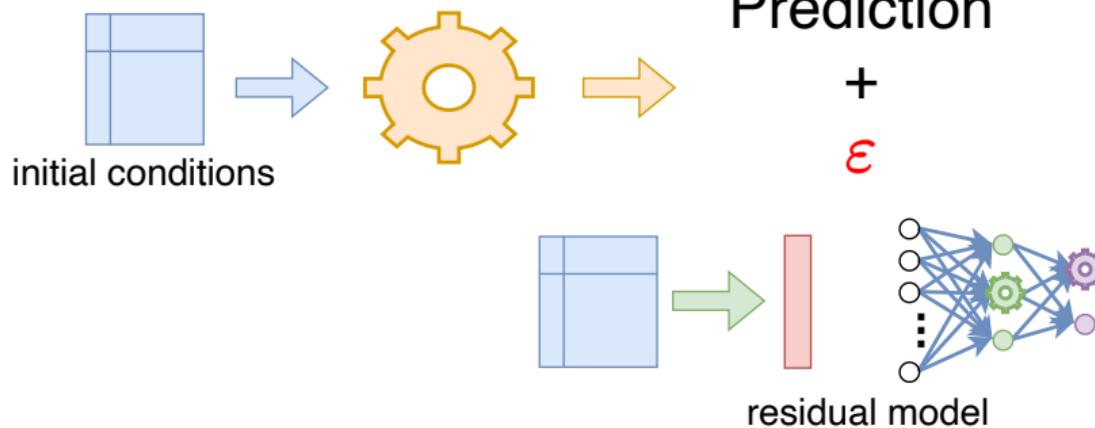
Weak component
Not enough model hypotheses



Data driven

Data-driven vs Modeling

Mecanistic model / simulation

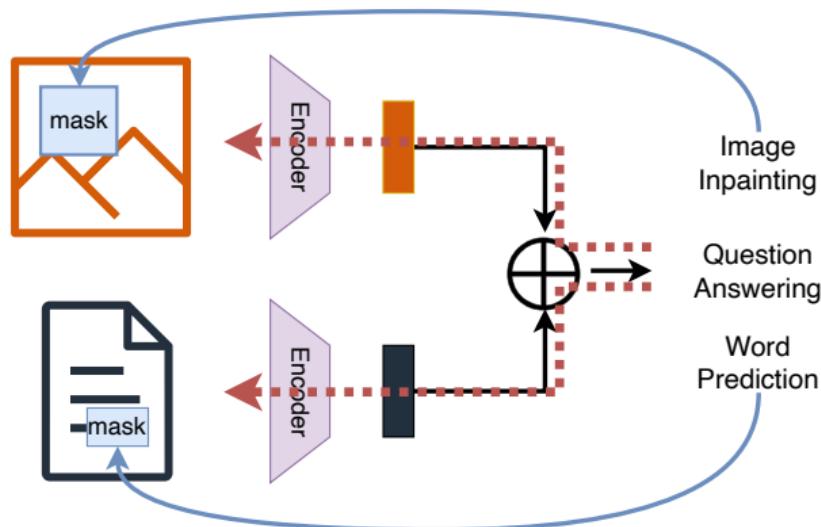


Data driven



Multi-Modality

- Construction of multimodal representation spaces = *grounding*
- Image \Rightarrow Text: *Captioning, Visual Question Answering*
- Text \Rightarrow Image: *mid-journey, dall-e, ...*

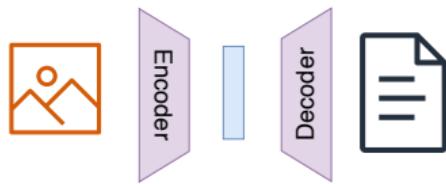


Alignment of representation spaces

Word	Teraword	Knext
Spoke	11,577,917	372,042
Laughed	3,904,519	179,395
Murdered	2,843,529	16,890
Inhaled	984,613	5,617
Breathed	725,034	41,215

Multi-Modality

- Construction of multimodal representation spaces = *grounding*
- Image ⇒ Text: *Captioning, Visual Question Answering*
- Text ⇒ Image: *mid-journey, dall-e, ...*

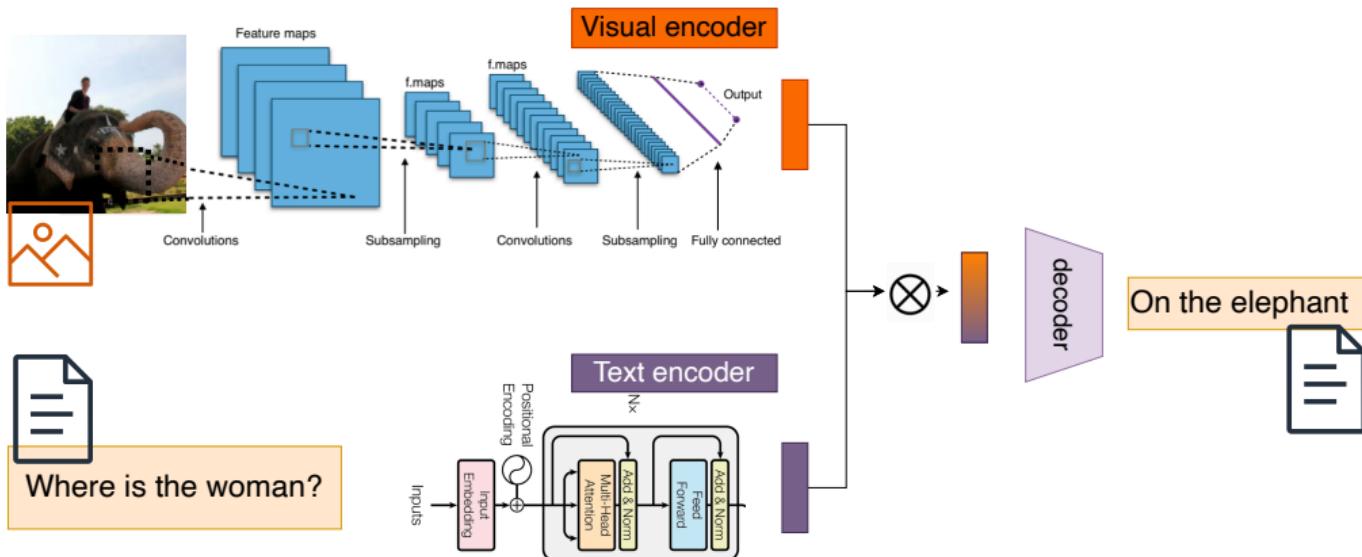


Show and Tell: image captioning open sourced in TensorFlow, Chris Shallue , Google Research, 2016
36/79



Multi-Modality

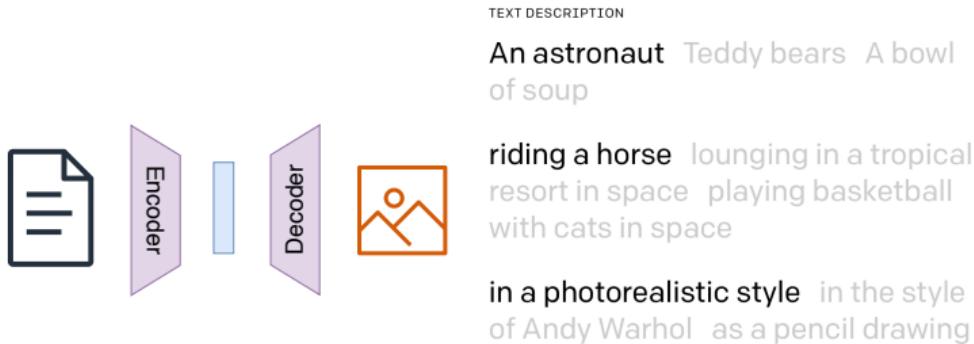
- Construction of multimodal representation spaces = *grounding*
- Image \Rightarrow Text: *Captioning, Visual Question Answering*
- Text \Rightarrow Image: *mid-journey, dall-e, ...*



Vqa: *Visual question answering*, ICCV, 2015
Antol et al.

Multi-Modality

- Construction of multimodal representation spaces = *grounding*
- Image ⇒ Text: *Captioning, Visual Question Answering*
- Text ⇒ Image: *mid-journey, dall-e, ...*



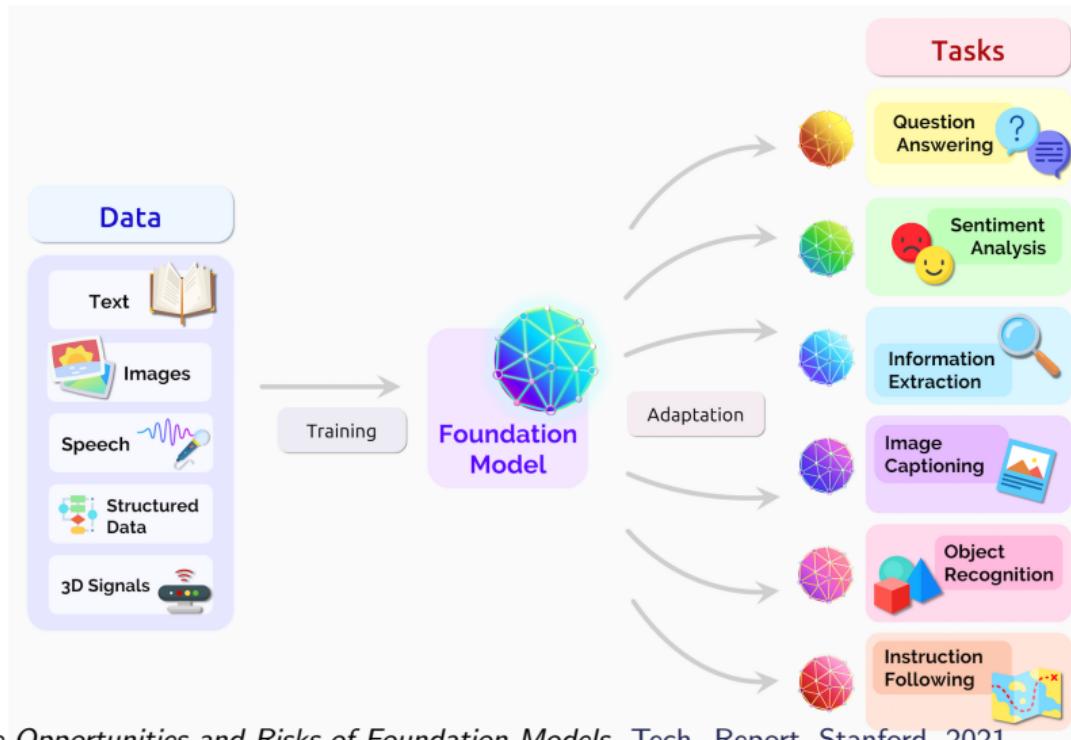
DALL-E 2





Towards Larger Foundation Models?

- Let the modalities enrich each other



On the Opportunities and Risks of Foundation Models, Tech. Report, Stanford, 2021
Bommasani et al.

Conclusion

The main challenges of multimodality

- New applications
 - at the interface between text, image, music, voice, ...
- Performance improvement
 - Better encoding, disambiguation, context encoding
- Explainability (through dialogue)
 - IoT / RecSys / Intelligent Vehicle / ...

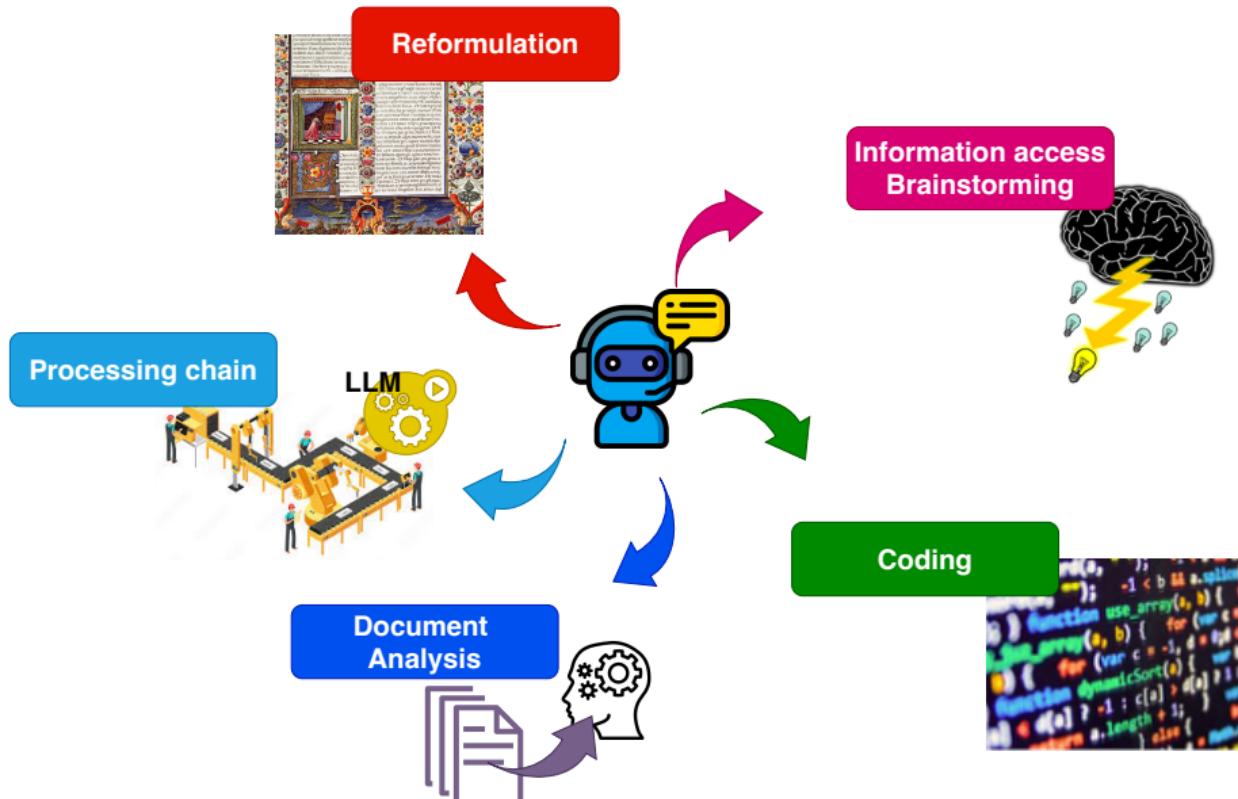


Dall-e

LARGE LANGUAGE MODELS USES

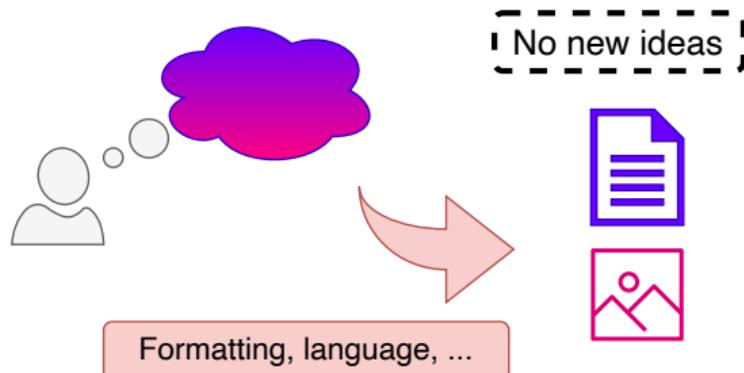


Key uses in 5 pictures



(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, ...

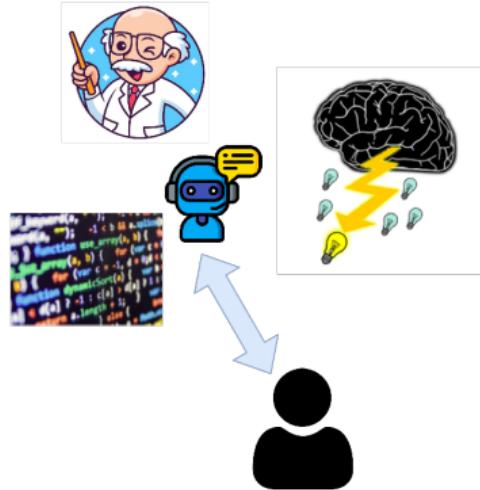
⇒ No new information, just writing, cleaning up, ...



(2) Brainstorming / Course Planning / Statistics Review

- **Find** inspiration [writer's block syndrome]
- **Organize** ideas quickly
- **Avoid omissions** / increase confidence
- **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?

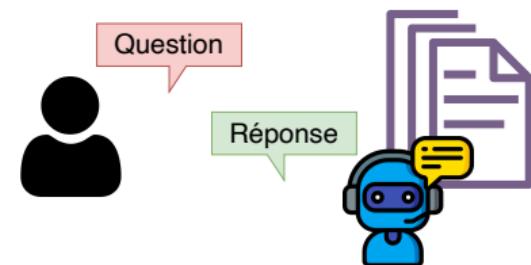
What about the carbon footprint?
- What is the risk of error propagation?

```
sentiment.ts    -∞ write.sql.go    parse_expenses.py    addresses/b
1 import datetime
2
3 def parse_expenses(expenses_string):
4     """Parse the list of expenses and return the list of triples (date,
5     Ignore lines starting with #.
6     Parse the date using datetime.
7     Example expenses_string:
8         2016-01-02 -34.01 USD
9         2016-01-03 2.59 DKK
10        2016-01-03 -2.72 EUR
11
12    expenses = []
13    for line in expenses_string.splitlines():
14        if line.startswith("#"):
15            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



Wi-Fi NotebookLM

**Think Smarter,
Not Harder**

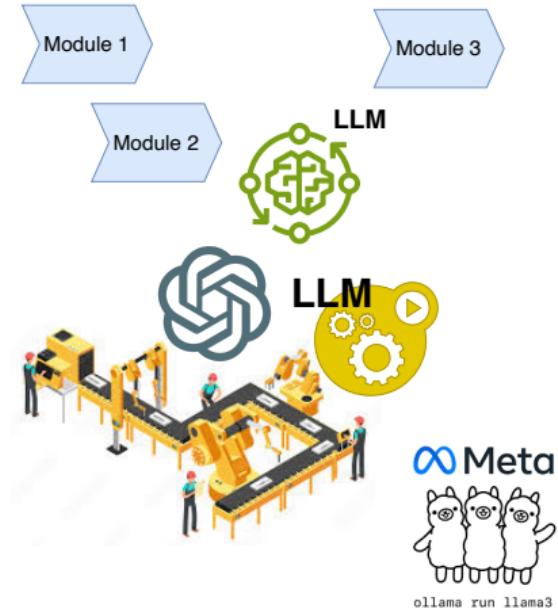
Try NotebookLM

- 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
 - 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**

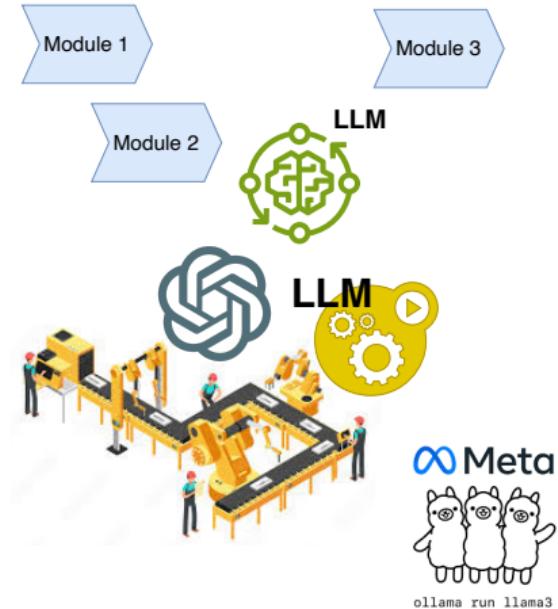


- Can I train models on generated data?
- How much does it cost? (\$ + CO₂) Need for GPUs?
- How good are open-weight models?



(5) LLM in a Production Pipeline / Agentic AI

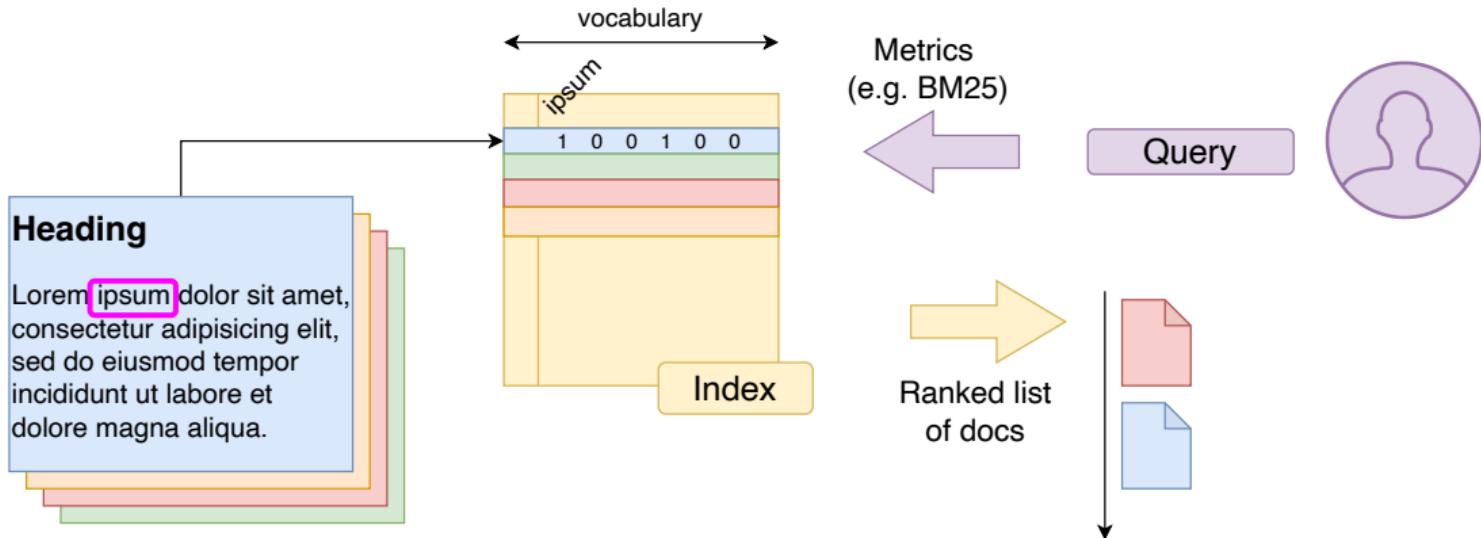
- Run LLM locally
 - Extract knowledge
 - Sort documents / generate summaries
 - Generate examples to train a model
[Teacher/student - distillation]
 - Generate variants of examples ↗↗ increase dataset size
[Data augmentation]
- ⇒ Integrate the LLM into a processing pipeline
= little/less supervision = **Agentic AI**



- Can I train models on generated data?
- How much does it cost? (\$ + CO₂) Need for GPUs?
- How good are open-weight models?

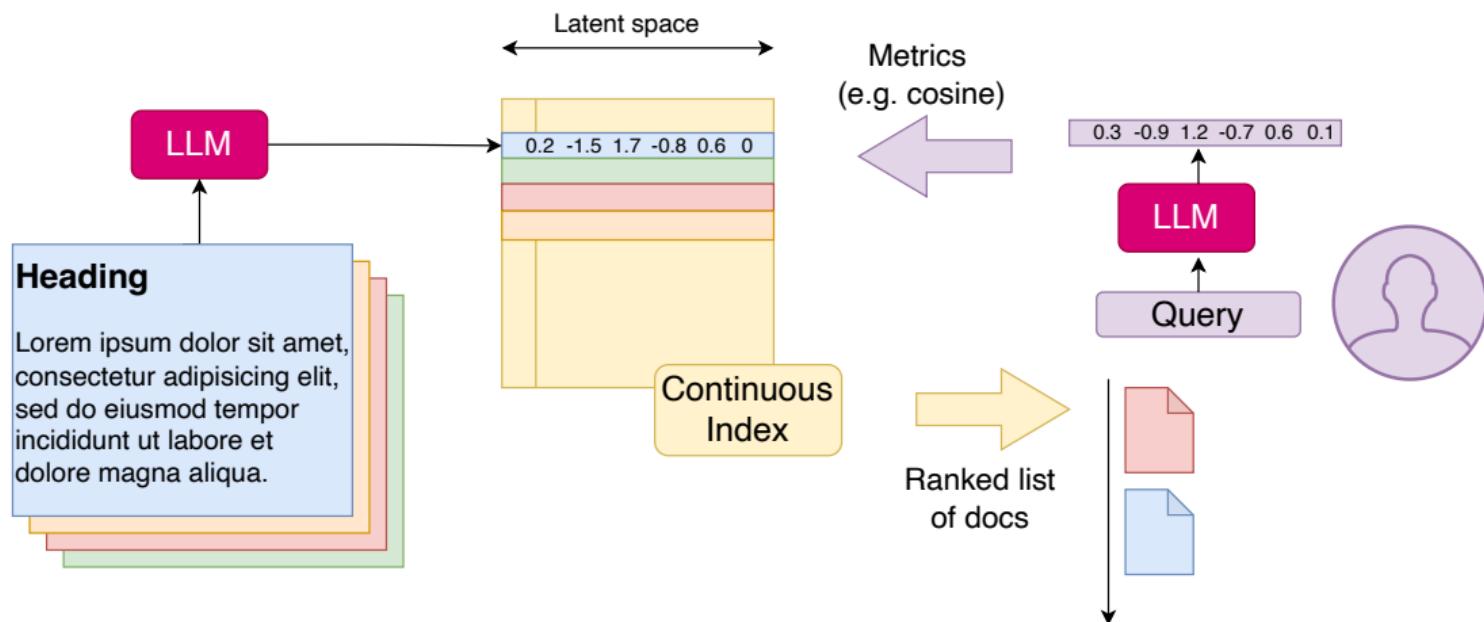


LLM vs Information Retrieval



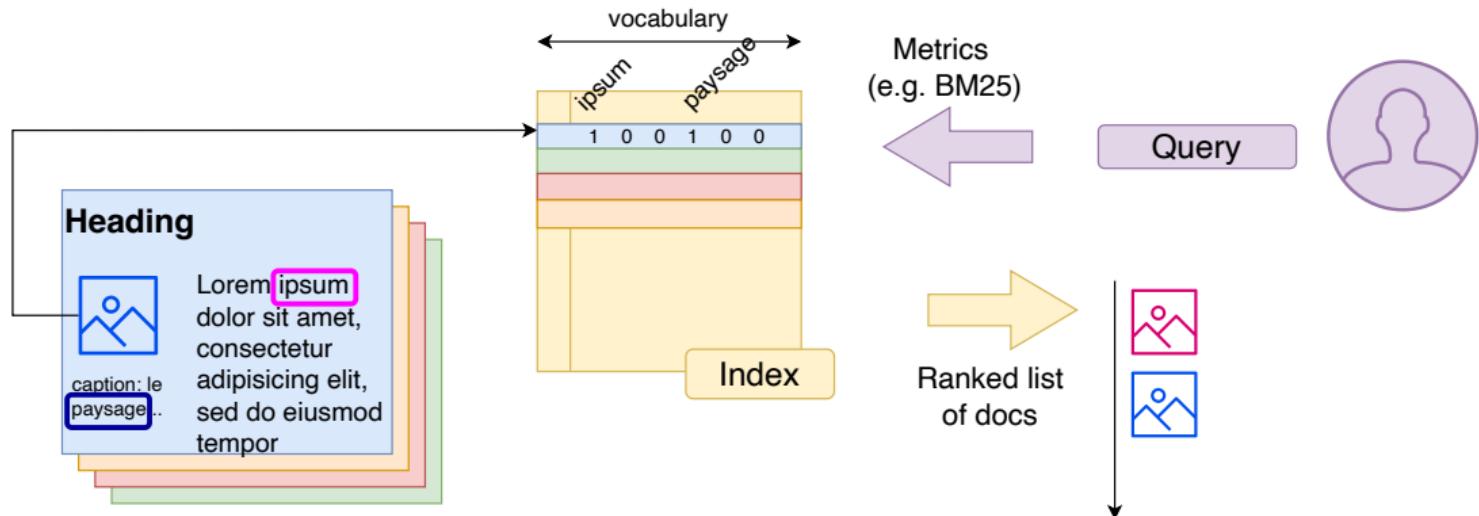


LLM vs Information Retrieval



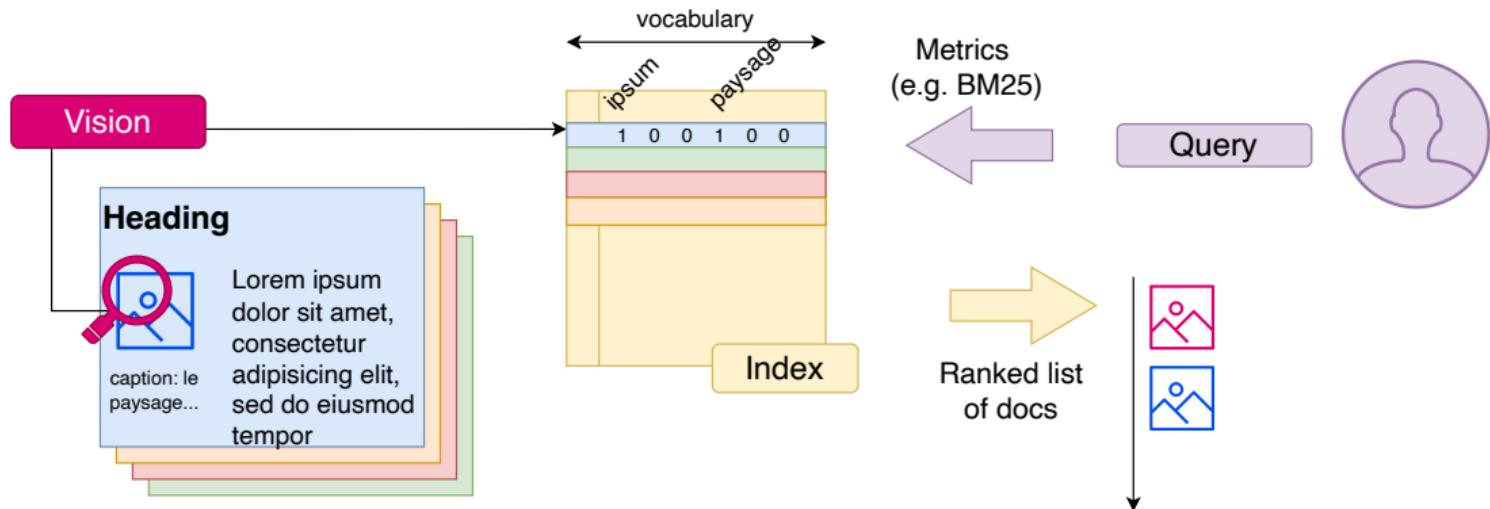


LLM vs Information Retrieval



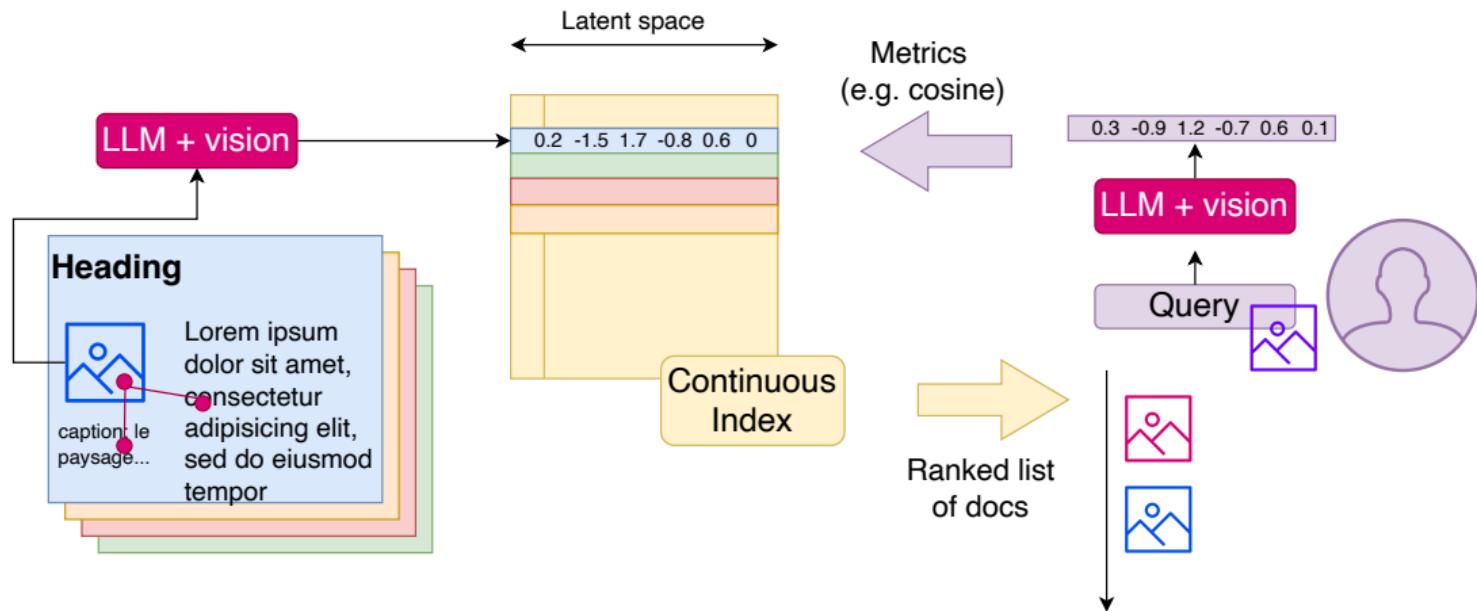


LLM vs Information Retrieval



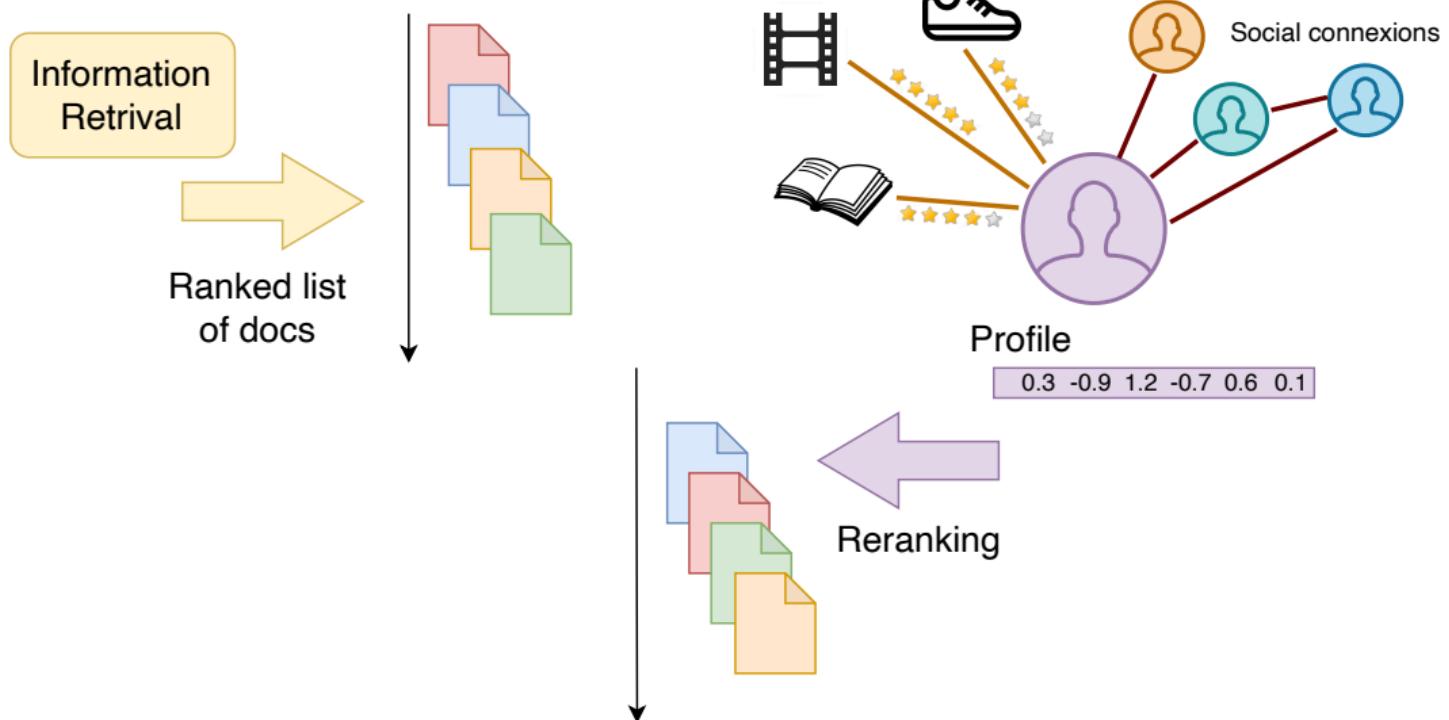


LLM vs Information Retrieval





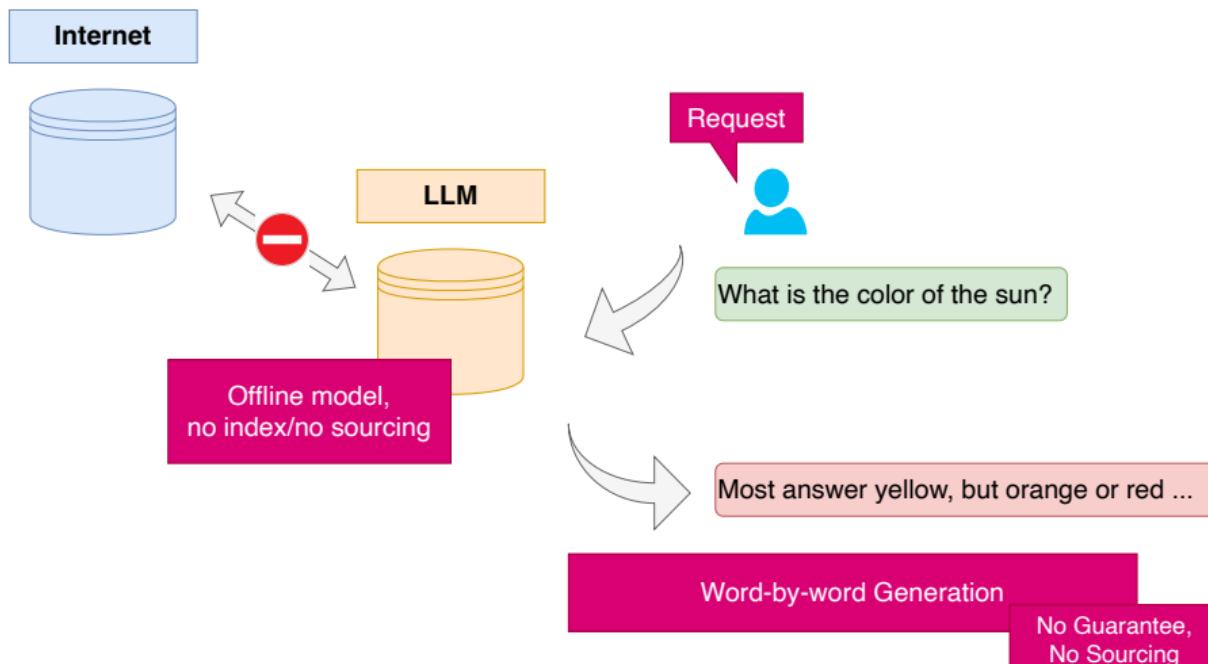
LLM vs Information Retrieval





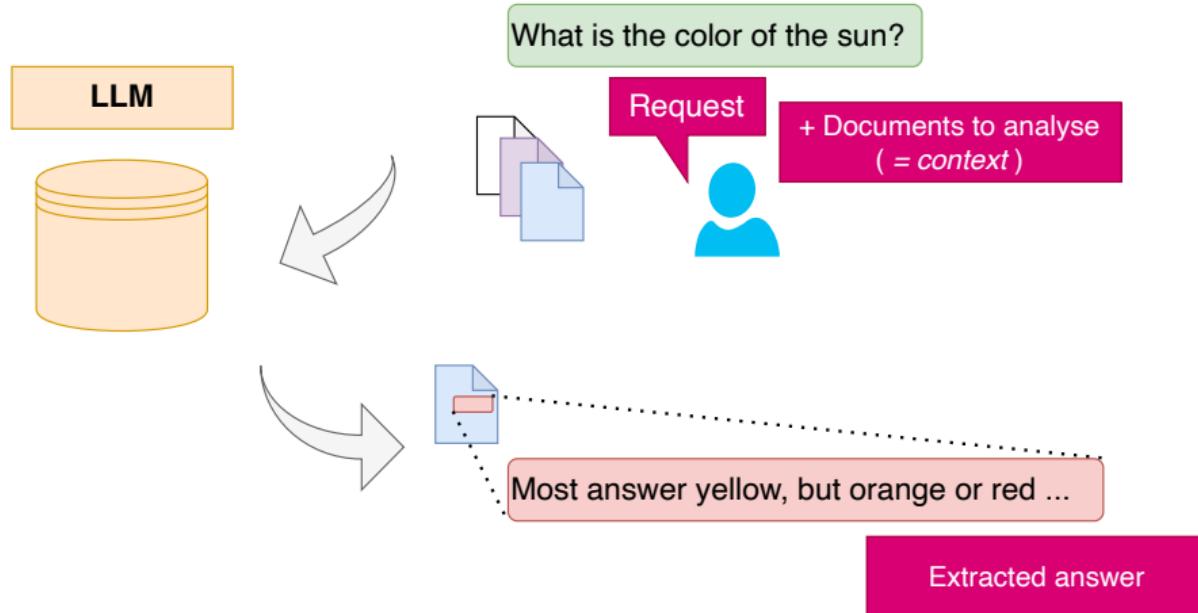
LLMs \Rightarrow RAG : parametric memory vs Info. Extraction

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





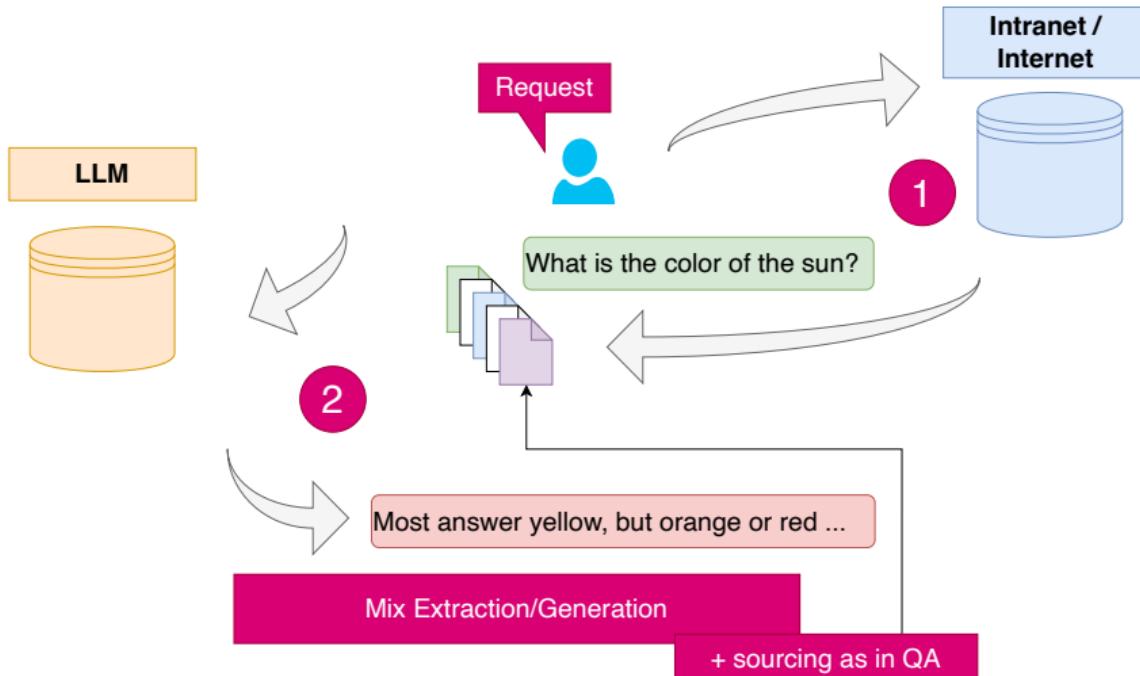
LLMs \Rightarrow RAG : parametric memory vs Info. Extraction



- 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 \Rightarrow RAG : parametric memory vs Info. Extraction



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

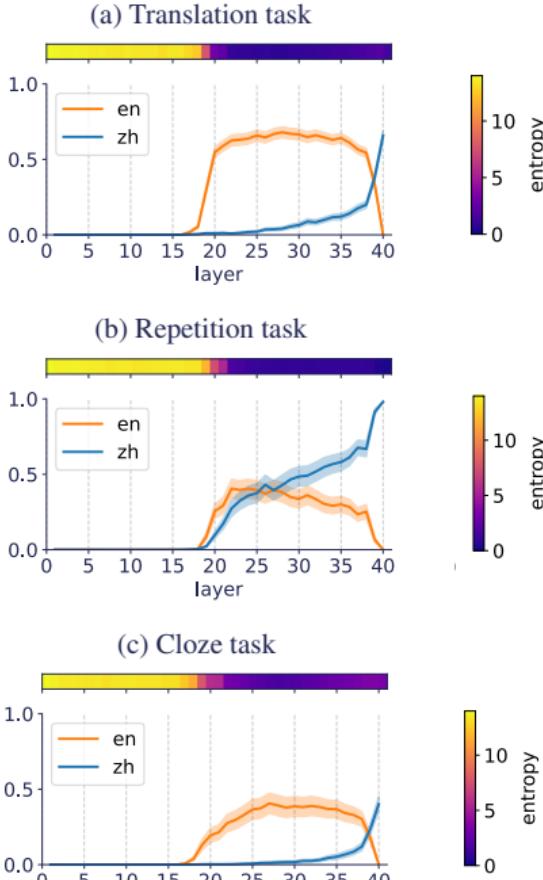


Language Handling

- Language models are (mostly) multilingual:

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

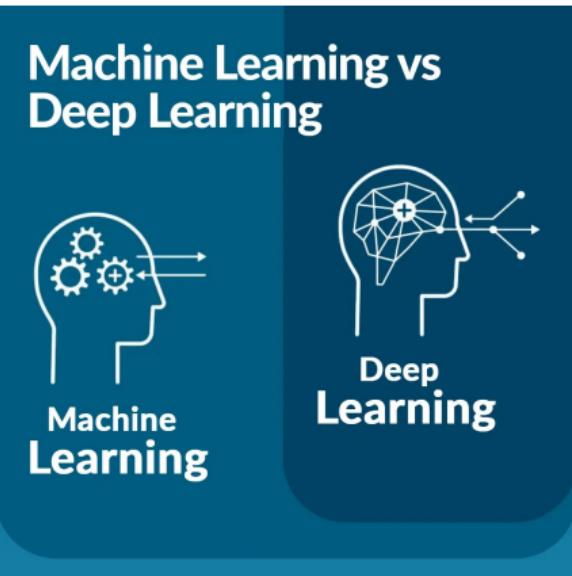
[Wendler et al. 2024] Do Llamas Work in English?
On the Latent Language of Multilingual Transformers





IA générative vs IA classique

- Prédiction de séries temporelles, maintenance prédictive
- Prédiction des prix (voitures, immobilier, ...)
- Diagnostic médical sur des données numériques, EEG, ECG, ...
- Systèmes de recommandation

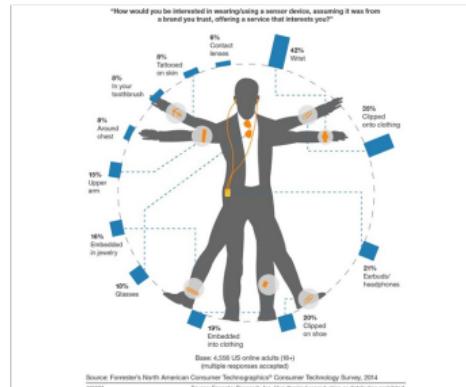


Why do tree-based models still outperform deep learning on typical tabular data?

L Grinsztajn, E Oyallon, G Varoquaux, NeurIPS 22

Des très nombreuses application d'IA embarquée

1 Bracelet connecté, vêtements, lunettes



Smart
wristbands



Smart
clothing



mart
watch



Head-mounted devices

- Séries temporelles, diagnostic, recherche d'anomalie
 - Médecine ou gadget?



Des très nombreuses application d'IA embarquée

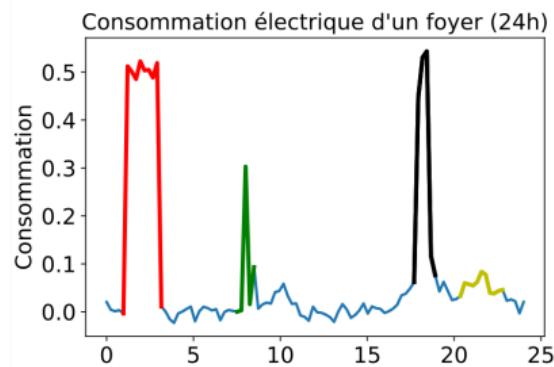
- 1 Bracelet connecté, vêtements, lunettes
- 2 Assistant intelligent, Chatbot





Des très nombreuses application d'IA embarquée

- 1 Bracelet connecté, vêtements, lunettes
- 2 Assistant intelligent, Chatbot
- 3 Compteur intelligent (e.g. Linky)





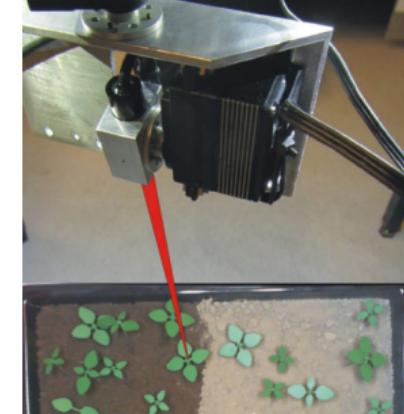
Des très nombreuses application d'IA embarquée

- 1 Bracelet connecté, vêtements, lunettes
- 2 Assistant intelligent, Chatbot
- 3 Compteur intelligent (e.g. Linky)
- 4 Cabine télémédecine



Des très nombreuses application d'IA embarquée

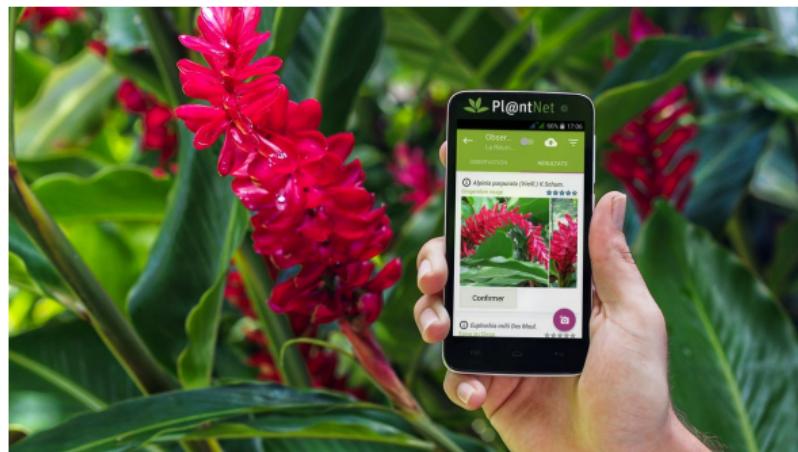
- 1 Bracelet connecté, vêtements, lunettes
- 2 Assistant intelligent, Chatbot
- 3 Compteur intelligent (e.g. Linky)
- 4 Cabine télémédecine
- 5 Robotique



A

Des très nombreuses application d'IA embarquée

- 1 Bracelet connecté, vêtements, lunettes
 - 2 Assistant intelligent, Chatbot
 - 3 Compteur intelligent (e.g. Linky)
 - 4 Cabine télémédecine
 - 5 Robotique
 - 6 ... Et plein d'autres choses !
Smartphone?



MACHINE LEARNING LIMITS



chatGPT and the relationship with truth

1 **Likelihood** = grammar, agreement, tense concordance, logical sequences...
 ⇒ Repeated knowledge

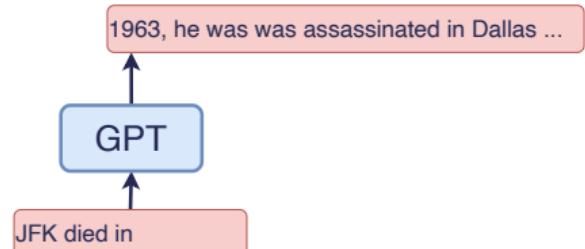
2 Predict the most **plausible** word...
 ⇒ produces **hallucinations**

3 **Offline** functioning

4 chatGPT ≠ **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?

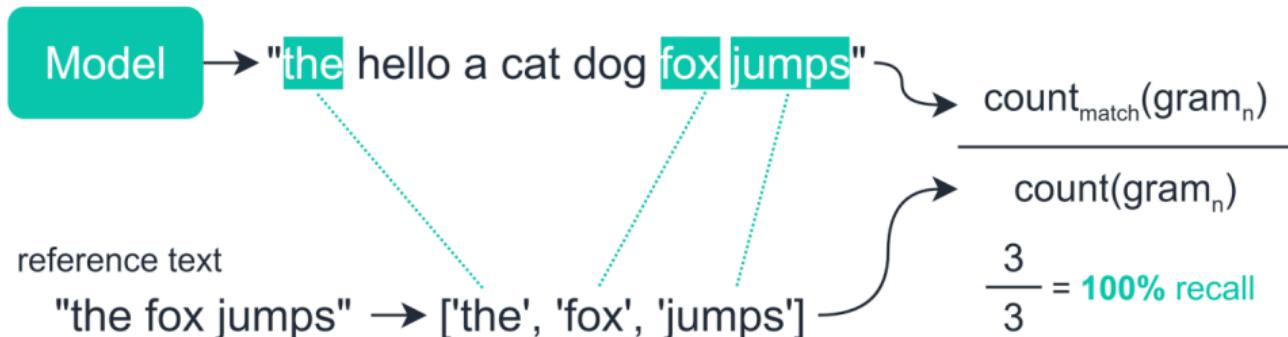
- recurrents.
2. "Variational Recurrent Auto-encoders" par Chung et al. (2016) - Cette étude introduit une approche VAE pour la modélisation de séquences temporelles en utilisant des réseaux de neurones récurrents et une méthode de maximisation de la vraisemblance pour la phase d'entraînement. [Variational recurrent auto-encoders](#)
 3. "Generative Modeling for Time Series with Variational Auto-Encoder" par Bao et al. (2017) - Cette étude propose un modèle qui combine les forces des RNNs et SGVB: le Variational Recurrent Auto-Encoder (VRAE). Such a model can be used for efficient, large scale ...
 4. "Deep Variational Bayes Filters: Unsupervised Learning of State Space Models from ..."
- [Enregistrer](#) [Cler](#) [Cité 302 fois](#) [Autres articles](#) [Les 2 versions](#) [PDF](#)



Generative AI: how to evaluate performance?

The critical point today

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

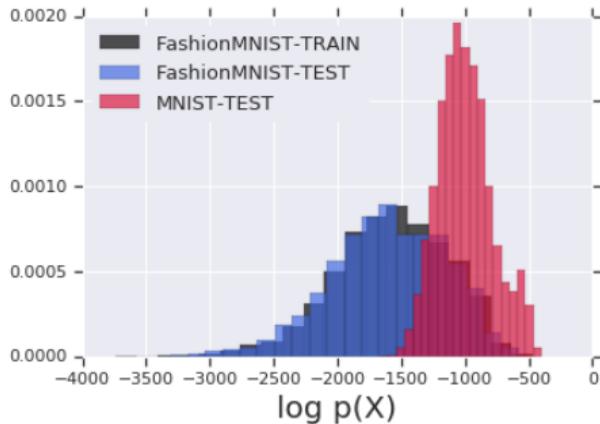




Generative AI: how to evaluate performance?

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



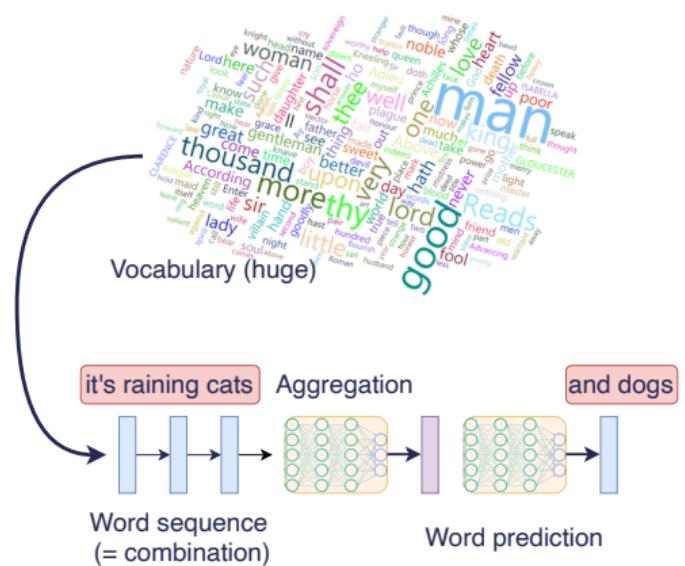
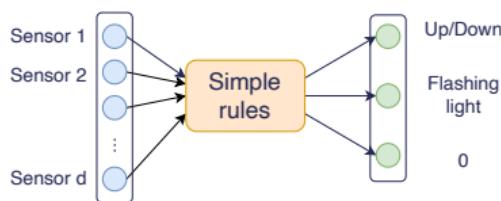
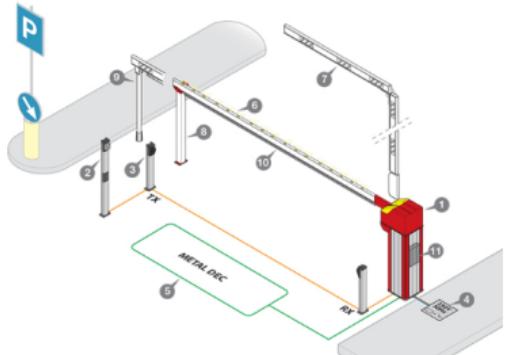
V how old is obama?
==

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

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 : 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

		Meta	BigScience	OpenAI	stability.ai	Google	ANTHROPiC	cohere	AI21labs	Inflection	amazon	Average	
		Llama 2	BLOOMZ	GPT-4	Stable Diffusion 2	PaLM 2	Claude 2	Command	Jurassic-2	Inflection-1	Titan Text		
Major Dimensions of Transparency	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%	0%	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%	

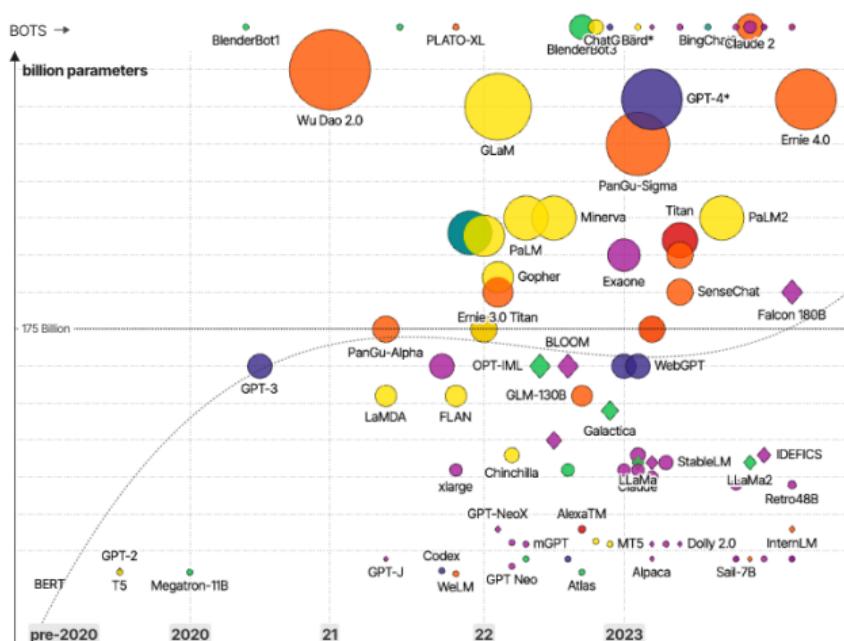


Costs / Frugality

The Rise and Rise of A.I.

Large Language Models (LLMs) & their associated bots like ChatGPT

● Amazon-owned ● Chinese ● Google ● Meta / Facebook ● Microsoft ● OpenAI ● Other

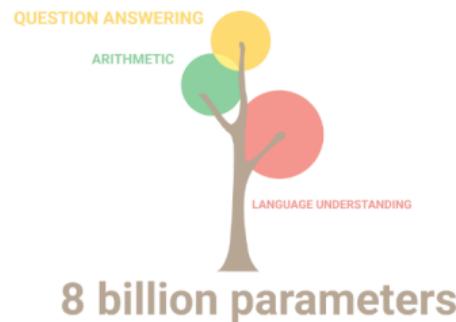


Parameters

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

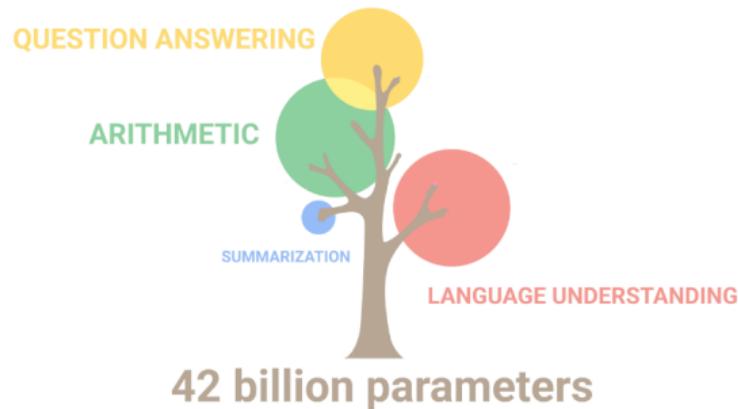
Costs / Frugality

Emergent Capabilities



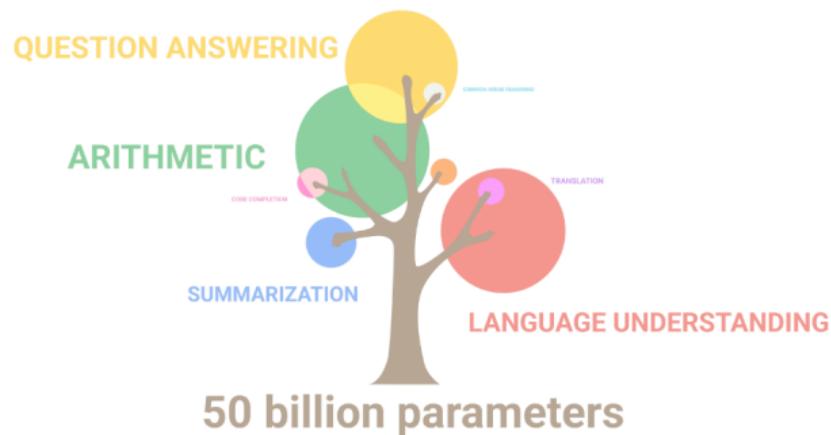
Costs / Frugality

Emergent Capabilities



Costs / Frugality

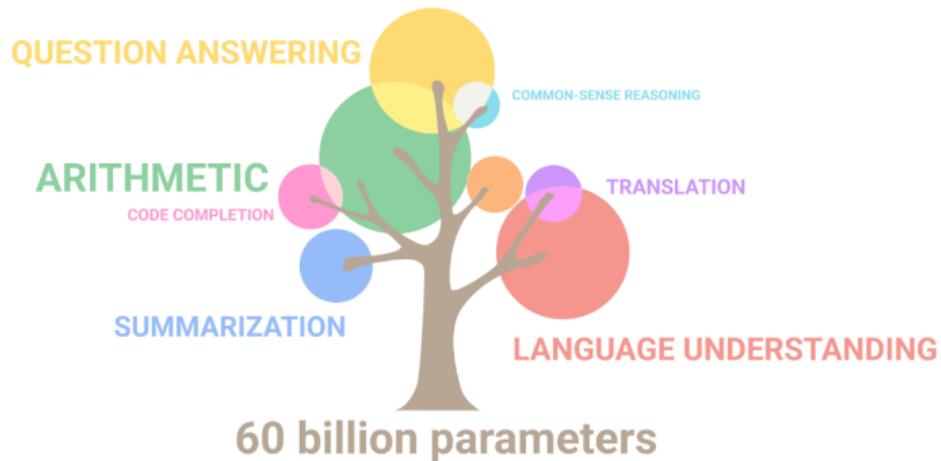
Emergent Capabilities





Costs / Frugality

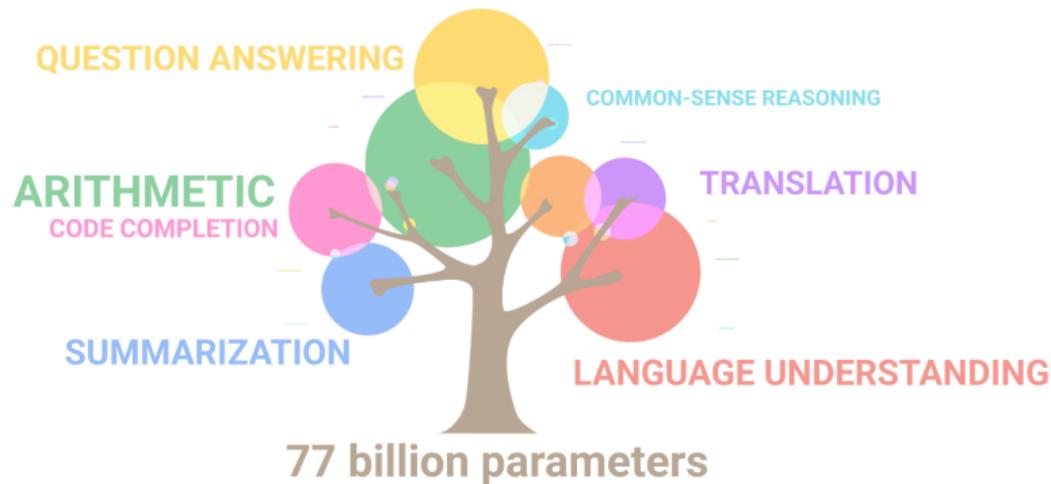
Emergent Capabilities





Costs / Frugality

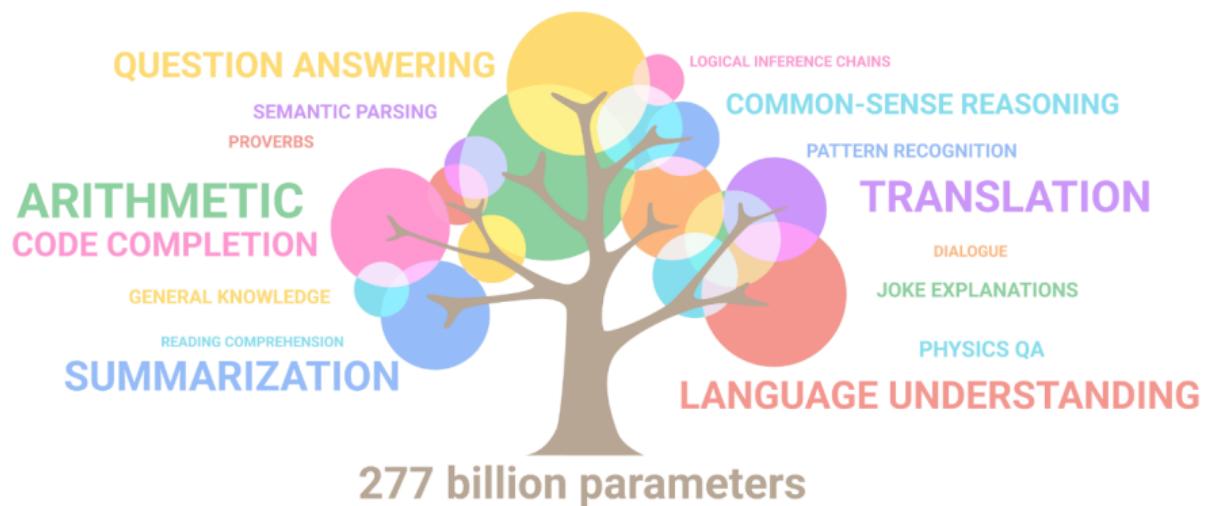
Emergent Capabilities





Costs / Frugality

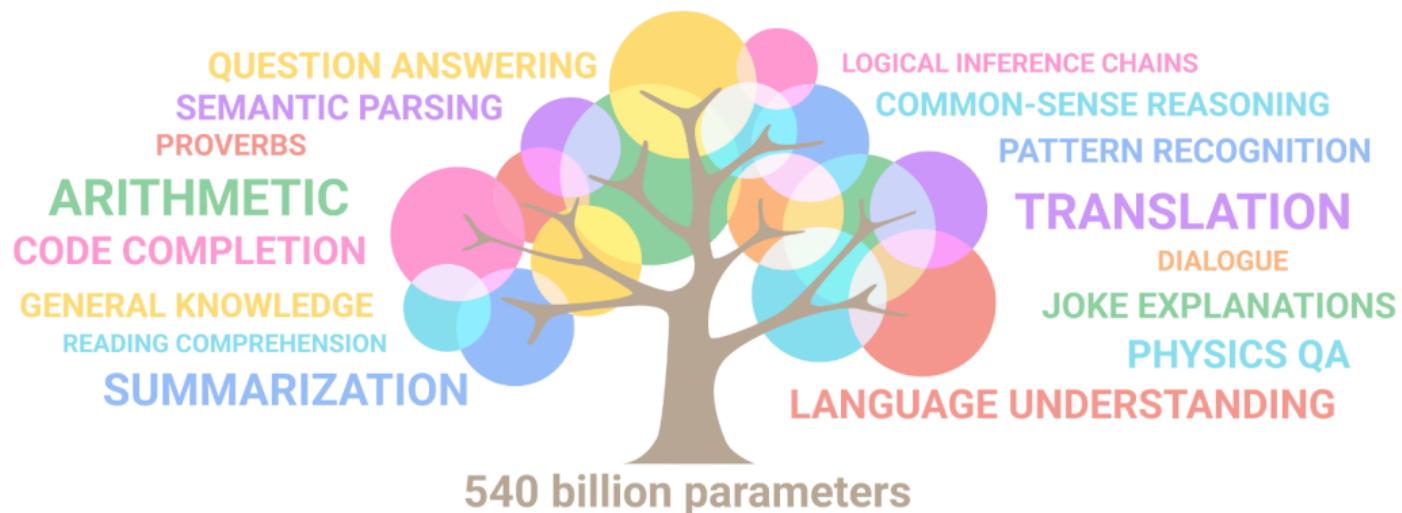
Emergent Capabilities





Costs / Frugality

Emergent Capabilities



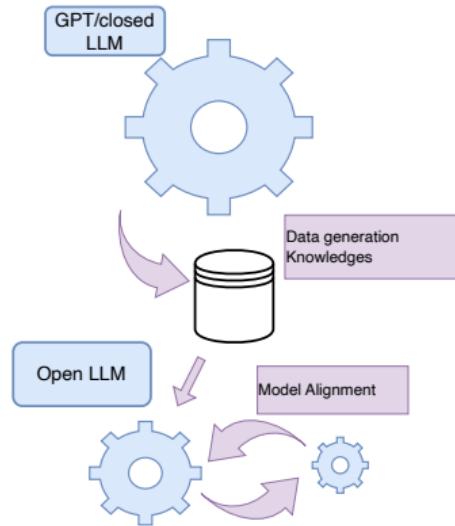


LLMs & Frugality

Distillation

Pruning Quantization

Mixture of Experts

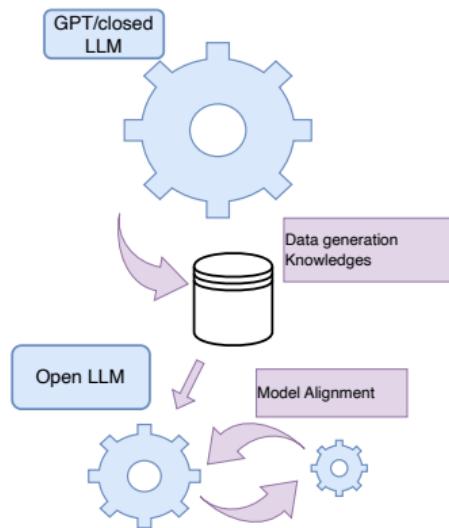


Frugality... Model size **x1000** in 3y... Then **optimization x1/100** in 2y

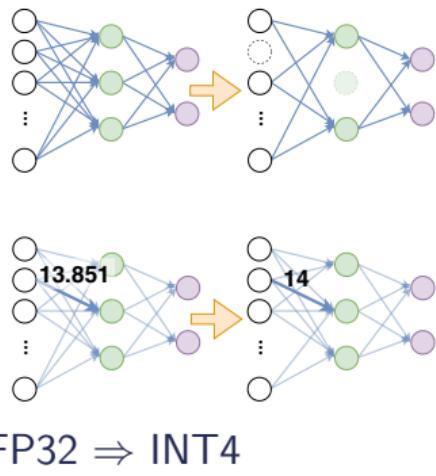


LLMs & Frugality

Distillation



Pruning Quantization



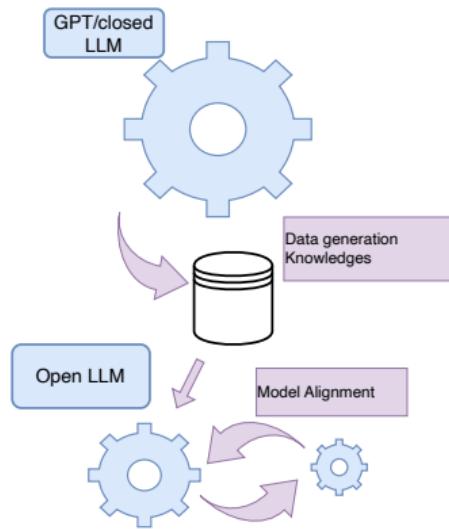
Mixture of Experts

Frugality... Model size **x1000** in 3y... Then **optimization x1/100** in 2y

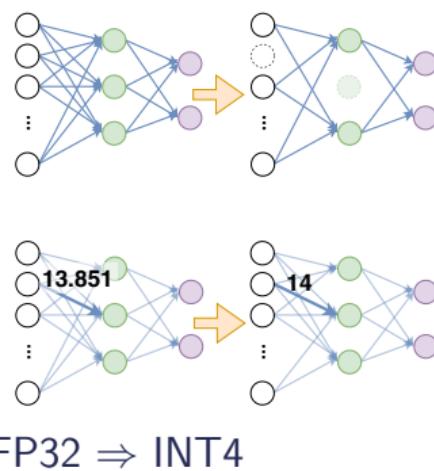


LLMs & Frugality

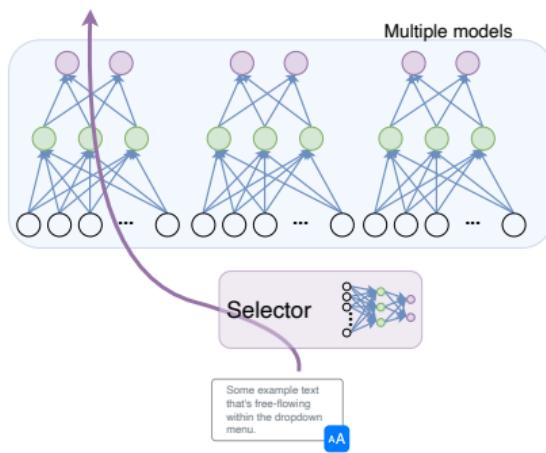
Distillation



Pruning Quantization



Mixture of Experts



+ Code industrialization

Frugality... Model size **x1000** in 3y... Then **optimization x1/100 in 2y**



Different behaviors, different costs

Les IA sont démasquées !

Mistral/Minstral

SEMI-OUVERT 8 MDS DE PARAMÈTRES SORTIE 10/2024

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

Impact énergétique de la discussion

$$\begin{matrix} \text{8 milliards param.} \\ \text{taille du modèle} \end{matrix} \times \begin{matrix} \text{128 tokens} \\ \text{taille du texte} \end{matrix} = \begin{matrix} \text{0.30 Wh} \\ \text{énergie consu.} \end{matrix}$$

Ce qui correspond à :

0.30g
CO₂ émis

5min
ampoule LED

33s
vidéos en ligne

Voir plus

DeepSeek/DeepSeek v3

SEMI-OUVERT 671 MDS DE PARAMÈTRES SORTIE 12/2024

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

Impact énergétique de la discussion

$$\begin{matrix} \text{671 milliards param.} \\ \text{taille du modèle} \end{matrix} \times \begin{matrix} \text{225 tokens} \\ \text{taille du texte} \end{matrix} = \begin{matrix} \text{6Wh} \\ \text{énergie consu.} \end{matrix}$$

Ce qui correspond à :

6g
CO₂ émis

2h
ampoule LED

12min
vidéos en ligne

Voir plus



Different behaviors, different costs

Different costs for different users/languages

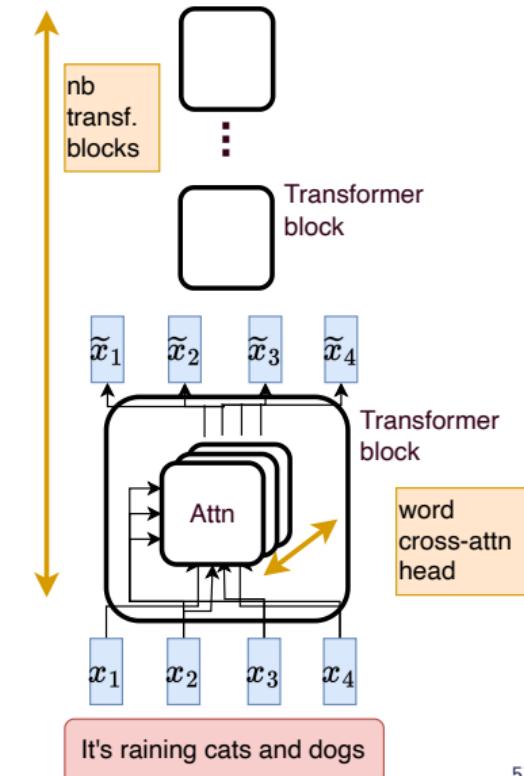
Pour un texte significatif en Français

and the same in English

TOKENS CHARACTERS
17 63

<|s> Pour un texte significatif en Français

and the same in English





Different behaviors, different costs

Different costs for different users/languages

The Tokenizer Playground

Experiment with different tokenizers (running locally in your browser).

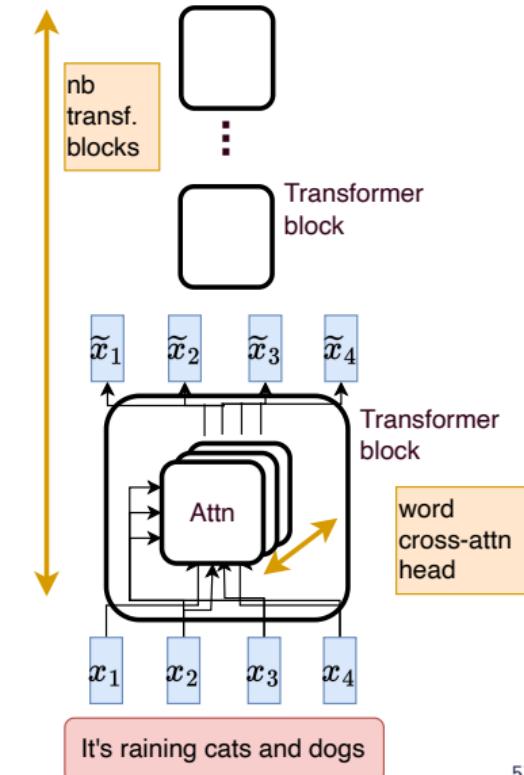
gpt-4 / gpt-3.5-turbo / text-embedding-ada-002 ▾

124578 * 963

TOKENS CHARACTERS
5 12

124578 * 963

● Text ○ Token IDs ○ Hide

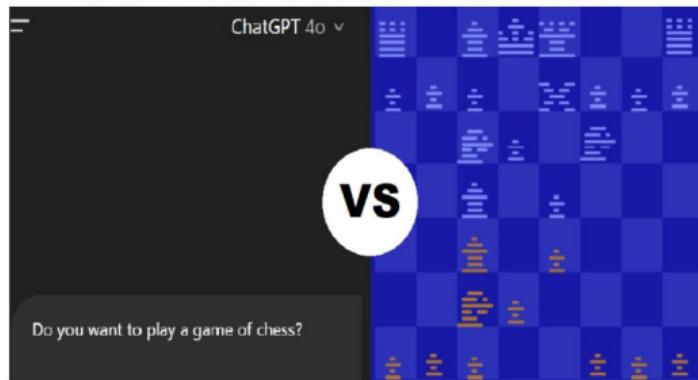




Everything beyond the LLM's capabilities/training

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

ATARI 2600 SCORES STUNNING VICTORY OVER CHATGPT



WHEN YOU UNDERESTIMATE A 1977 CHESS ENGINE... AND IT HUMBLES YOU IN FRONT OF THE WHOLE INTERNET

(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.

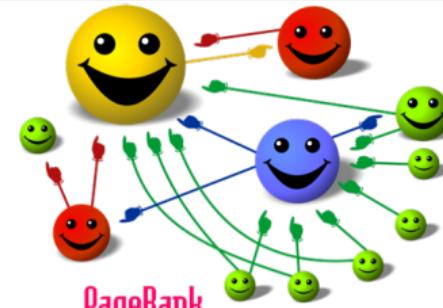


Automation, access and environmental harms

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 \Rightarrow bias in the responses

Machine learning is based on extracting statistical biases...

\Rightarrow Fighting bias = manually adjusting the algorithm

Machine Learning & Bias



Stereotypes from *Pleated Jeans*

A screenshot of the Google Translate web interface. At the top, there's a navigation bar with a menu icon, the "Google Traduction" logo, and a settings gear icon. Below the navigation bar are four tabs: "Texte" (selected), "Images", "Documents", and "Sites Web". The main area shows a translation from English to French. On the left, the English sentence "The nurse and the doctor" is inputted. On the right, the translated French sentence "L'infirmière et le médecin" is displayed. There are dropdown menus for source ("Anglais") and target ("Français") languages, and a dropdown for "Detecter la langue" (Detect language). A green circular button with a white checkmark is visible in the top right corner.

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

Bias in the data \Rightarrow 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

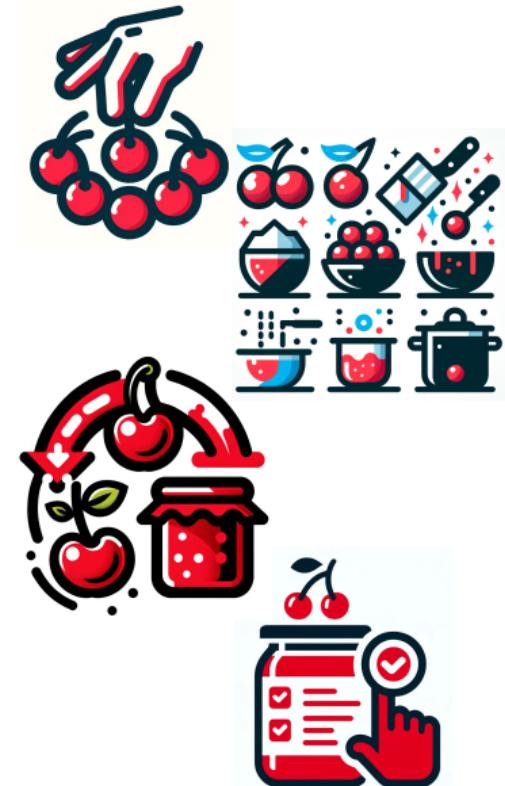
3 Prior knowledge

- Balance, loss, a priori, operator choices...

4 Output filtering

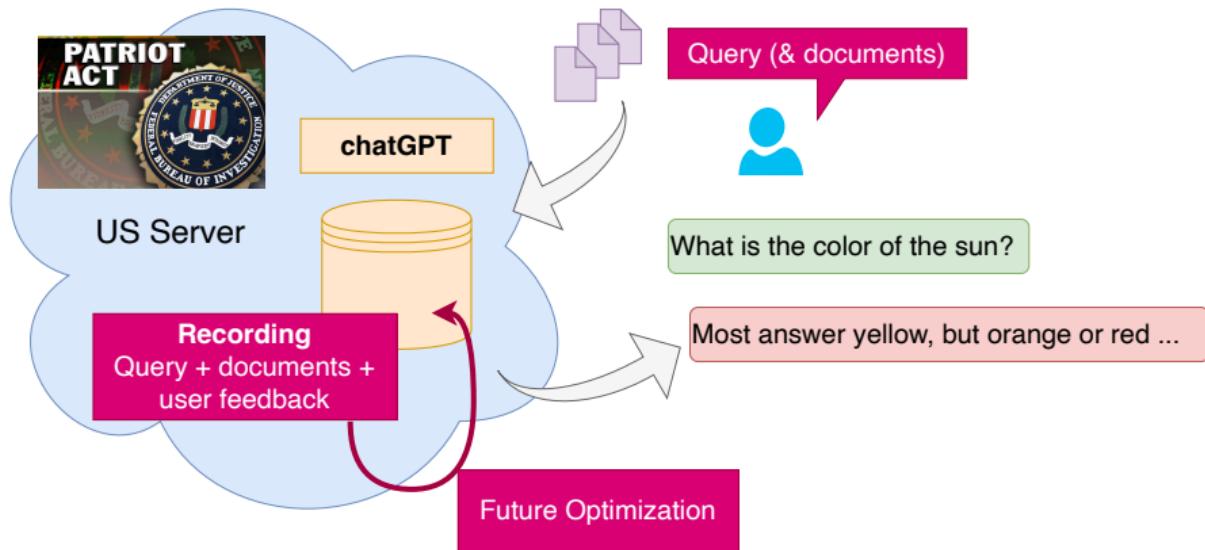
- Post processing
- Censorship, redirection, ...

⇒ Choices that influence algorithm results





Data Leak(s): different security levels



- 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

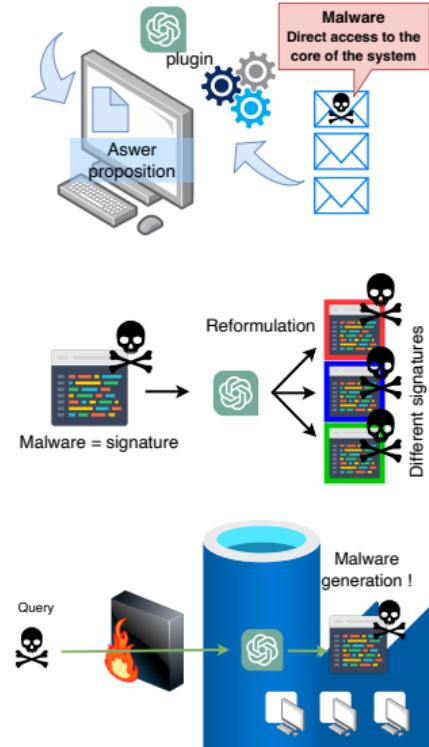
Use of a locally operated LLM, with no data transferred over the web.

HuggingFace, Ollama, ...



Security Issues

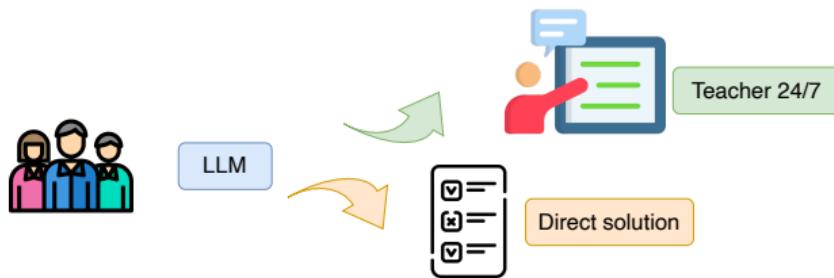
- Plug-ins ⇒ Often significant security vulnerabilities for users
 - Email access / transfer of sensitive information etc...
- Management issues for companies
 - Securing (very) large files
- Increased opportunities for malware signatures
 - ≈ software rephrasing
- New problems!
 - Direct malware generation



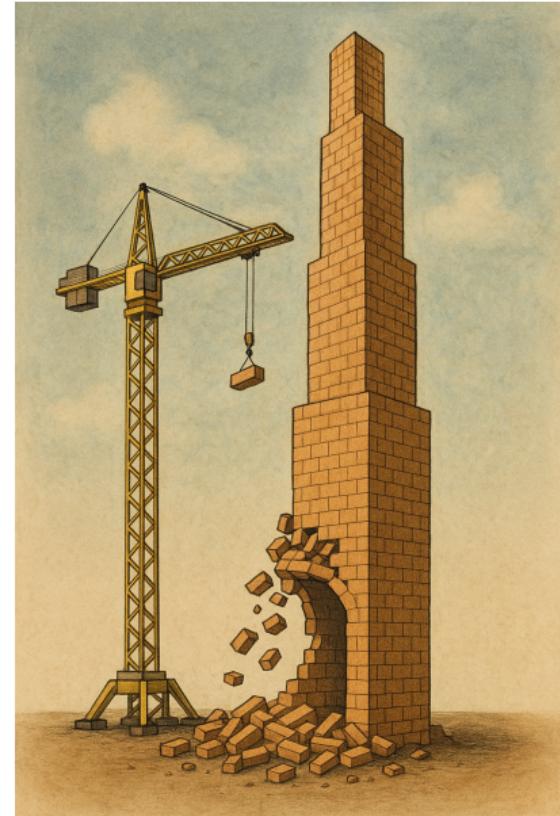


Educational Challenges

- 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



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



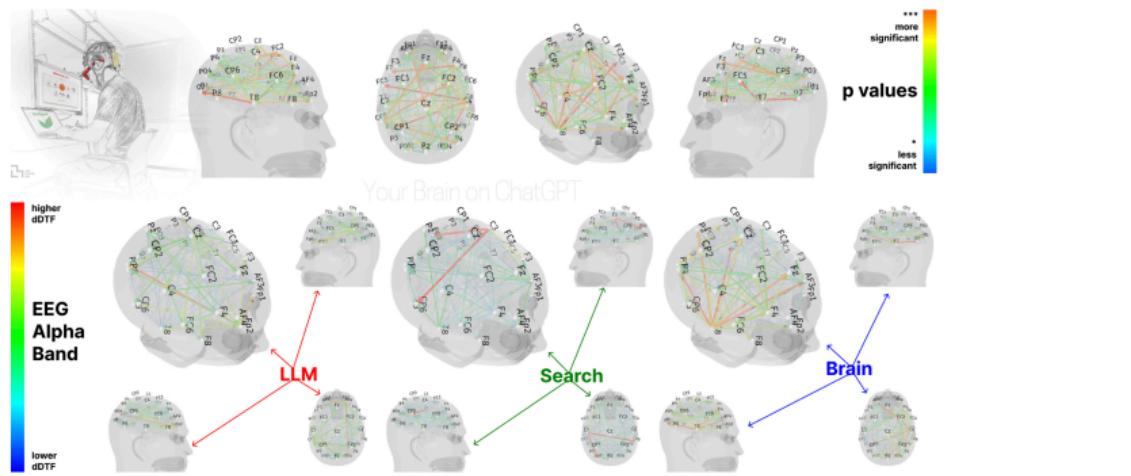


Decline / Evolution of Cognitive skills

Our brain will evolve with these new tools...

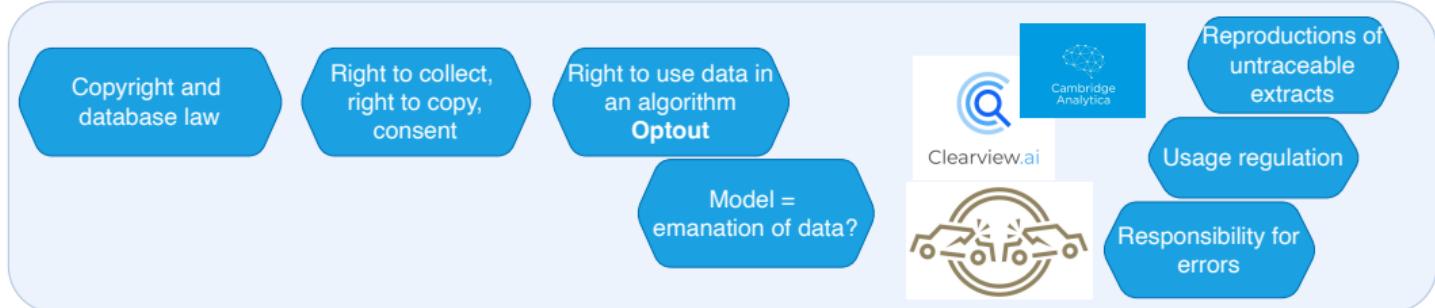
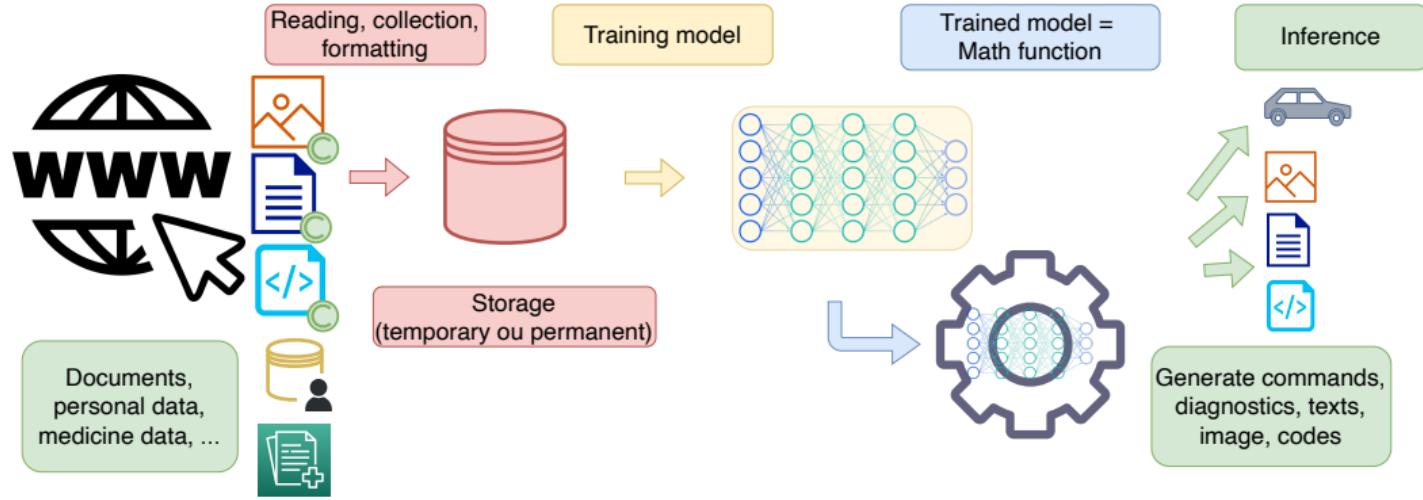
What is the scope of these transformations? What will be the consequences?

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





Legal Risks/Questions



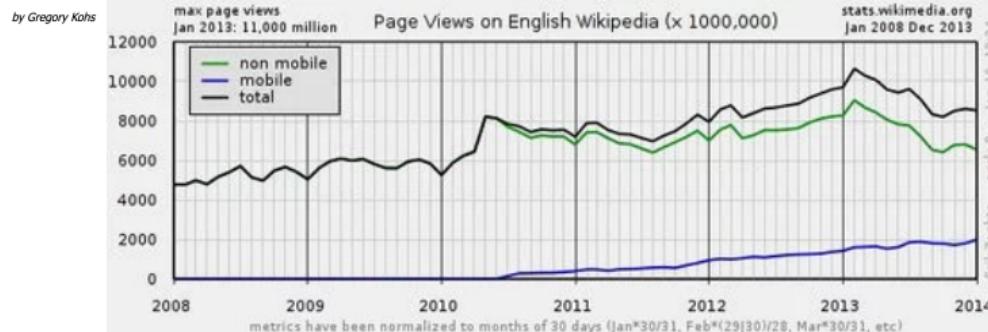


Economic Questions

- 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?



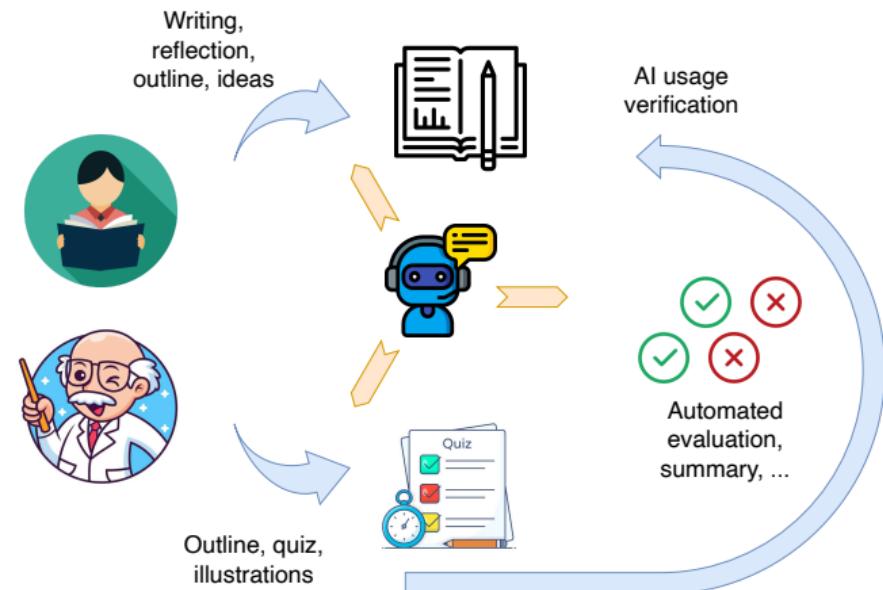
\Rightarrow Who does benefit from the feedback? [StackOverFlow]



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 **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?

- 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?



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

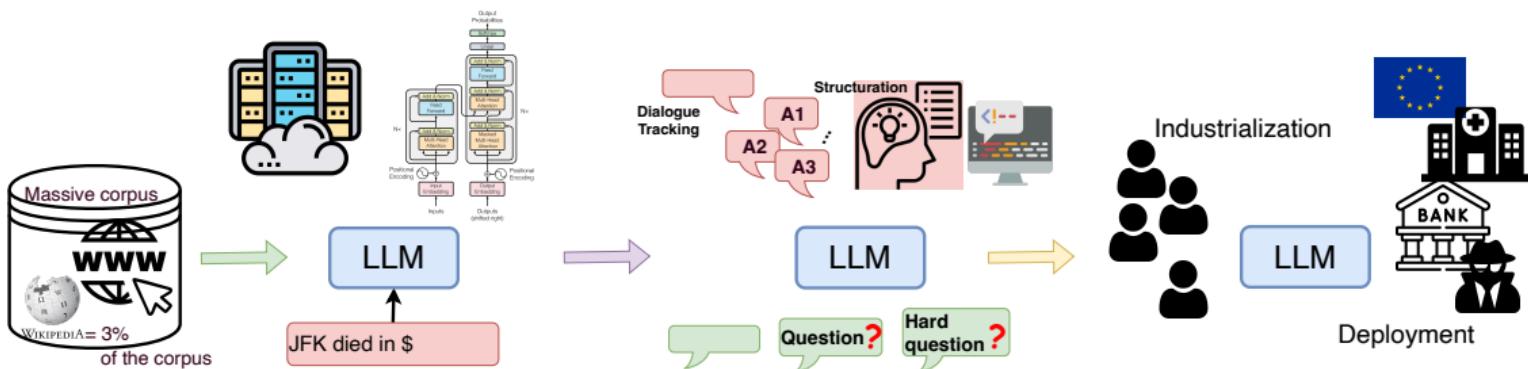
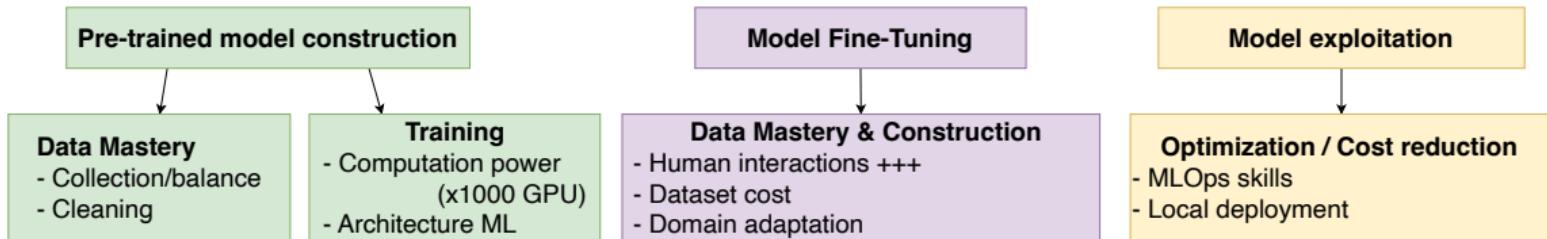
- **Environments:** Jupyter, Visual Studio Code, ...
 - **Machine Learning** Scikit-Learn: blocks to assemble
 - Training: 1 week
 - Project completion: few hours to few days
 - **Deep Learning** pytorch, tensorflow: building blocks... but more complex
 - Training: 2-5 weeks
 - Project completion: few 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*
 - 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: ≥ 1 year

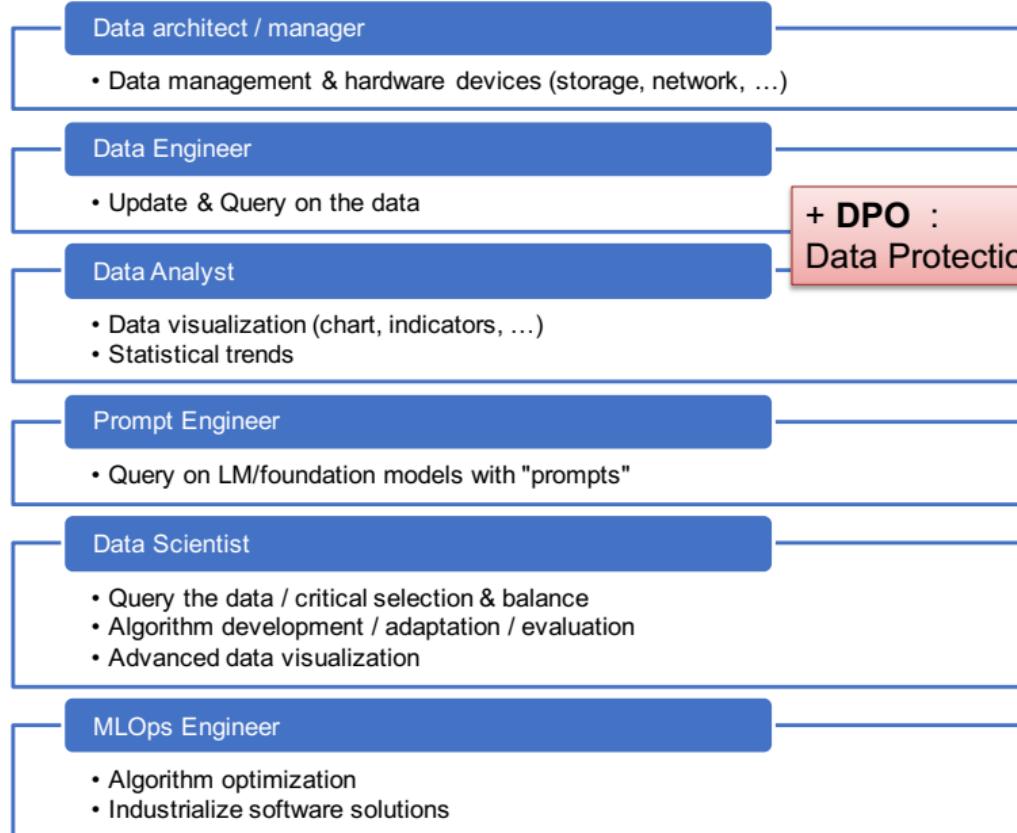


Digital Sovereignty: the Entire Chain





A Multitude of Professions



+ DPO :
Data Protection Officer





Factors of Acceptability for Generative AI

1 Utilitarianism:

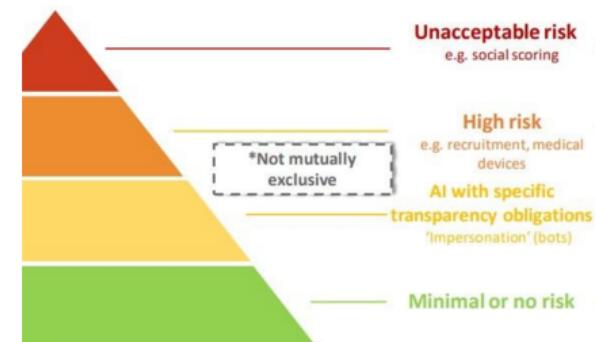
- Performance (acceptance factor of chatGPT)
- Reliability / Self-assessment

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 ⇒ 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



Sam Altman ✅
@sama

...

ChatGPT launched on wednesday. today it crossed 1 million users!

8:35 AM · Dec 5, 2022

3,457 Retweets 573 Quote Tweets 52.8K Likes

■ Multiplication of initiatives: GPT, LaMBDA, PaLM, BARD, BLOOM, Gopher, Megatron, OPT, Ernie, Galactica...

■ Public involvement,
impact on information access

Temporary page!

\LaTeX was unable to guess the total number of pages correctly. As there was some unprocessed data that should have been added to the final page this extra page has been added to receive it.

If you rerun the document (without altering it) this surplus page will go away, because \LaTeX now knows how many pages to expect for this document.