PANORAMA DE L'IA & RÉFLEXIONS SUR LES IMPACTS EN RECHERCHE

Mercredi 24 juin 2026, Ile de Ré
Assises du département AlimH, INRAe

Vincent Guigue
https://vguigue.github.io

Introduction

Digital & Artificial Intelligence

- Two related but distinct concepts
- AI: Different Definitions

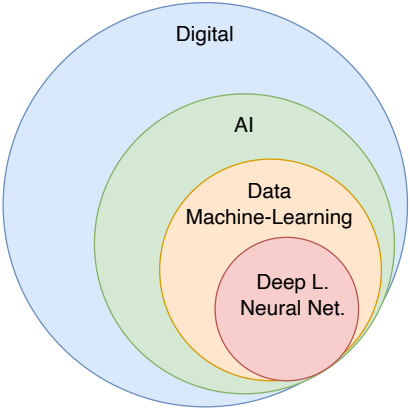
1956 Any algorithm / program
1960-2012 Expert systems and logical reasoning
2012- Data & neural networks

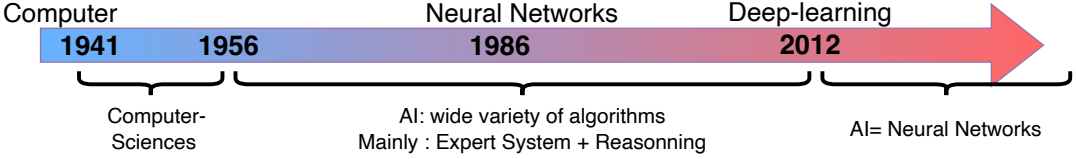

A. Turing


Marvin Minsky

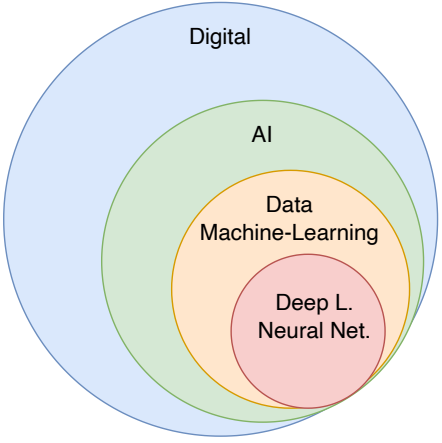

G. Hinton


Y. Lecun





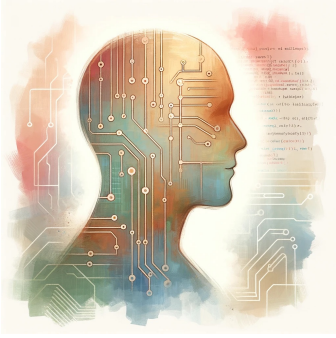
AI's Place in the Digital Domain



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

A lot of digital tools transform the job market...
$\Rightarrow$ A global view is required to understand the transformation

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, for now, are more satisfactorily performed by humans because they require high-level mental processes.

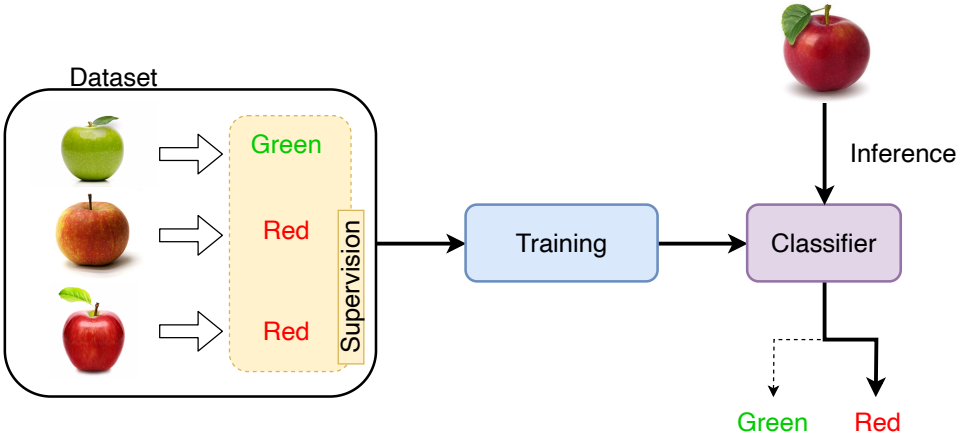
Marvin Lee Minsky, 1956

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

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

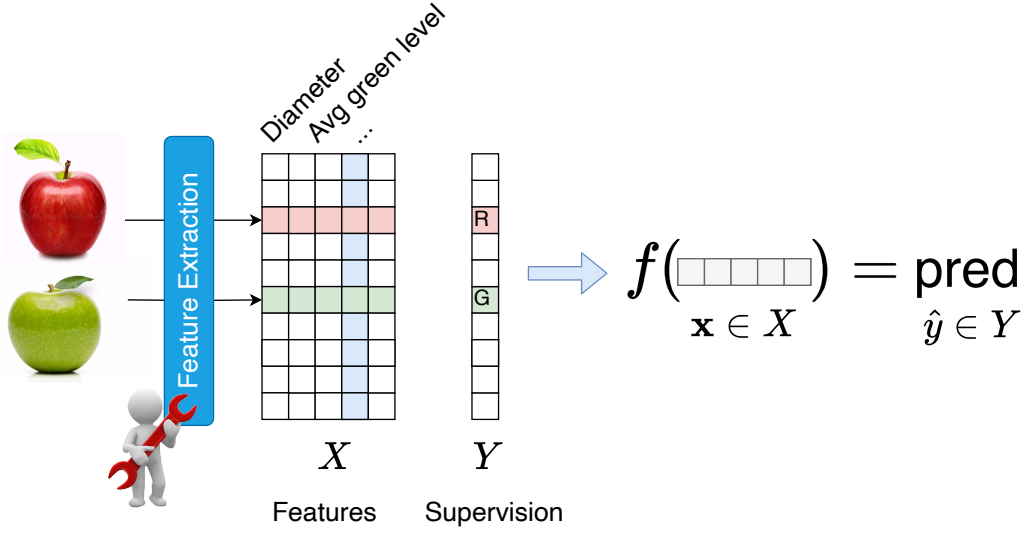
Andrew Ng, 2015

Supervised Processing Chain & Models



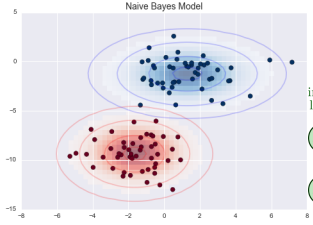
- Promise = building a model *solely* from observations

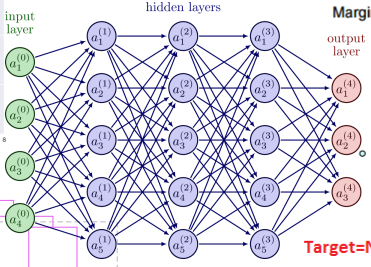
Supervised Processing Chain & Models

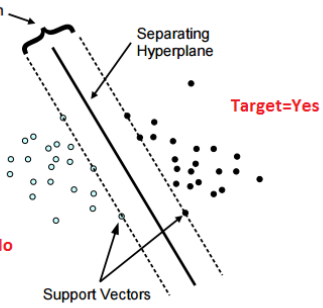


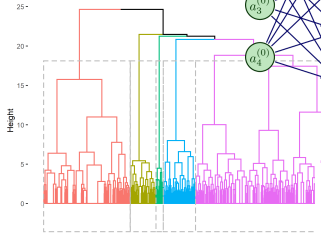
- Think about the feature space = observation description
- ... And about required supervision

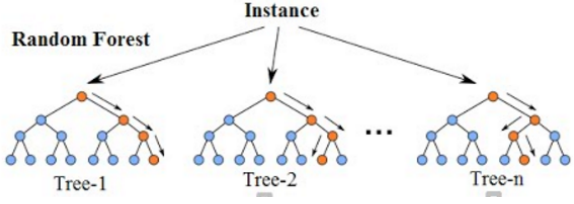
Supervised Processing Chain & Models



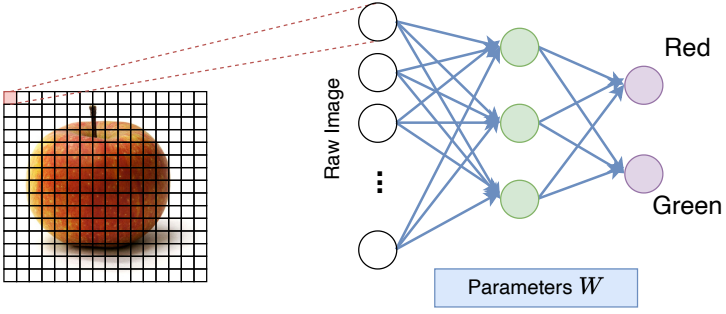








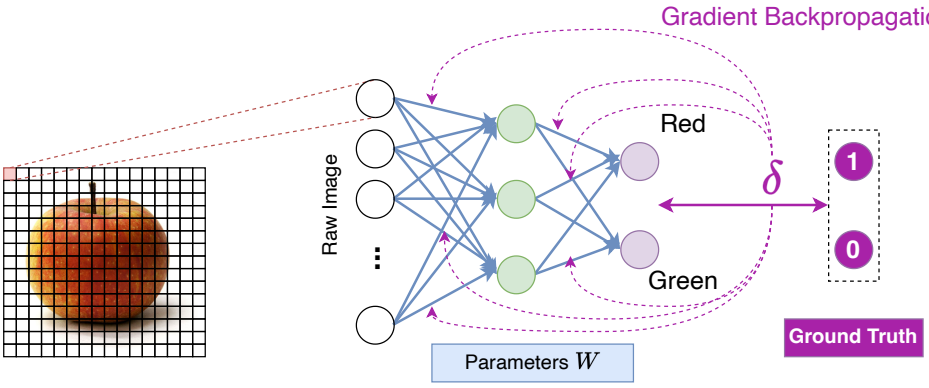
Supervised Processing Chain & Models



- Random initialization...

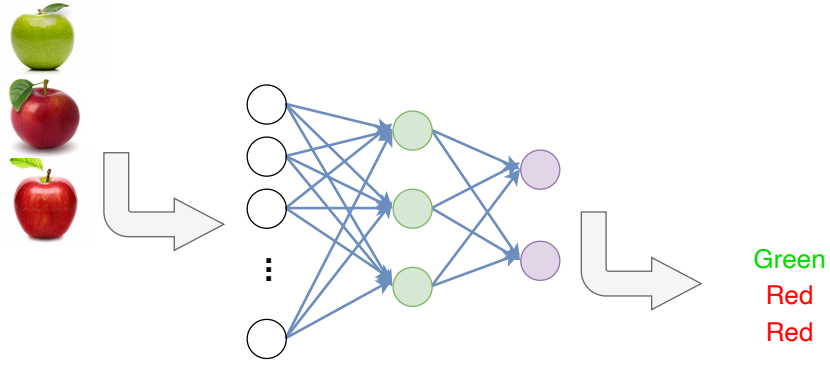
And random decision-making (at first!)

Supervised Processing Chain & Models



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

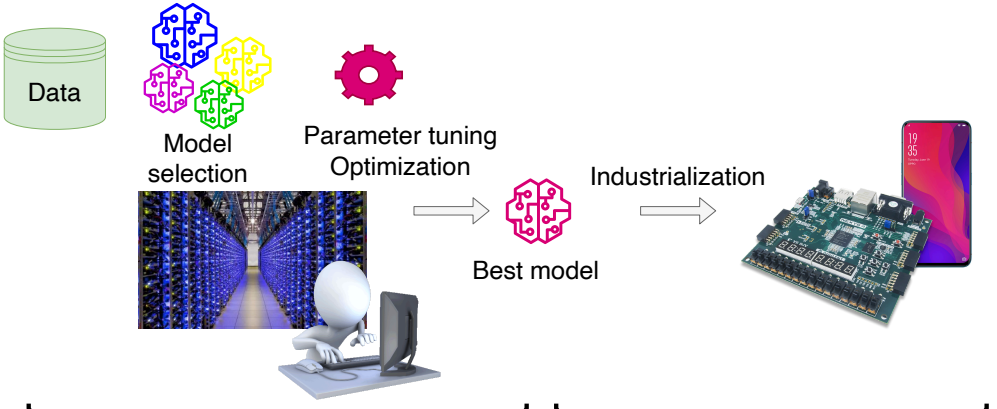
Supervised Processing Chain & Models



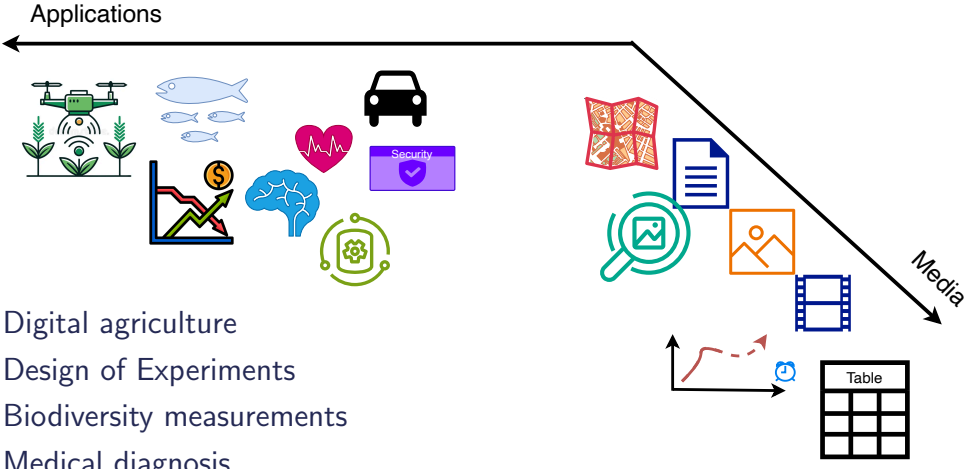
- **Training** is slow and costly
- **Inference** is (much) faster

Supervised Processing Chain & Models

Clearly separate the different steps in machine-learning



What can we do with AI?



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

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

What is the cost to access AI (today)?

Adapt **vision system** to detect objects

Integrate LLMs in processing chains

Use Chatbots

Learn machine learning models on **tabular data**

AlphaFold exploitation (structure / function prediction)

Fintune LLM to **new tasks**

Deep **Vision** adaptation

Time-series analysis

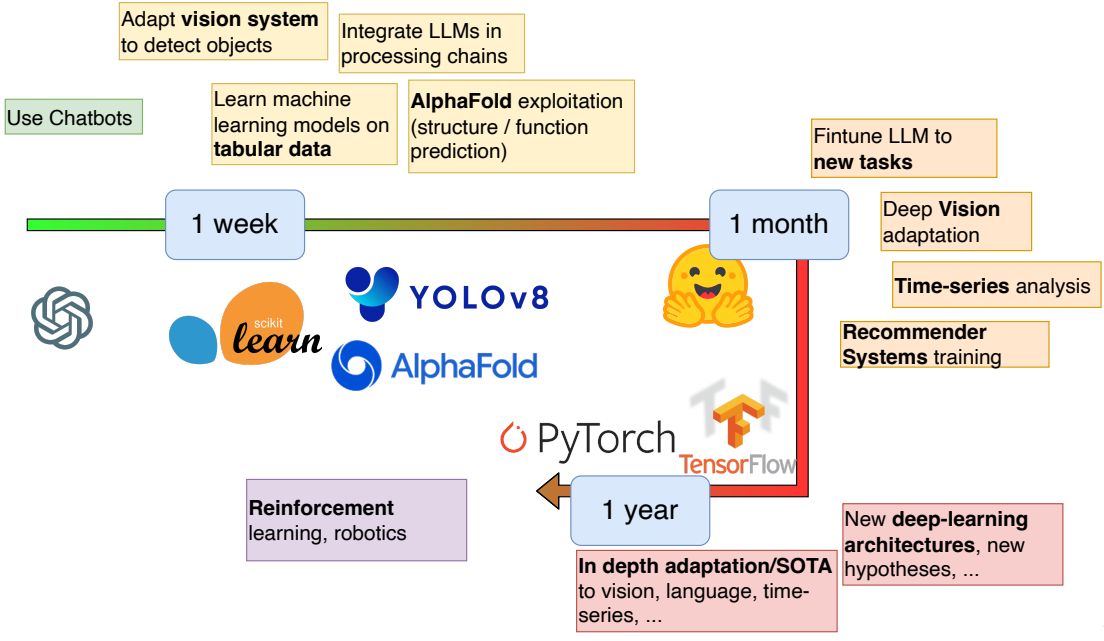
Recommender Systems training

Reinforcement learning, robotics

New **deep-learning architectures**, new hypotheses, ...

In depth adaptation/SOTA to vision, language, time-series, ...

What is the cost to access AI (today)?



What is the cost of AI?

Sensor monitoring, basic decision

Autonomous driving system

Local plateform to build & deploy AI models

Answering LLM calls

Training a vision system

Training an LLM

0 - 10 W

20 - 150 W

0.2 - 1.5 kW

1 - 5 kW

0.5 - 50 MW

- Electricity, water, rare-earth elements, CO_2 emissions, financing costs, etc.

$\Rightarrow$ Most costs scale proportionally

So, let's stop speaking of AI !

Machine-learning

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

Deep-learning

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

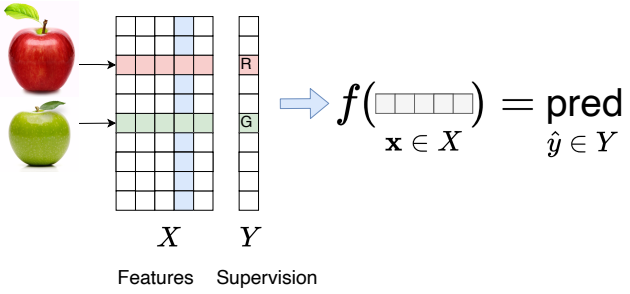
LLM

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

$\Rightarrow$ Choose your keywords more carefully!

From Artificial Intelligence to Deep-Learning & Generative AI

Great opportunity of Machine-Learning



Numerous opportunities

- Experimental design
- Biological interaction prediction
- Nutritional value prediction for animal feed
- Medical diagnosis
- ...

Requirements: what you observe, what you expect

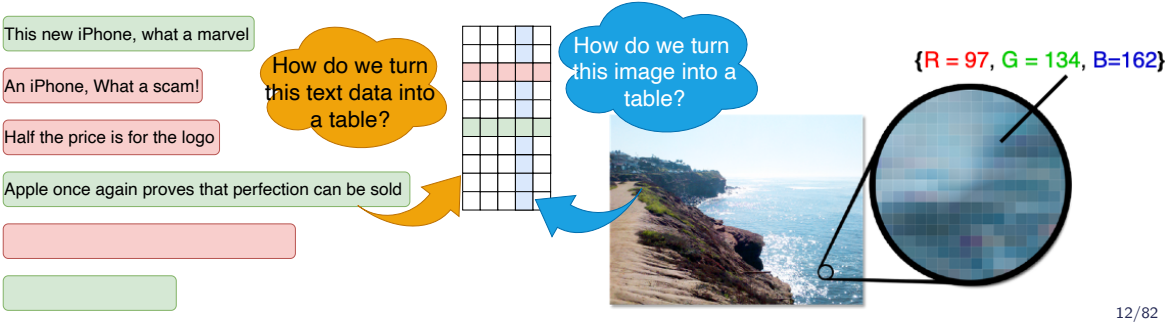
- X : Observations
  - Historical data + perimeter (= what do we need to observe)
  - Variations
- Y : Targets
  - Historical + Synchronized with observations

Various questions: is it possible/easy for a human to make the prediction? Can you explain the reasoning? What format is the data in?

Recent evolution in Machine-Learning

Machine-learning Limitations

1. Performance ambiguity
   - 90% recognition : great, interesting... or unusable?
   - Performance on the end-user task... Or on a specific domain?
2. Range of validity (Climate change, patient change, location shift...)
3. Operational integration (UX, ergonomics, training requirement)
4. How to represent complex input in a table?

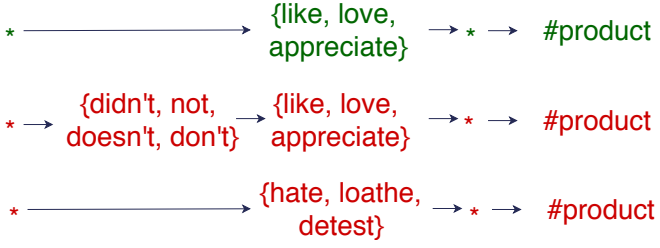


AI + Textual Data: Natural Language Processing (NLP)

NLP = largest scientific community in AI

Linguistics [1960-2010]

Rule-based Systems:

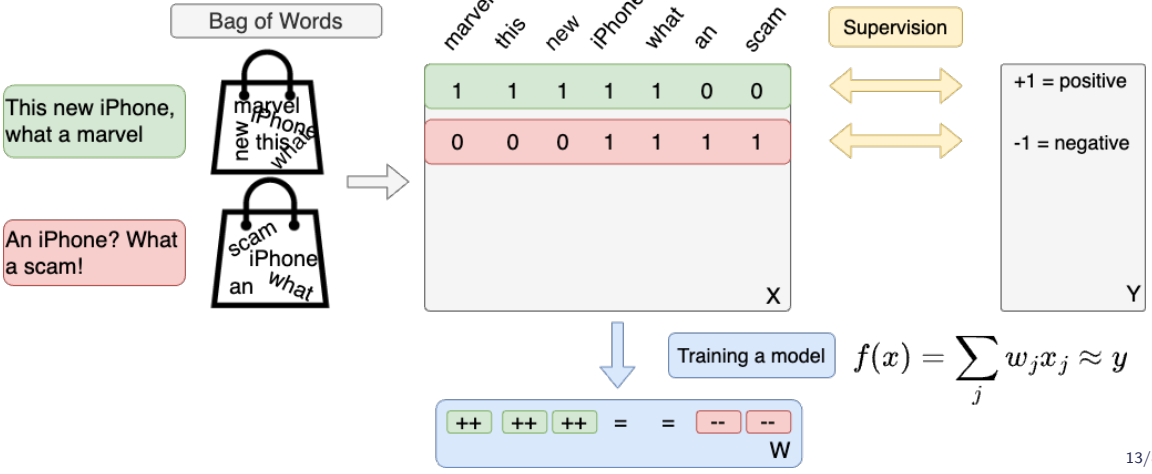


- Requires expert knowledge
- Rule extraction $\Leftrightarrow$ very clean data
- Very high precision
- Low recall
- Interpretable system

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

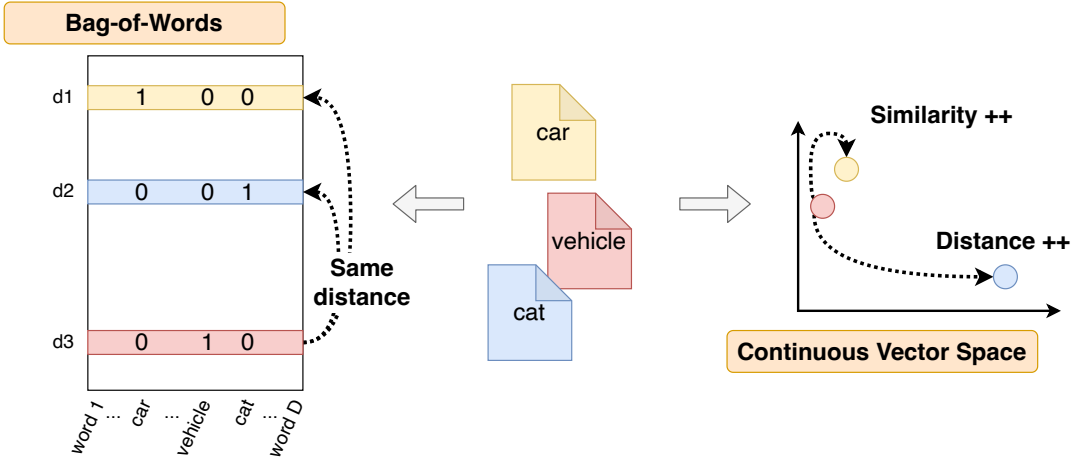
$\approx$ Less interpretable system
− Lower precision
\+ Better recall

Precision = criterion for acceptance by industry

→ Link to metrics

Deep-learning = representation learning

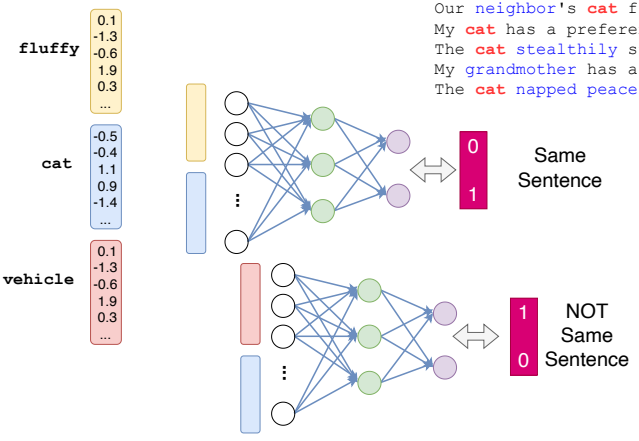
- 1 item = 1 vector $\mathbf{z} = [z_1, \ldots, z_d]$
- 2d illustration... But high dim. in reality $(100 < d < 100,000)$

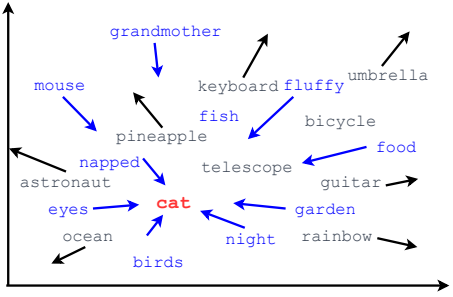


LeCun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

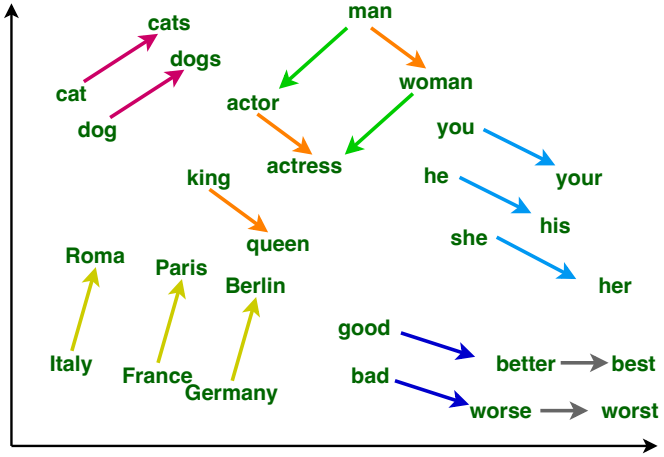
Deep-learning = representation learning

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-learning = representation learning

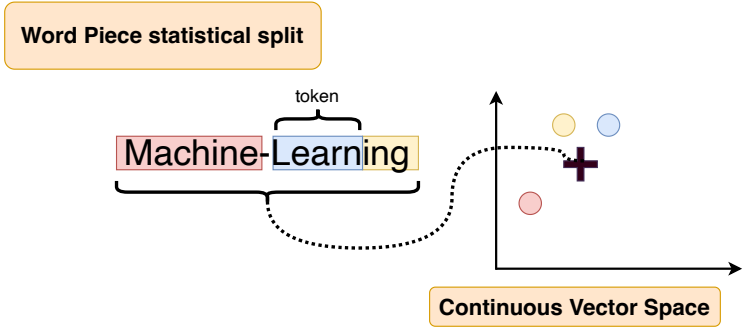


- Semantic Space:
  similar meanings
  $\Leftrightarrow$
  close positions
- Structured Space:
  grammatical regularities,
  basic knowledge, ...

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

Deep-learning = representation learning

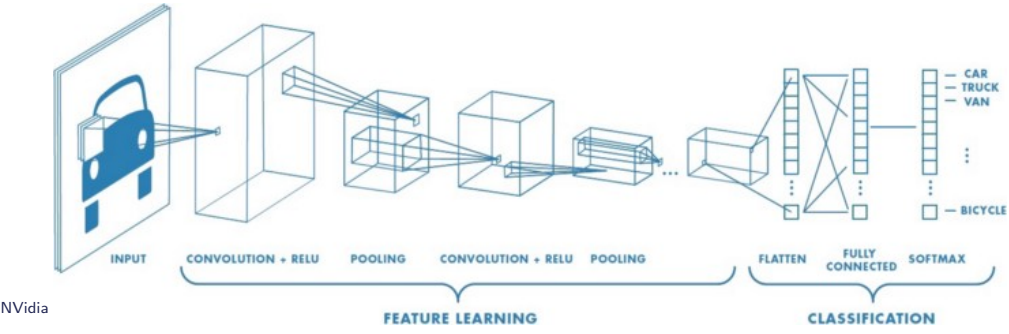
From Words to Tokens



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

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

Image representation

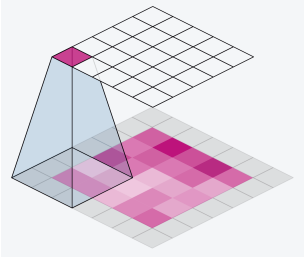


NVidia

Convolution

Sliding filters

- Few parameter (relatevly frugal)
- Pattern extraction
- Hierarchical information extraction/aggregation

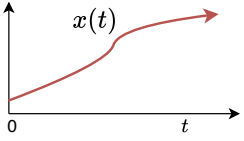


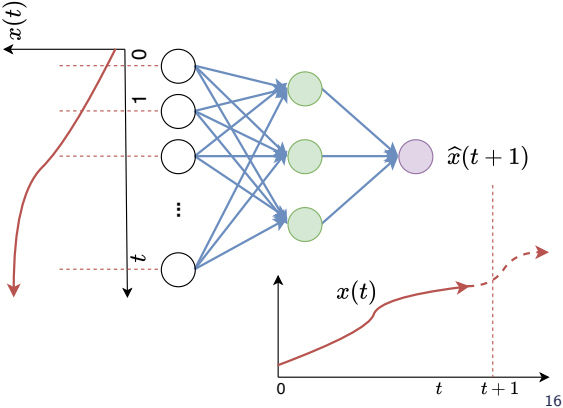
Time-series representation

Any dynamical system prediction

- Plant/animal growth
- Stock market
- Temperature evolution
- ...

- Various possible architecture
- Pattern extraction





Once again :
time series = vector representation

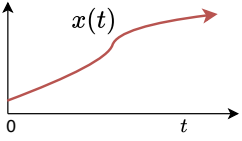
- Distance / similarity
- Integration with other data

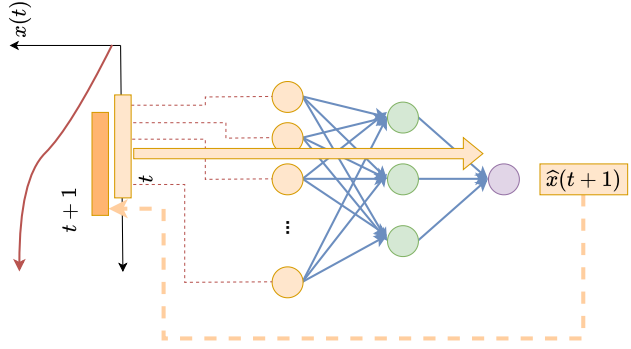
Time-series representation

Any dynamical system prediction

- Plant/animal growth
- Stock market
- Temperature evolution
- ...

- Various possible architecture
- Pattern extraction



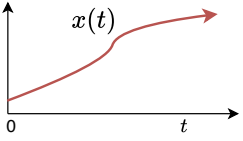


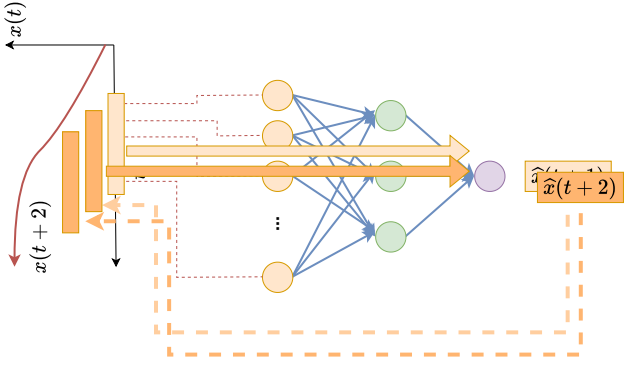
Time-series representation

Any dynamical system prediction

- Plant/animal growth
- Stock market
- Temperature evolution
- ...

- Various possible architecture
- Pattern extraction



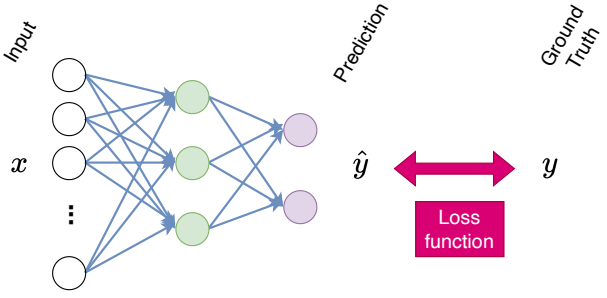


Deep-learning = taylor made solutions

Combining different objectives at different scales

- Triplet loss
- Self supervision / masking
- Profiling to tackle missing information
- ...

Taylor made = more expensive

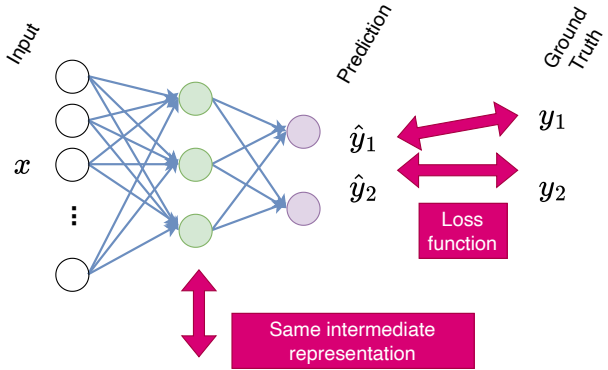


Deep-learning = taylor made solutions

Combining different objectives at different scales

- Triplet loss
- Self supervision / masking
- Profiling to tackle missing information
- ...

Taylor made = more expensive

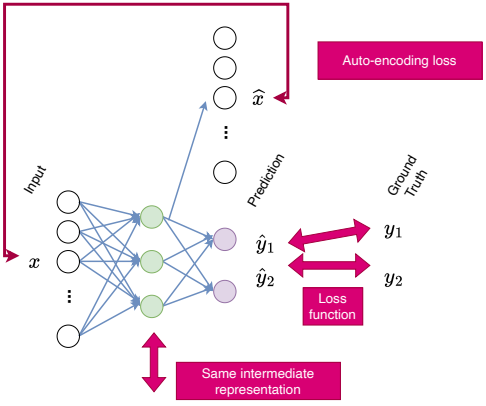


Deep-learning = taylor made solutions

Combining different objectives at different scales

- Triplet loss
- Self supervision / masking
- Profiling to tackle missing information
- ...

Taylor made = more expensive

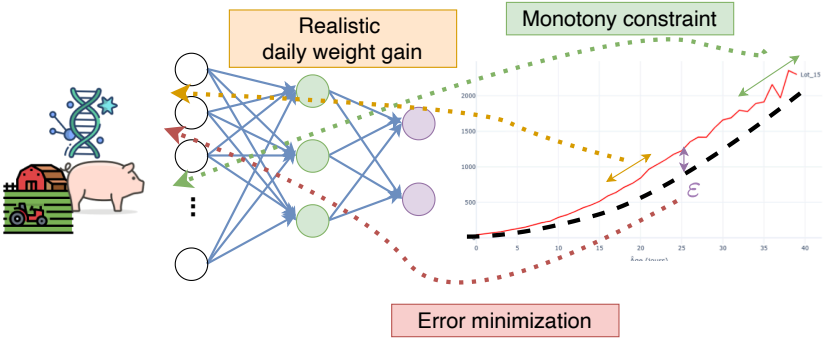


Deep-learning = taylor made solutions

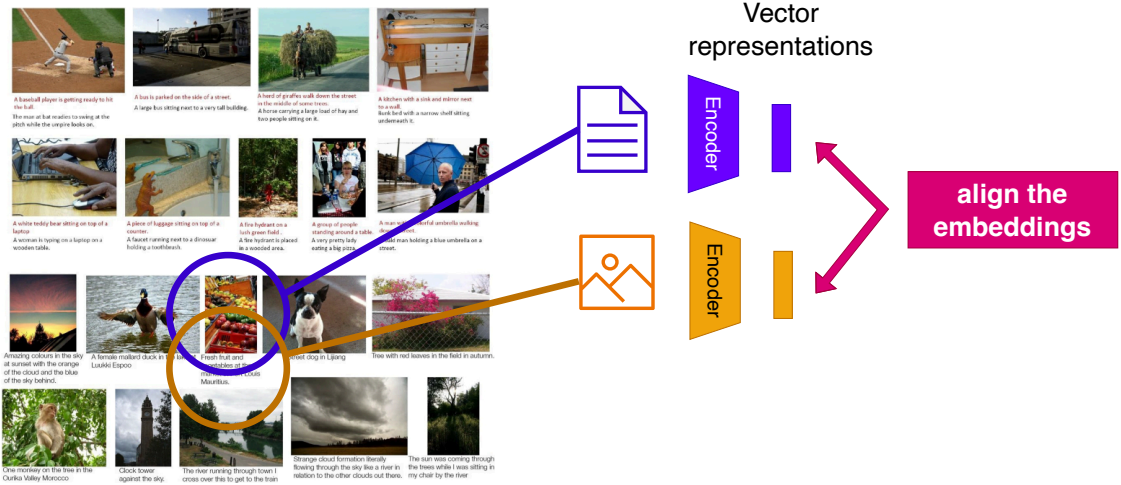
Combining different objectives at different scales

- Triplet loss
- Self supervision / masking
- Profiling to tackle missing information
- ...

Taylor made = more expensive



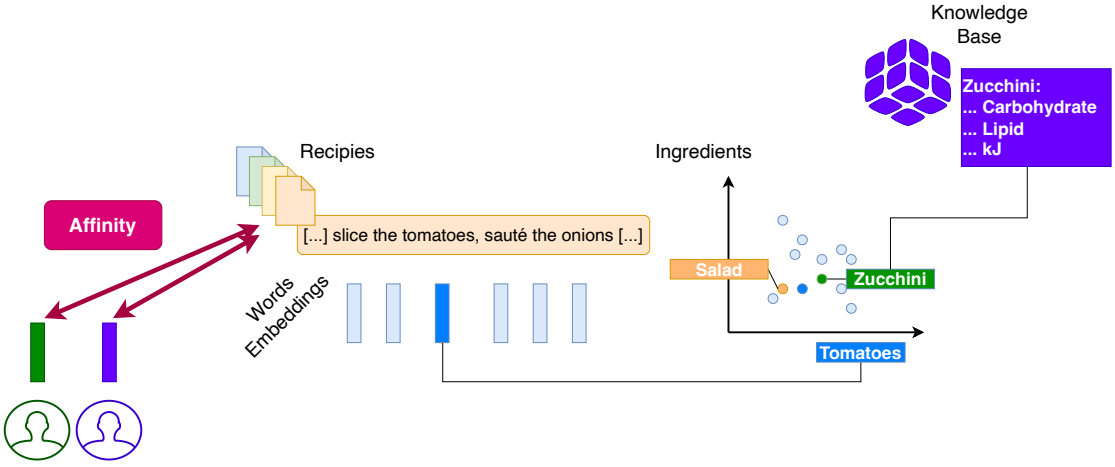
Deep-learning = deal with different modalities



- Aligning languages (translation), aligning modalities
- New applications: image description, handwritten transcription, ...

Word, music, speech, image, user embeddings

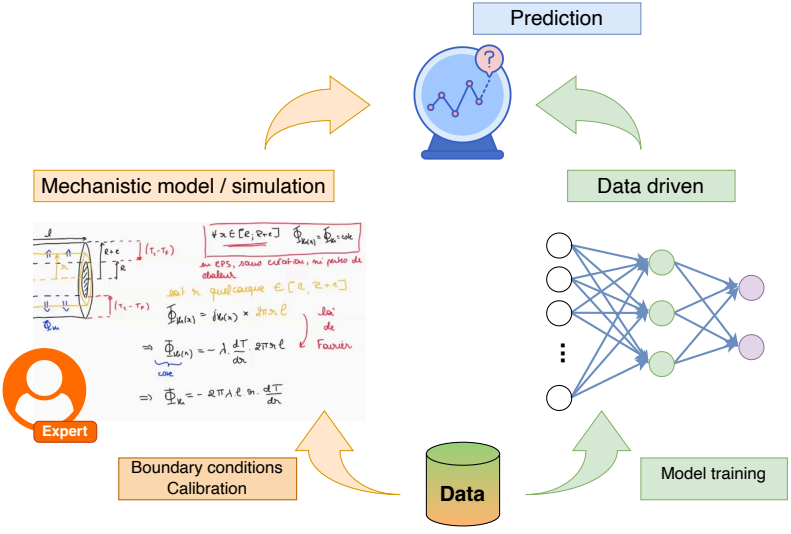
- User embedding ⇒ recommender system
- Knowledge embedding ⇒ neuro-symbolic architecture



Warning: each arrow = aligned dataset required !

Physics Informed Machine-Learning (PIML / PINNs)

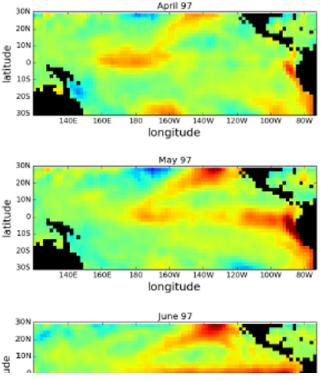
- Historically, two very different approaches
- Huge ambiguity about modeling !

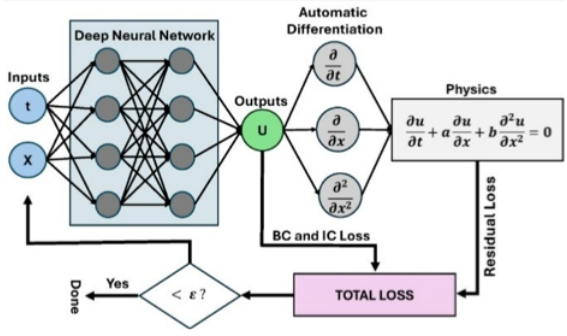


Physics Informed Machine-Learning (PIML / PINNs)

- Physics Informed Neural Networks
  = integrating ODE/PDE inside NN architecture
- e.g. (1) temperature estimation (with NN), (2) heat diffusion (ODE)

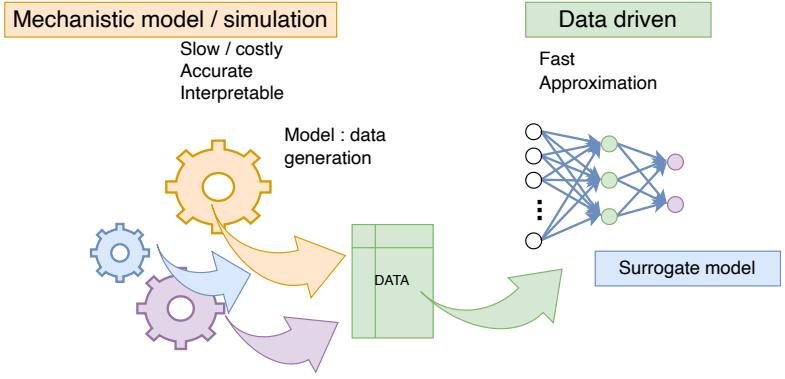
Physical constraints
Differential equation modeling
inside neural architecture





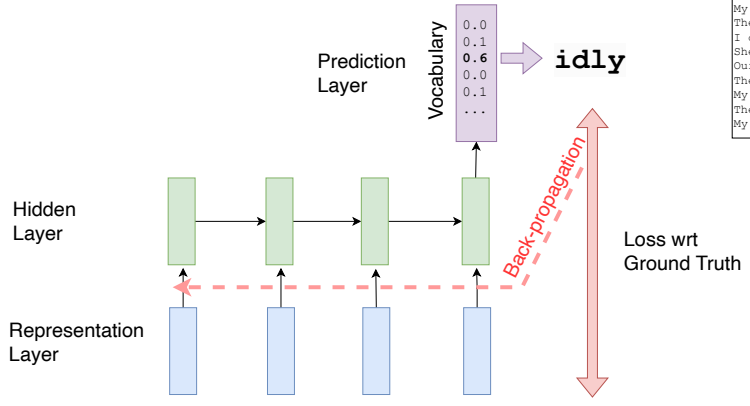
Physics Informed Machine-Learning (PIML / PINNs)

- Surrogate model
- ↗ speed, ↘ input req., new calibration options, ...



Aggregating word representations: towards generative AI

- Generation & Representation
- New way of learning word positions

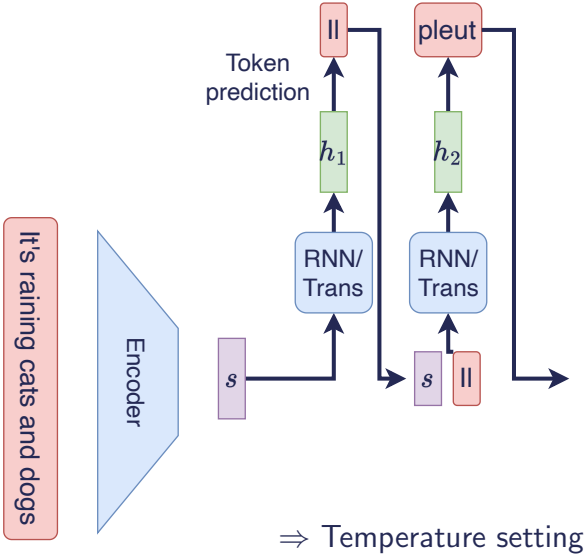


The fluffy cat napped lazily in the sunbeam.
I adopted a stray cat from the shelter last week.
My cat loves to chase after toy mice.
The black cat stealthily crept through the dark alley.
I often find my cat perched on the windowsill, watching birds.
She gently stroked her cat's fur as it purred contentedly.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

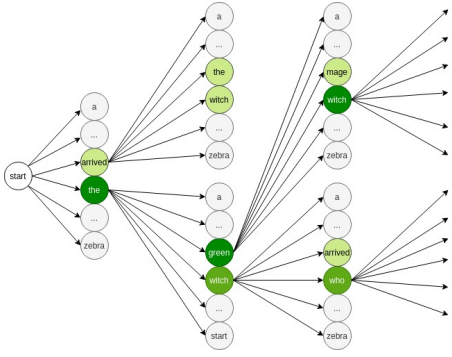
Sequence to Sequence Learning with Neural Networks, Sutskever et al. NeurIPS 2014

Inference & Beam Search



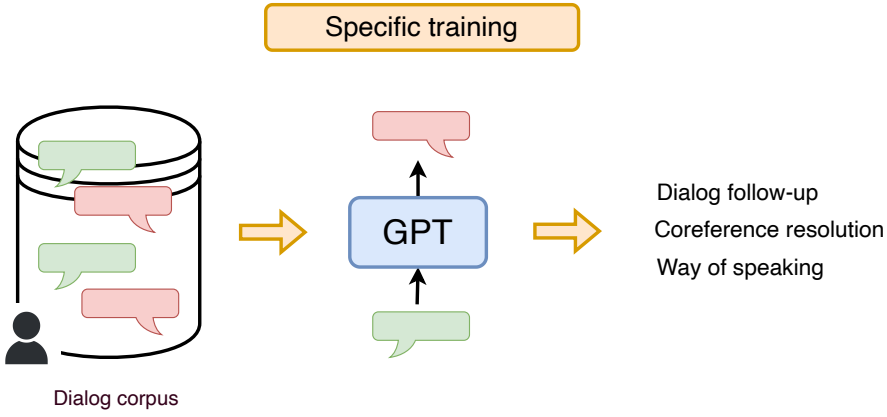
$\Rightarrow$ Temperature setting

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



The additional ingredients of chatGPT

Dialogue Tracking

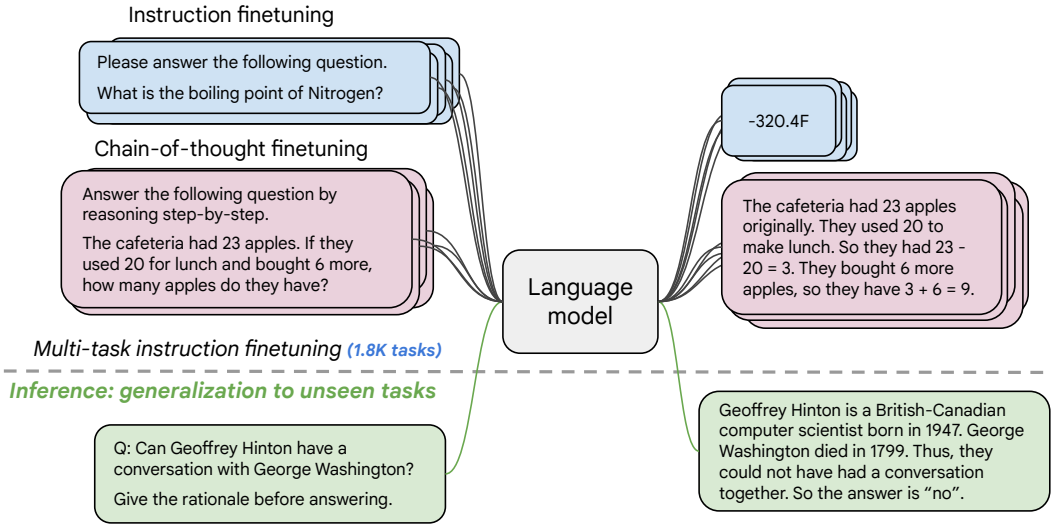


- **Very clean** data

Data generated/validated/ranked by humans

The additional ingredients of chatGPT

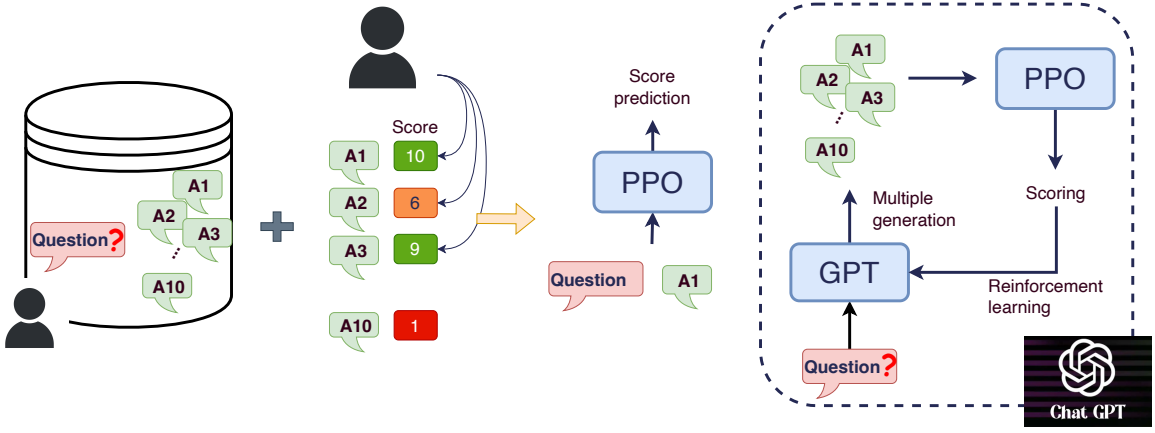
Fine-tuning on different (±) complex reasoning tasks



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

The additional ingredients of chatGPT

Instructions + answer ranking

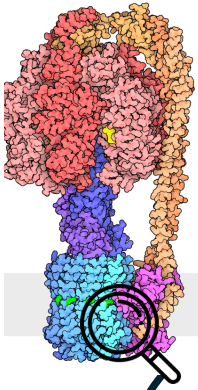


- Database created by humans
- Response improvement

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

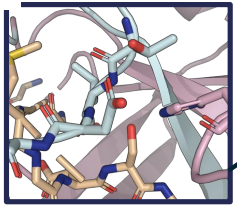
Training language models to follow instructions with human feedback, Ouyang et al., 2022

DNA: yet another language?



Fonction

Structure



Séquence

Méthodes expérimentales

Prédictions physiques et statistiques

Quignot, Postic, **Bret** et al. *Bioinformatics*, 2021

1ers réseaux convolutifs pour les protéines

Réseaux type transformers

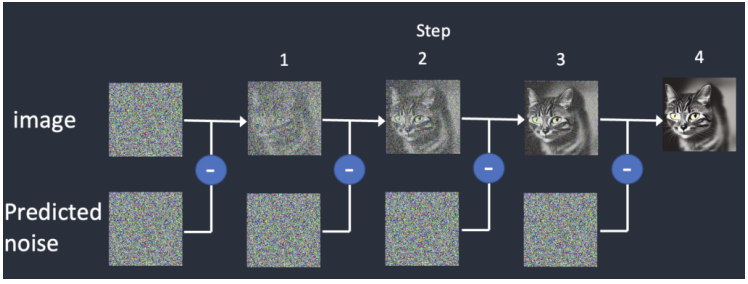
AlphaFold2

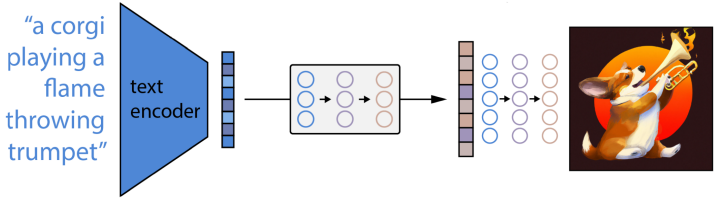
Prix Nobel de Chimie 2024

Jumper et al. *Nature* 2021

Image generation

- Learning to remove noise from images
- + Aligning image representation to textual prompt

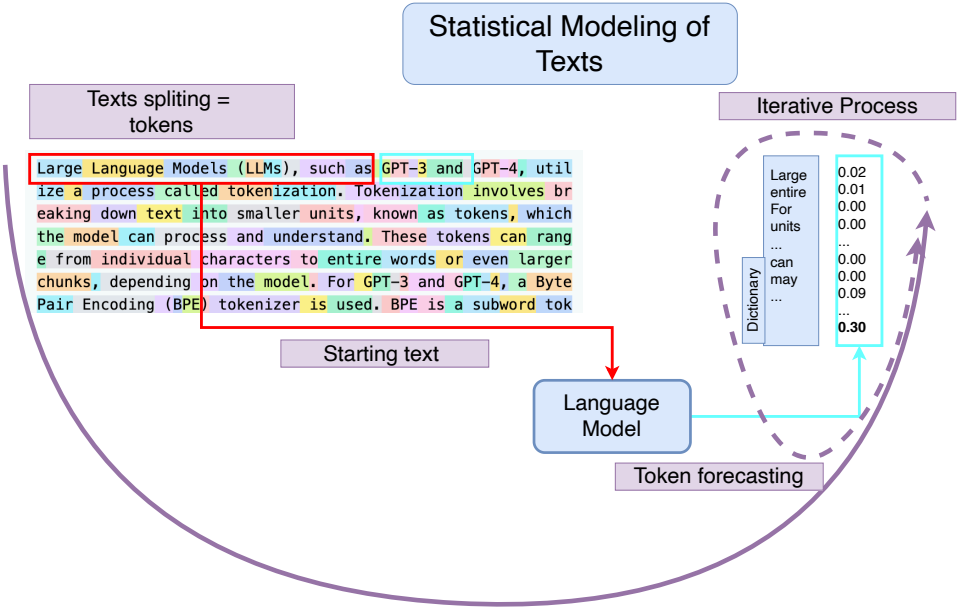




Denoising Diffusion Probabilisti Models, NeurIPS, 2020
Ho, J., Jain, A., & Abbeel, P.

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

ChatGPT: theory & uses



Conclusion

- Some systems are already fully available (0-shot)
  - chatbot, AlphaFold, ...

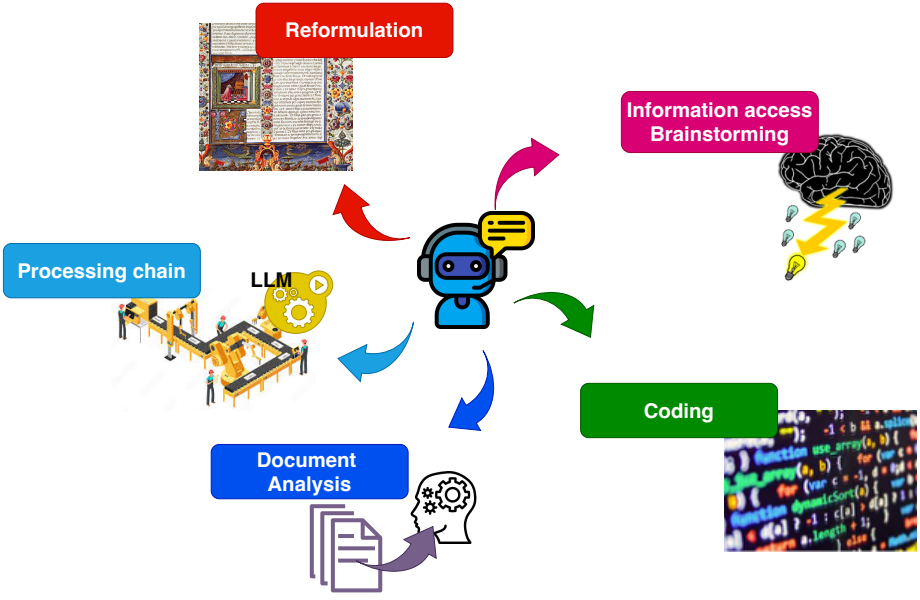
  $\Rightarrow$ Let's try it, integrate it in processing chains
- Some systems are cheap to develop
  - Fine-tuning vision / LLM
  - Machine-Learning & tabular data

  $\Rightarrow$ Let's focus on application (before data)
- Some systems are more expensive to develop
  - Advanced time-series, specific language/vision model, heterogeneous data integration...

  $\Rightarrow$ Require PhD grant / research project

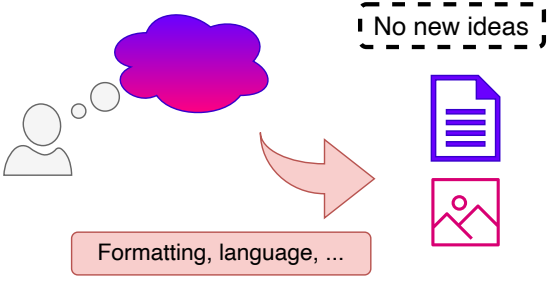
What can we do with LLMs (chatbots)?

Key uses in 5 pictures



(1) Formatting information

A fantastic tool for **formatting**

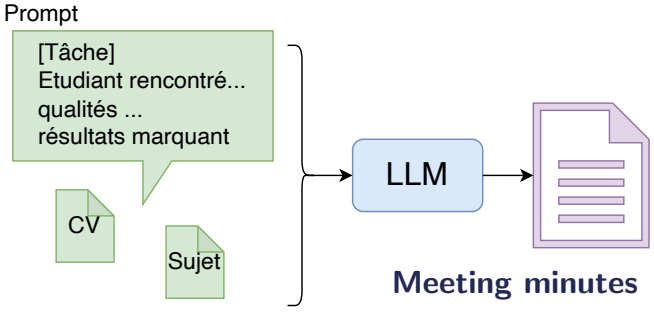


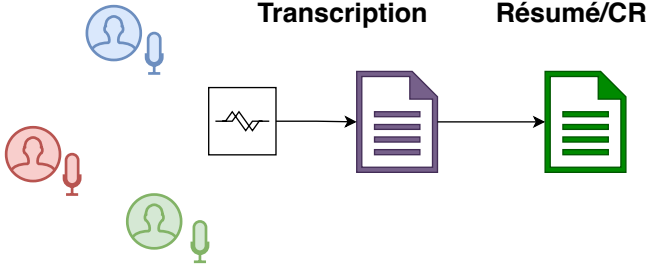
- Personal assistant
  - Standard letters, recommendation letters, cover letters, termination letters
  - Translations
- Meeting reports
  - Formatting notes
- Writing scientific articles
  - Writing ideas, in French, in English

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

(1) Examples of Data Formatting

Building a recommendation letter





(1) Writting a scientific article

- Writing a paragraph from a bullet point list
- **Translating** a paragraph
- **Rephrasing** a paragraph
  - Improving the style, testing multiple formulations
  - Saving 10% of words in a paragraph

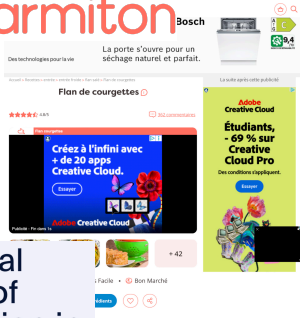
  $\Rightarrow$ *If you haven't already, you should give it a try!*
- Generating the abstract
- Formatting a table, adding a column (in a latex tabular !)

$\Rightarrow$ Pure surface change... [we will also explore other usages in the next slides]

(1) Nutrition use : Input standardization

⇒ opportunity to fuse heterogeneous information

marmiton



Food.

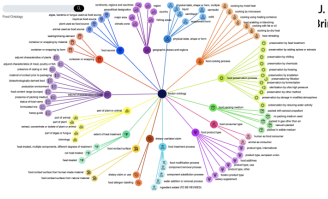
INCA2 Individual national study of food consumption in France Database

anses
agence nationale de sécurité sanitaire alimentation, environnement, travail

ARTICLE OPEN

FoodOn: a harmonized food ontology traceability, quality control and data ir

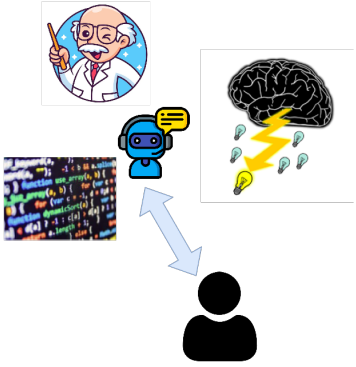
J. Griffiths[2,8], Gurinder S. Gosal[1], Pier L. Buttigieg[3], Ro Irinkman[2] and William W. L. Hsiao[1,2,7]



(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

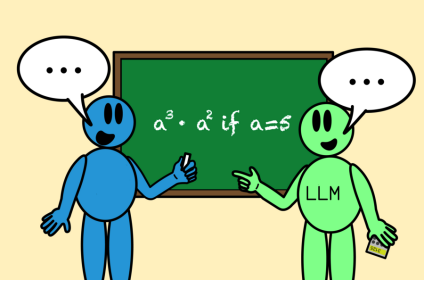
- In which areas are LLMs reliable?
- What are the risks for primary information sources?
- What societal risks for information?

(2) In a scientific context: a new research partner

- **Testing** an idea (strength, weakness, required experiments, ...)
  - Be careful not to mistake LLM flattery for validation!
- Proposing different schedules
- Searching for an **attracting acronym** with specific words
- Asking for weak point, asking for questions a reviewer would write...

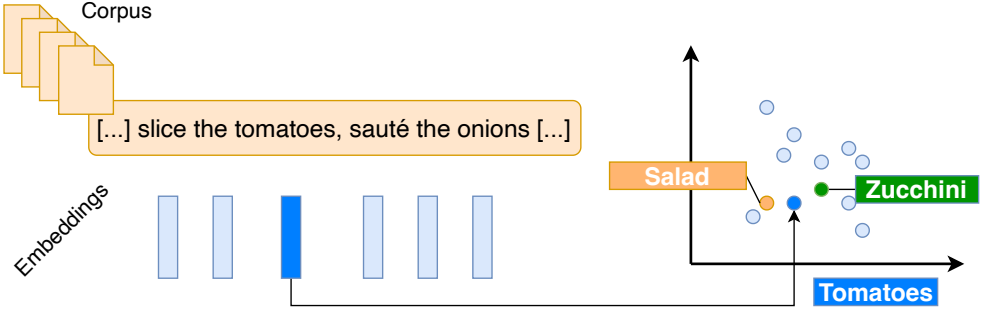
  $\Rightarrow$ Try it on a paragraph or a section
- Ask for a whole review, ask for weak and strong point...

  Check how you paper is seen by a chatbot



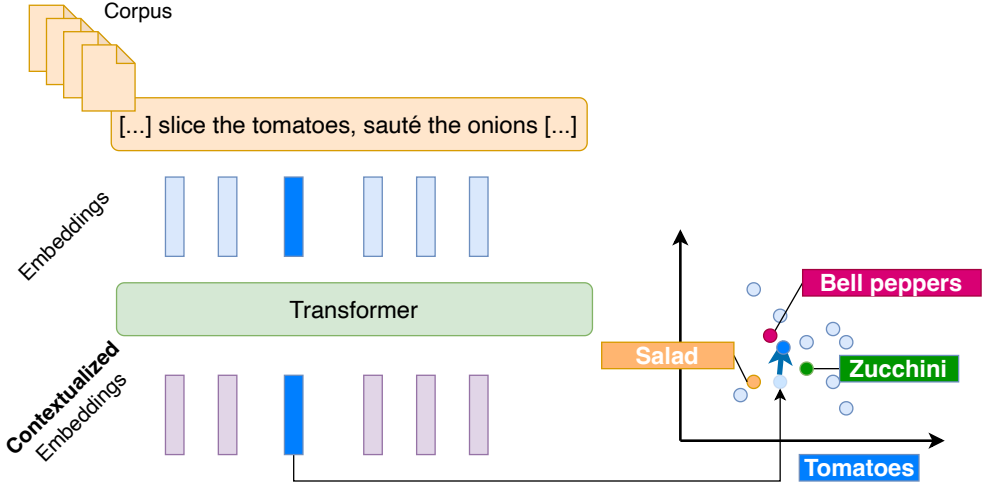
(2) Internal knowledge exploitation for nutrition

- Brainstorming in the kitchen: which application for cooking?
- Ingredient substitution... At every scale: Ingredient, Food, Dish



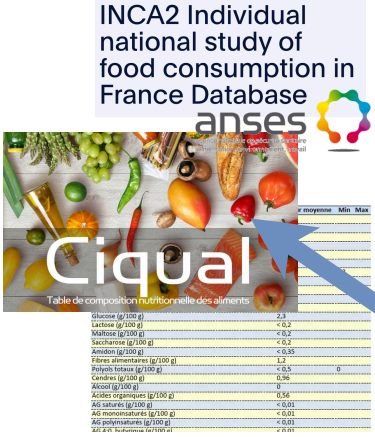
(2) Internal knowledge exploitation for nutrition

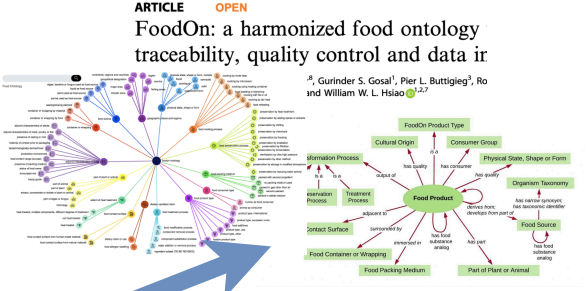
- Brainstorming in the kitchen: which application for cooking?
- Ingredient substitution... At every scale: Ingredient, Food, Dish
- ++ Upgrade by contextualization



(2) Internal knowledge exploitation for nutrition

- Brainstorming in the kitchen: which application for cooking?
- Ingredient substitution... At every scale: Ingredient, Food, Dish
- ++ Upgrade by contextualization
- Interoperability and ontologies





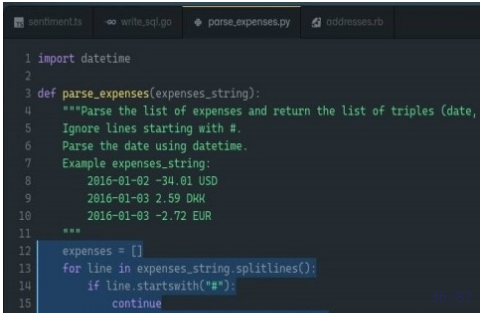
(3) Coding: Different Tools, Different Levels

- 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?



(3) Issues: Scope & Integration

- Classic question, coding-oriented, discovery, *how-to*...
- Validation on existing code
- Reasoned response + code!

⇒ Ideal for **learning**, **fixing** small bugs, proposing **basic solutions**

3️⃣ Script de tirage aléatoire + génération LaTeX

Principe

- Lecture du fichier GIFT
- Tirage **N questions aléatoirement**
- Conversion vers un **environnement** `enumerate` **LaTeX**

Exemple de script Python (fonctionnel)

python — Copier le code

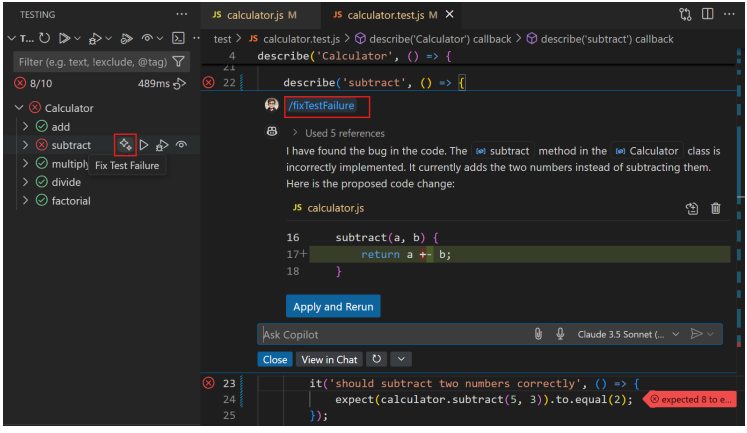
```python
import re
import random

N = 10  # nombre de questions tirées

with open("banque.gift") as f:
    content = f.read()

questions = re.findall(
    r"::.*?::\s*(.*?)\n\{\s*(.*?)\}",
    content,
```

(3) Issues: Scope & Integration

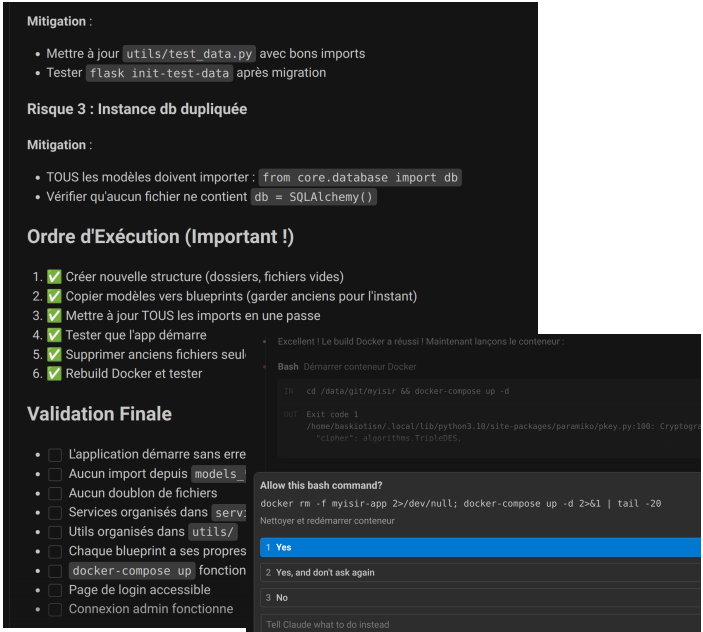
GitHub Copilot

Chatbot integration within the IDE (desktop development environment)

⇒ Coding **faster** for **developers**



(3) Issues: Scope & Integration



CLAUDE CODE

Thinking at the **project** level

- Architecture proposal,
- Methodology validation,
- Code suggestions... + validation interface

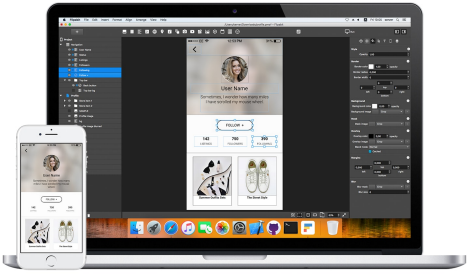
$\Rightarrow$ The developer lets the chatbot write the code but validates it as they go

(3) What about *no-code* (or *low-code*) approaches?

No-Code

Pre-defined patterns/templates for: websites (various), basic applications, ...

Promises that are (mostly) effective, but in **fairly limited use cases**



Exemple de script Python (fonctionnel)

python — Copier le code

```python
import re
import random

N = 10  # nombre de questions tirées

with open("banque.gift") as f:
    content = f.read()

questions = re.findall(
    r"::.*?::\s*(.*?)\n\{\s*(.*?)\}",
    content,
```

Low-code

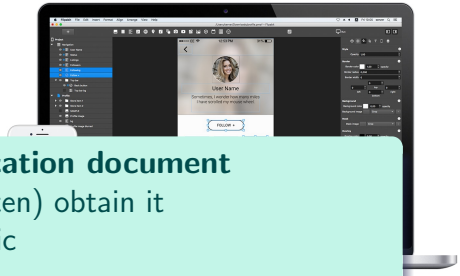
LLM requests for code generation
+ Fast integration with little to no verification
Speed & impression of mastery... But **taking risks** with development reliability

(3) What about *no-code* (or *low-code*) approaches?

No-Code

Pre-defined patterns/templates for: websites (various), basic applications, ...

Pron
fairl



Just remember that a **prompt is a specification document**
⇒ Users who know what they want will (often) obtain it
Wrong/incomplete specifications ⇒ Off topic

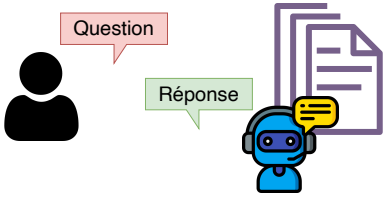
Exemple

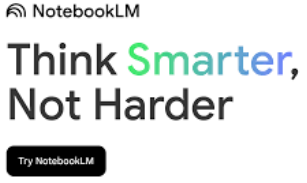
```python
import
import

N = 10  # nombre de questions tirées

with open("banque.gift") as f:
    content = f.read()

questions = re.findall(
    r"::.*?::\s*(.*?)\n\{\s*(.*?)\}",
    content,
```

verification
Speed & impression of mastery... But **taking risks** with development reliability

(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

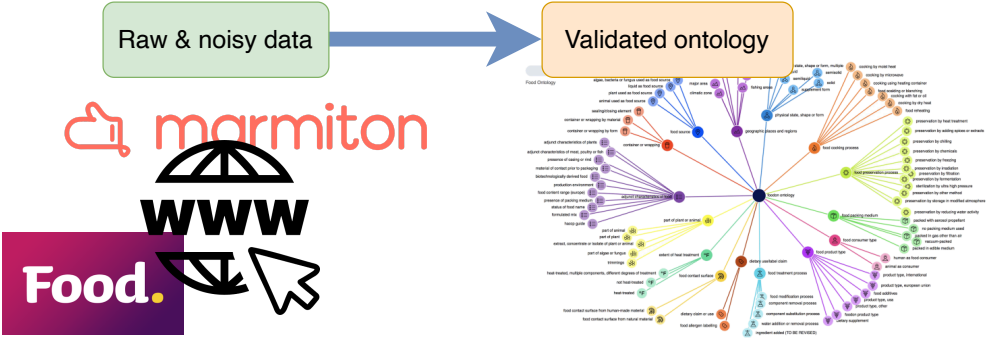




- Will articles still be read in the future?
  - Should we make our articles NotebookLM-proof?
- How to save time while remaining honest and ethical?

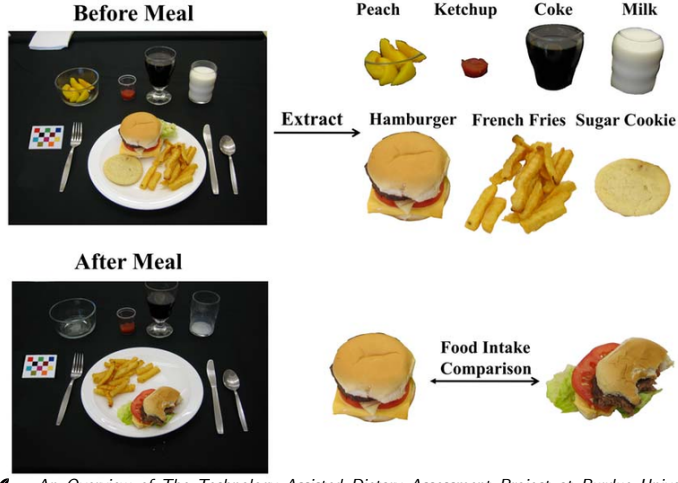
(4) Information Extraction in Nutrition

- Ontology enrichment/building (mostly textual data)



(4) Information Extraction in Nutrition

- Ontology enrichment/building (mostly textual data)
- Image analysis (new UX opportunities)

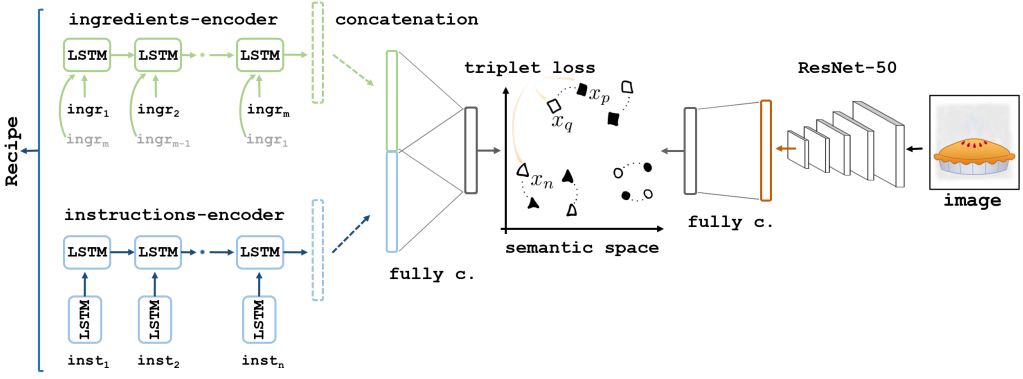


- Food recognition
- Segmentation
- Estimation of quantities

An Overview of The Technology Assisted Dietary Assessment Project at Purdue University., Khanna et al. , 2010

(4) Information Extraction in Nutrition

- Ontology enrichment/building (mostly textual data)
- Image analysis (new UX opportunities)
- Multimodal analysis + algorithmic process



Images & Recipes: Retrieval in the cooking context, SIGIR 2018
Carvalho et al.

(4) Information Extraction in Nutrition

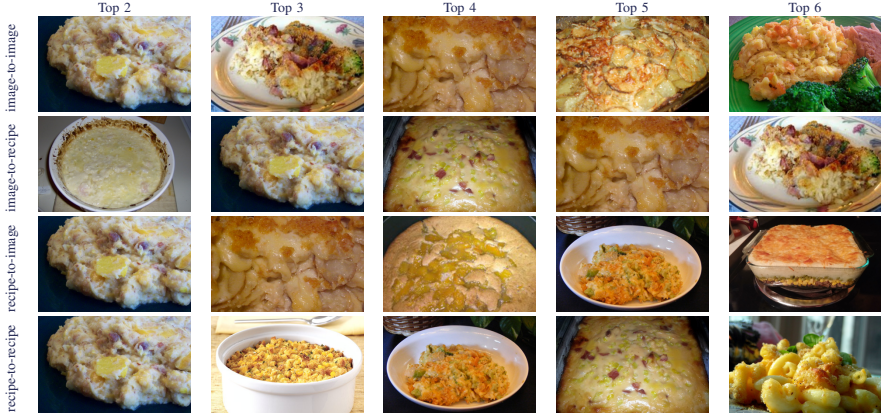
- Ontology enrichment/building (mostly textual data)
- Image analysis (new UX opportunities)
- Multimodal analysis + algorithmic process

	ingr (ingredients)	instr (cooking instructions)	image
Pizza	1) *pizza dough* 2) *hummus* 3) *arugula* 4) *cherry / grape tomatoes* 5) *pitted greek olives* 6) *crumbled feta cheese*	1) *Cut the dough into two 8-ounce sized pieces.* 2) *Roll the ends under to create round balls.* 3) *Then using a well-floured rolling pin, roll the dough out into 12-inch circles.* 4) *Place the dough circles on sheets of parchment paper.*	
Pecan Pie	1) *unsalted butter* 2) *eggs* 3) *condensed milk* 4) *sugar* 5) *vanilla extract* 6) *chopped pecans* 7) *chocolate chips*	1) *Preheat the oven to 375 degrees F.* 2) *In a large bowl, whisk together the melted butter and eggs until combined.* 3) *Whisk in the sweetened condensed milk, sugar, vanilla, pecans, chocolate chips, butterscotch chips, and coconut.*	

Images & Recipes: Retrieval in the cooking context, SIGIR 2018
Carvalho et al.

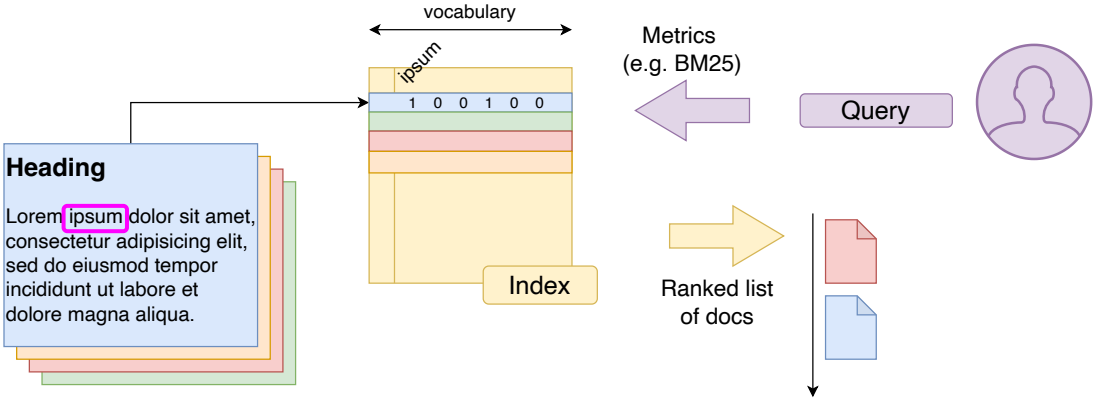
(4) Information Extraction in Nutrition

- Ontology enrichment/building (mostly textual data)
- Image analysis (new UX opportunities)
- Multimodal analysis + algorithmic process

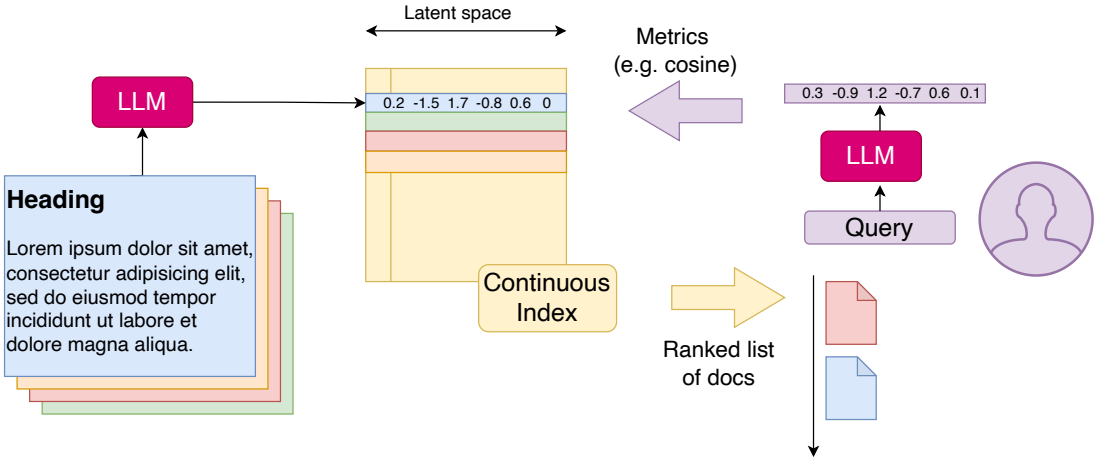


Images & Recipes: Retrieval in the cooking context, SIGIR 2018
Carvalho et al.

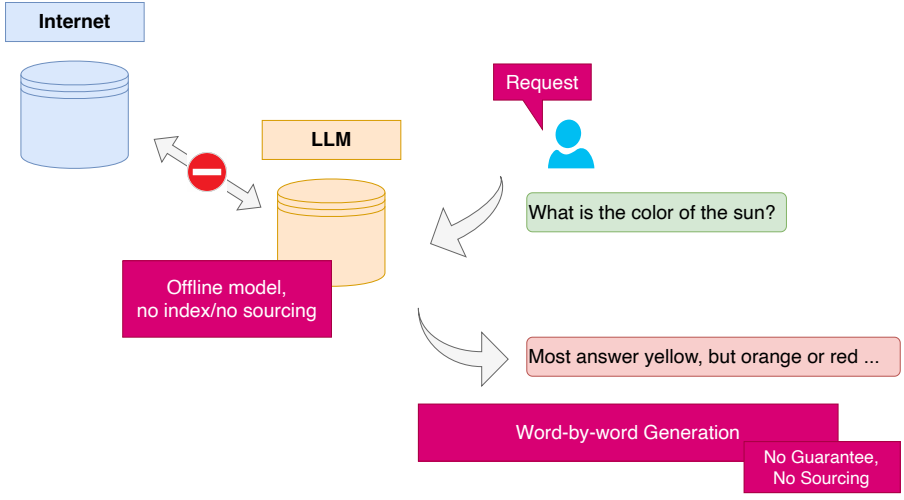
(4) Chat & RAG : a new way to access information



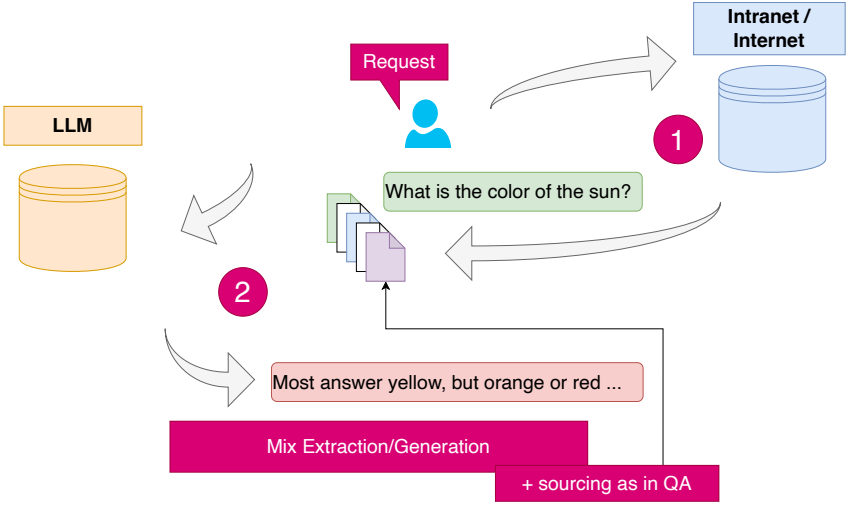
(4) Chat & RAG : a new way to access information



(4) Chat & RAG : a new way to access information



(4) Chat & RAG : a new way to access information



⇒ A way to build a *reliable* chatbot to advise users?

- Parametric memory *vs* Information Retrieval : opposite objectives?

(4) In a scientific context

- RAG in a scientific context
  should be called **bibliography** !
  - Consensus, Scopus.ai, opscidia, ...

  ⇒ great summary + analysis... On which articles?
  - NotebookLM : choose your articles (up to 50)...
    Then start the discussion

  ⇒ analysis, topic clustering, comparison, ...
- If you want to dialog with a manuscript...
  It is too long for an LLM

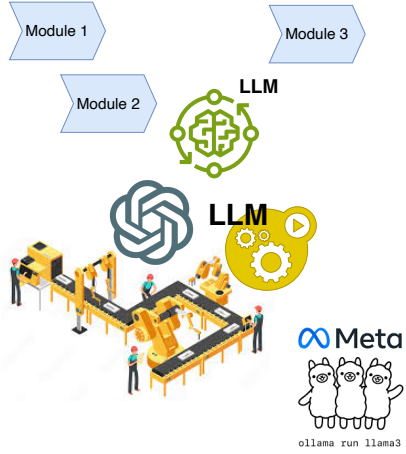
  ⇒ RAG again !

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

(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?

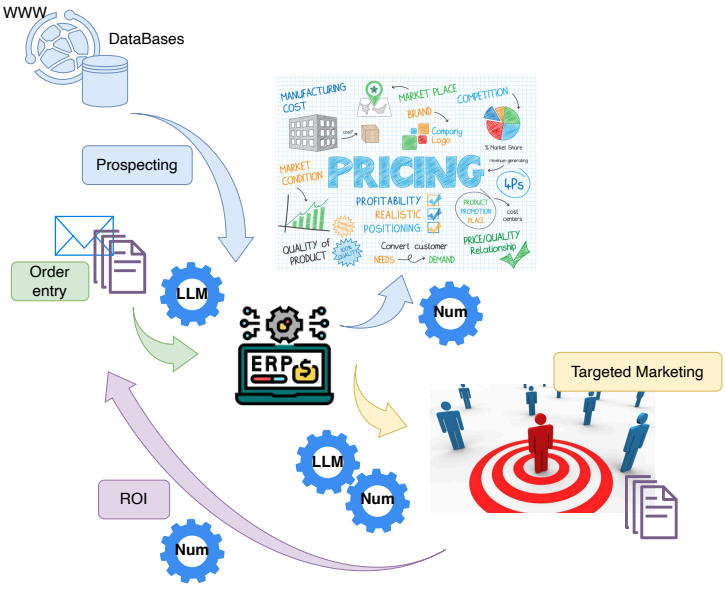
(5) LLM in a Production Pipeline / Agentic AI

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

⇒ Integrate t

=

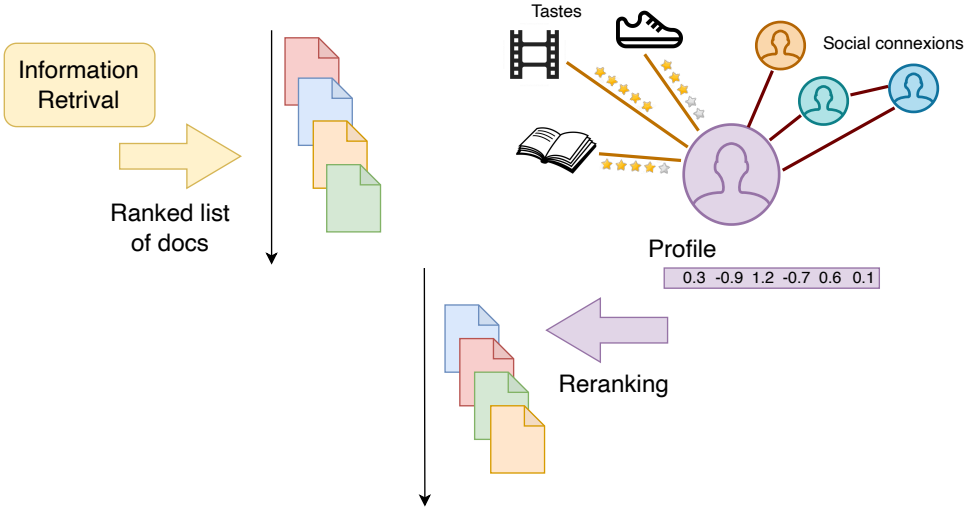
- Can I t
- How m
- How go



Module 3

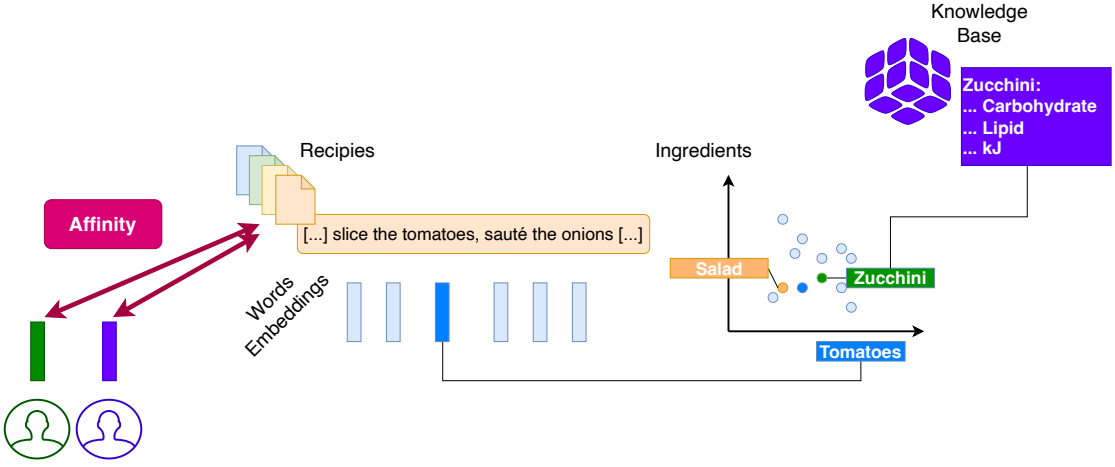
(5) What about Recommender System in Nutrition?

Profiling is roughly everywhere in Information Retrieval



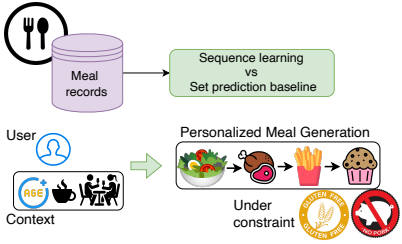
(5) What about Recommender System in Nutrition?

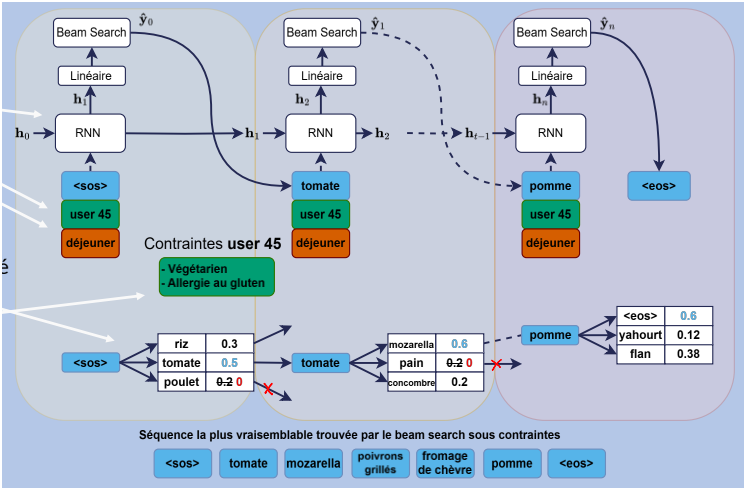
Opportunities in nutrition : modeling user preferences



(5) What about Recommender System in Nutrition?

Building consistent proposals... With expert constraints



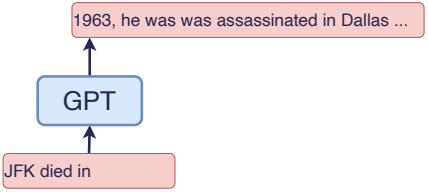


Personalized Sequence Generation in Food Recommender Systems, SAC 2026
Combeau et al.

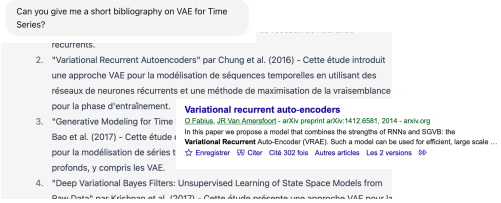
Discussion & conclusion

LLMs/ML and the relationship with truth

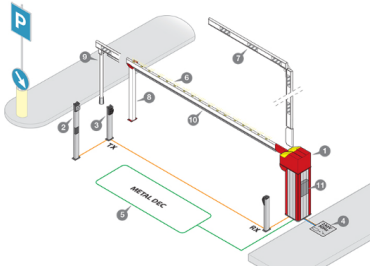
1. **Likelyhood** = grammar, agreement, tense concordance, logical sequences...
   ⇒ Repeated knowledge
2. Predict the most **plausible** word...
   ⇒ produces **hallucinations**
3. **Offline** functioning
4. chatGPT ≠ **knowledge graphs**
5. Brilliant answers...
   And silly mistakes!
   + we cannot predict the errors

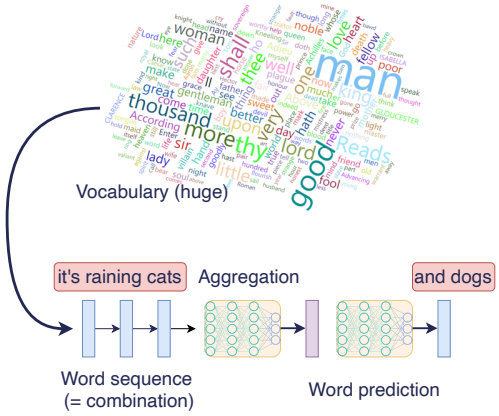


Example: producing a bibliography



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**

Machine Learning & Bias

Mustache, Triangular Ears, Fur Texture

Cat

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

Senior Executive

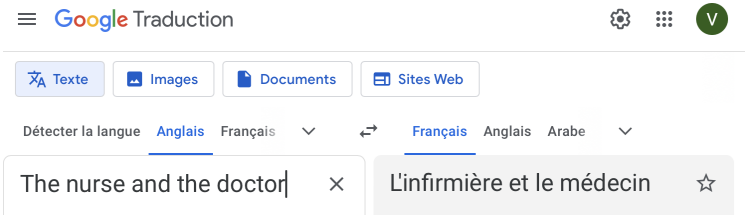
Bias in the data ⇒ bias in the responses

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:

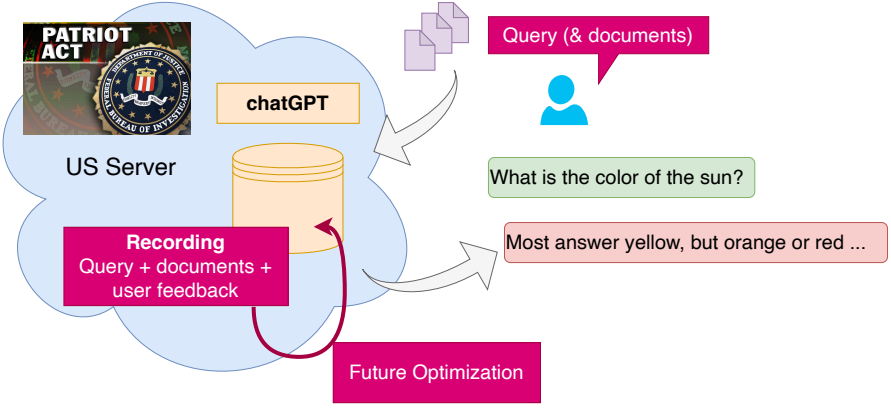
- Selection of specific data, rebalancing
- Censorship of certain information
- Censorship of algorithm results

⇒ 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?

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

Tools

Commercial tools,free to use
Variable licence

Commercial tools,
Paid licence
more guaranties vs patriot

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

Institutional LLMs
deployed within a
controlled perimeter

Local use
pre-trained/finetuned models

Data

Any document

Personal information

Ongoing project

Medical records

Conclusion

- Many opportunities (ML/DL):
  - user profiling, experimental design, BD completion
  - Frugal (or very frugal) systems
  - Easy to handle
- Many opportunities (generative AI):
  - unifying different systems (knowledge bases, user interactions, images, texts),
  - developing new interfaces (UX),
  - new applications

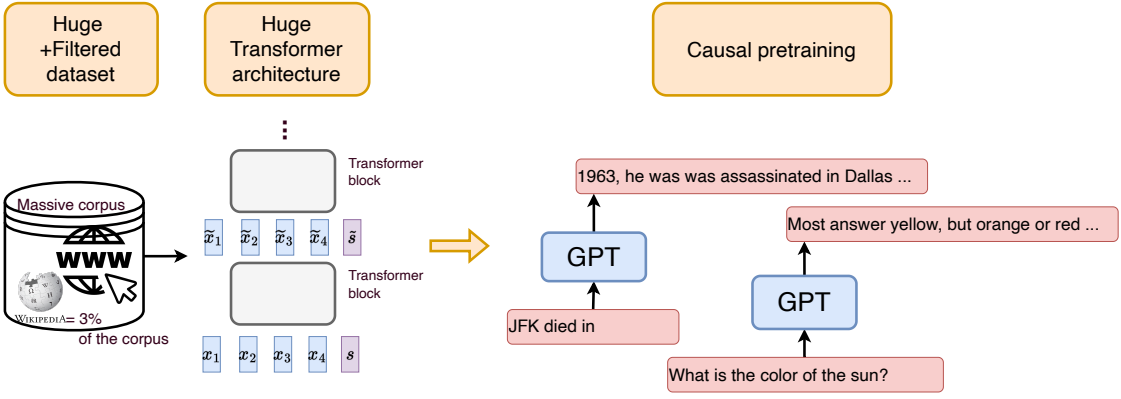
chatGPT

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)



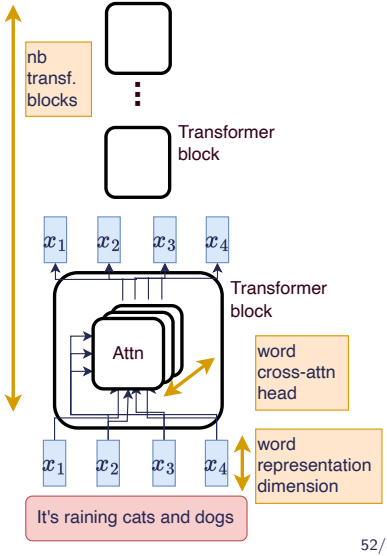
- Grammatical skills: singular/plural agreement, tense concordance
- Knowledges: entities, names, dates, places

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]

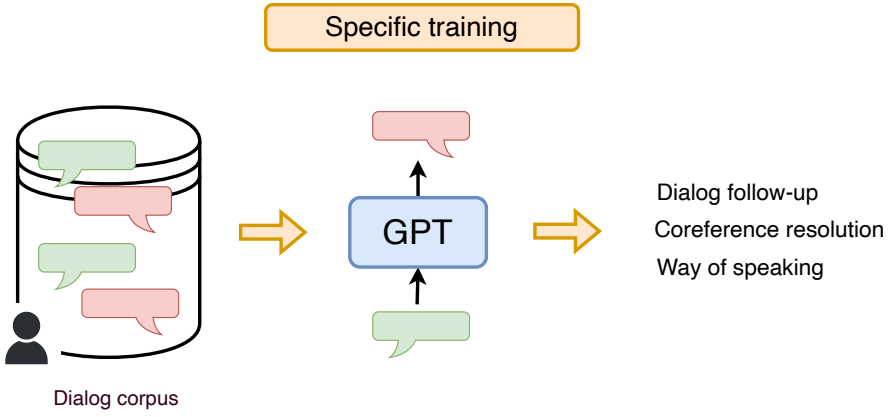


175 Billion parameters... What does it mean?

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

The Ingredients of chatGPT

2. Dialogue Tracking

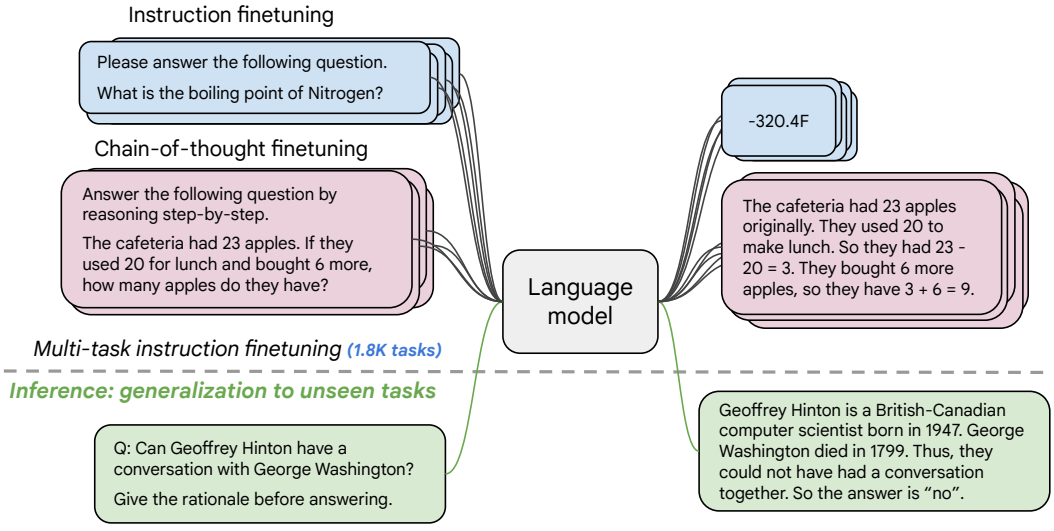


- **Very clean** data

Data generated/validated/ranked by humans

The Ingredients of chatGPT

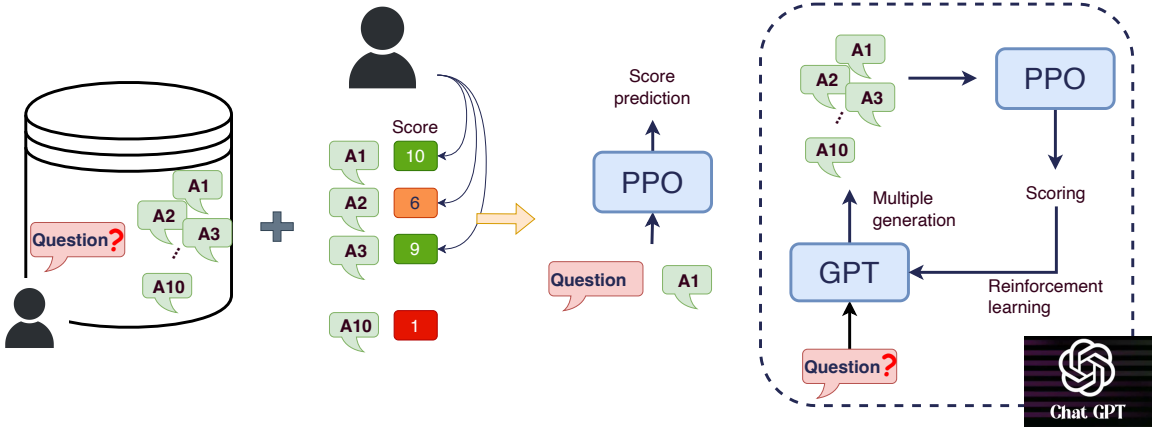
3. Fine-tuning on different (±) complex reasoning tasks



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

The Ingredients of chatGPT

4. Instructions + answer ranking



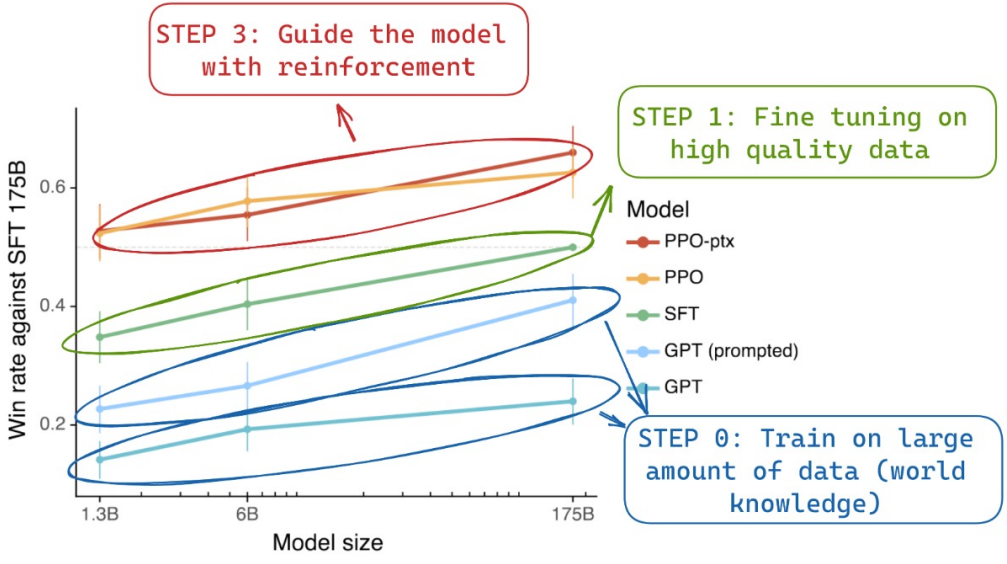
- Database created by humans
- Response improvement

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

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

Steps & Performance

Massive data ⇒ HQ data (dialogue) ⇒ Tasks ⇒ RLHF



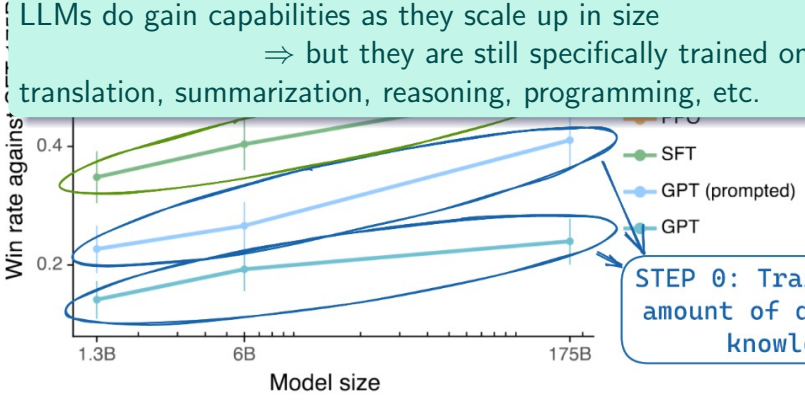
Steps & Performance

Massive data ⇒ HQ data (dialogue) ⇒ Tasks ⇒ RLHF

STEP 3: Guide the model with reinforcement

Warning about *emergent capabilities*

LLMs do gain capabilities as they scale up in size
⇒ but they are still specifically trained on all the tasks: translation, summarization, reasoning, programming, etc.



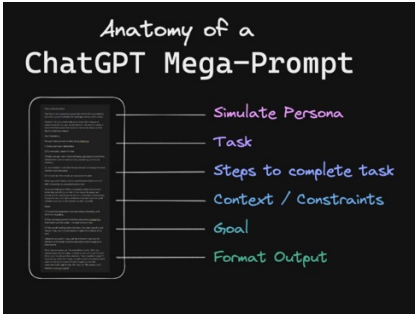
STEP 0: Train on large amount of data (world knowledge)

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/

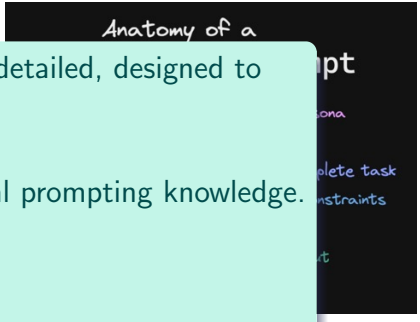
⇒ Need for **practice** [1 to 2 hours], discuss with colleagues

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 ...*
  - [illegible]
- Don'
  - [illegible]
  - [illegible]
  - [illegible]
- Reph
  - [illegible]
  - [illegible]
  - Add mistakes (!)

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

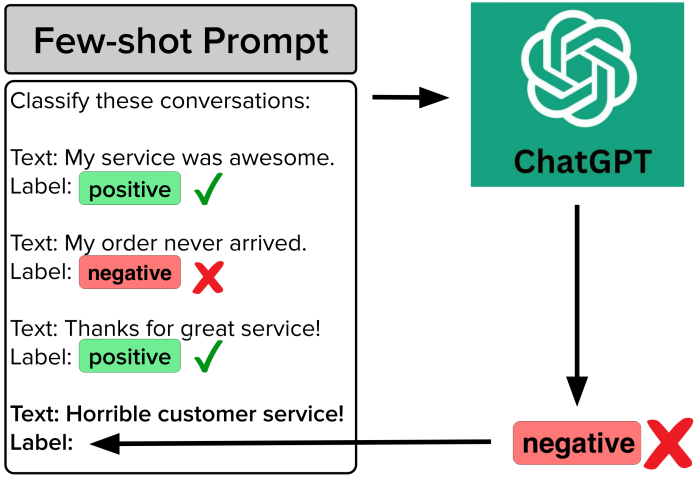
- **Prompt = specification document**: long, detailed, designed to remove ambiguity about intended directions.
- Prompting requires
  domain expertise >> general prompting knowledge.



-chatgpt-prompt/

Towards *few-shot learning*

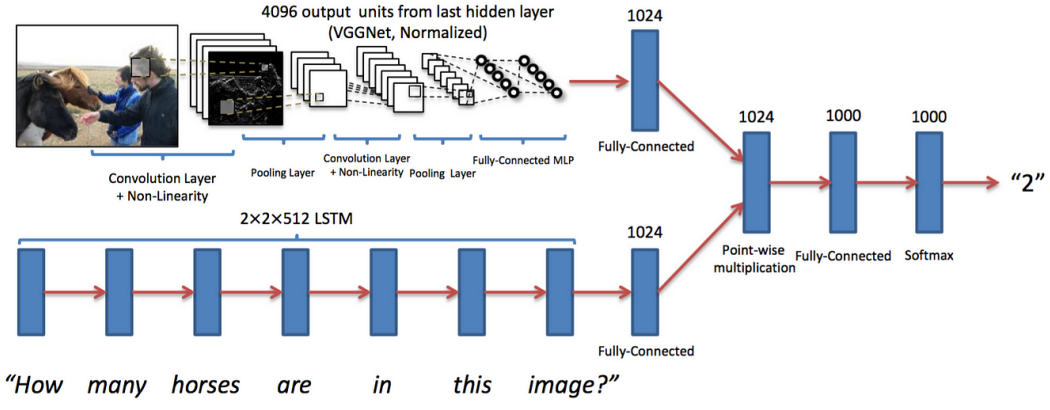
- Learning without modifying the model = examples in the prompt



GPT4 & Multimodality

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

The example of VQA: visual question answering

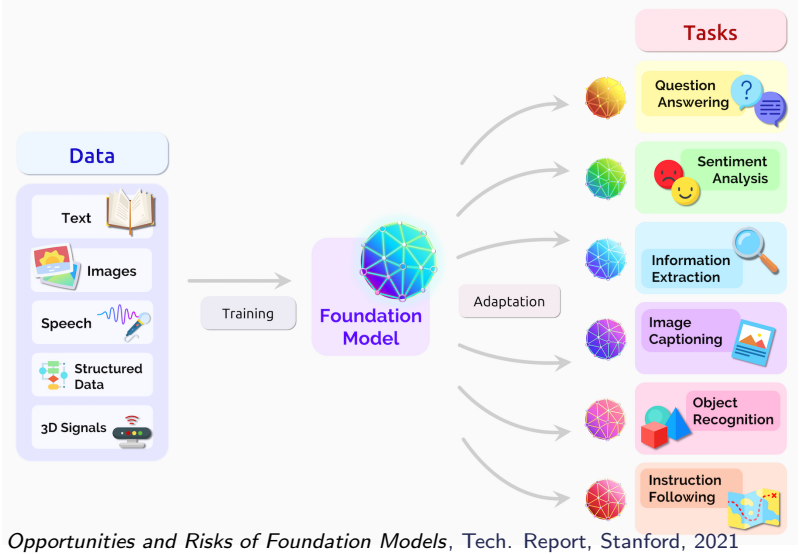


$\Rightarrow$ Backpropagate the error $\Rightarrow$ modify word representations + image analysis

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

Towards Larger Foundation Models?

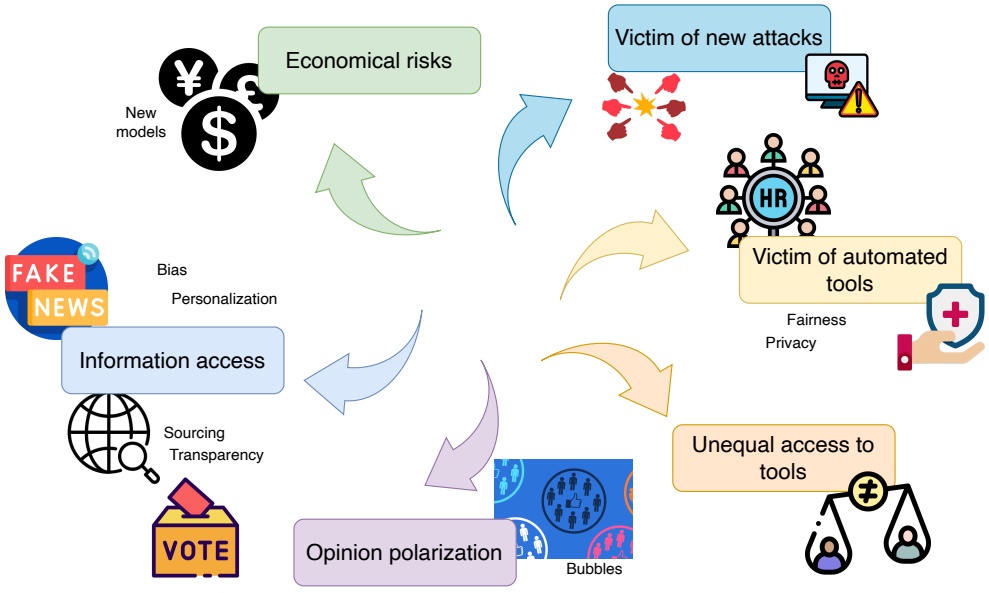
- Let the modalities enrich each other



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

(Main) Risks derived from ML & LLM

Risk topology



Access to Information

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

  $\Rightarrow$ A huge risk for democracy
- Risk of anthropomorphizing the algorithm
  - Distinguishing human from machine
  - Critical in the medical domain (i.e. robot-assisted therapy for autism)

Profiling / personalization to information access

- A good way to face **Information overload**
  - Personal filter / agent
  - Pro-active system : watch agent + notifications
- **Information bubbles** = cluster in social networks
  - Uniform point of view (tunnel vision)
  - Consensus illusion
- **Group polarization**
  - emotional reinforcement
  - Confirmation bias ↗↗↗
  - Disinformation rapid diffusion

Machine Learning & Bias

Mustache, Triangular Ears, Fur Texture

Cat

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

Senior Executive

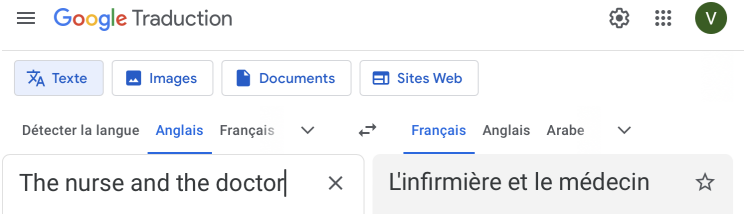
Bias in the data ⇒ bias in the responses

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 $\Rightarrow$ bias in the responses

Machine learning is based on extracting statistical biases...

$\Rightarrow$ Fighting bias = manually adjusting the algorithm

Bias Correction & Editorial Line

Bias Correction:

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

⇒ Editorial work... Done by whom?

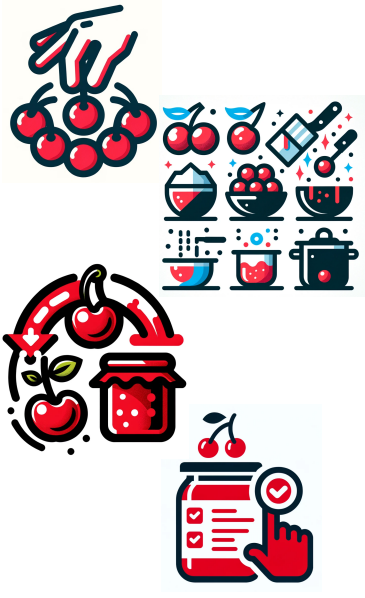
- Domain experts / specifications
- Engineers, during algorithm design
- Ethics group, during result validation
- Communication group / user response

⇒ What legitimacy? What transparency? What effectiveness?

Machine learning is never neutral

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

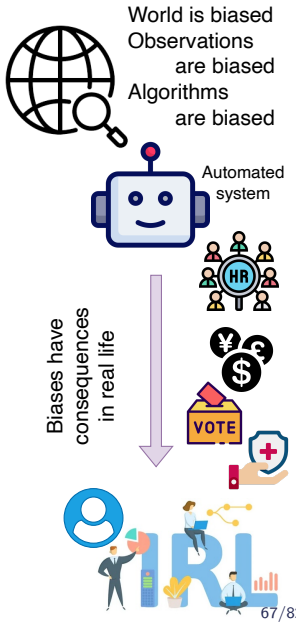
$\Rightarrow$ Choices that influence algorithm results



Bias consequences (& exploitation)

- People medical insurance cancelling (companies minimizing the risks)
- Automated (?) exclusion of certain categories of people from
  - a job offer,
  - social assistance,
  - housing,
  - credit, ...
- Targeted information stream
  - possibly without user consent (e.g. Google search portal)

  $\Rightarrow$ Create / Reinforce information bubble

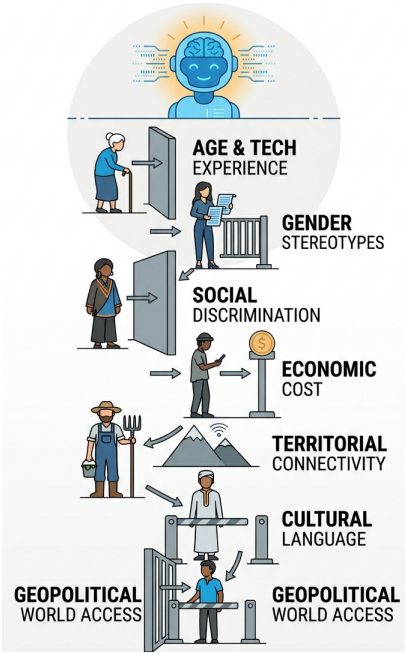


Multiplication of AI uses $\Rightarrow$ multiplication of the risks

(Human) Bias facing AI-Proposed Solutions

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

Unequal access to AI



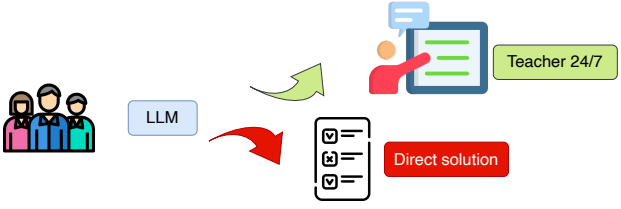
- Age gap
  - well-known for digital service
  - Mitigated through new user-friendly interfaces (?)
- Gender gap (toward computer science in general)
- Social gap
  - Access cost (especially since sept. 2025)
  - Educational ways to handle new tools

    the good, the bad, the ugly
- Territory gap
  - Internet access required
  - + information feedback, sensors density, POI, ...
- Language / culture gap
  - Tools are mainly adapted to english

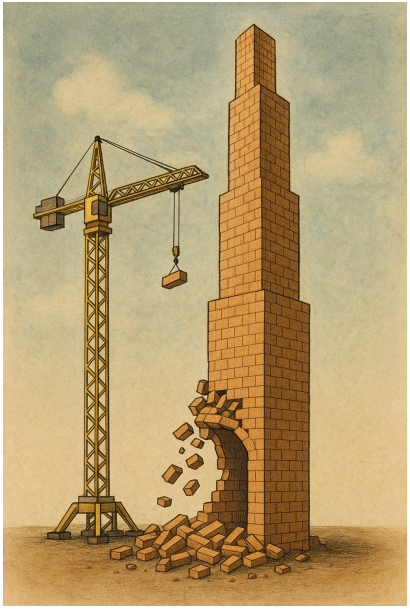
    + american culture

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
- Do not underestimate the psychological aspects



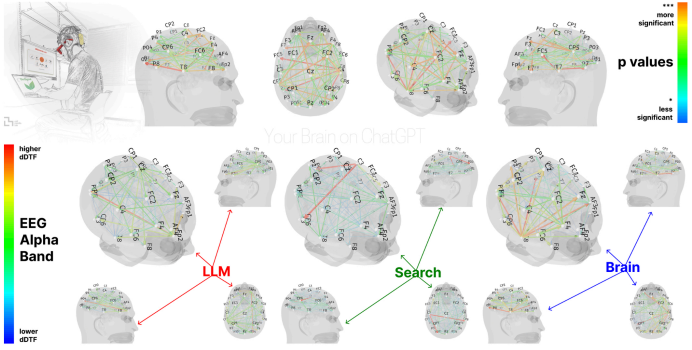
Decline / Evolution of Cognitive skills

Our brain will evolve with these new tools...

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

- Education sciences and psychology had conjectured it...

  cognitive sciences have measured it

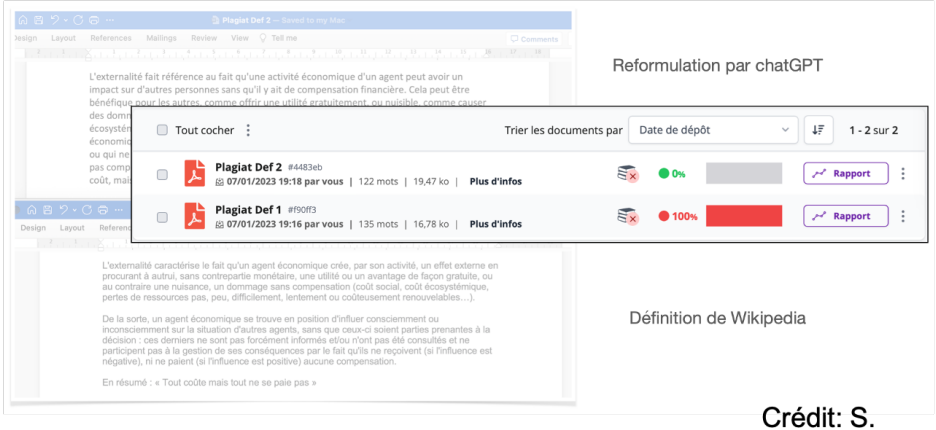


Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task, N. Kosmyna et al. arXiv 2025

Human Ressources

- **Applications** rely more & more on AI
  $\Rightarrow$ new emerging objectives (i.e. not to be mistaken with AI)
  What do you think of an job-searching-agent?
  Am I going to disappear behind massive data?
- **HR systems** rely more & more on AI
  - Matching objectives ($\approx$ recommender systems) - job center
  - Filtering application - employer
- Many risks
  - Bias risks (Amazon example, EU regulations, ...) $\Rightarrow$ **Fairness**
    - Unethical/Inefficient/illegal automatic filtering
  - But, biases are often related to preferences
    (e.g. diff. average expectations from men & women)
  - **Privacy** / who access to sensitive data, which archive policy?
    Right to be forgotten (GDPR): which guarantees?
  - LinkedIn !!! [=Information (& bias) Monopole?]

Detection of *texts generated by* chatGPT



Crédit: S. Pajak

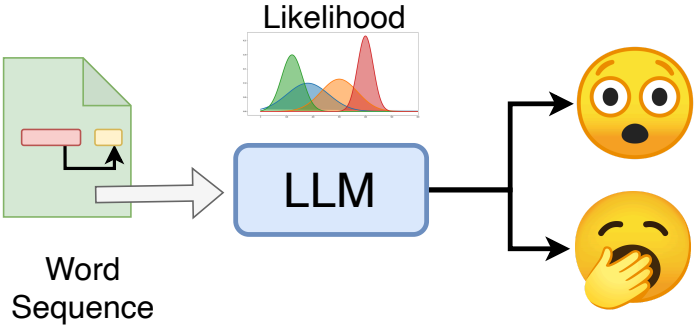
Detection of *texts generated by* chatGPT

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

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

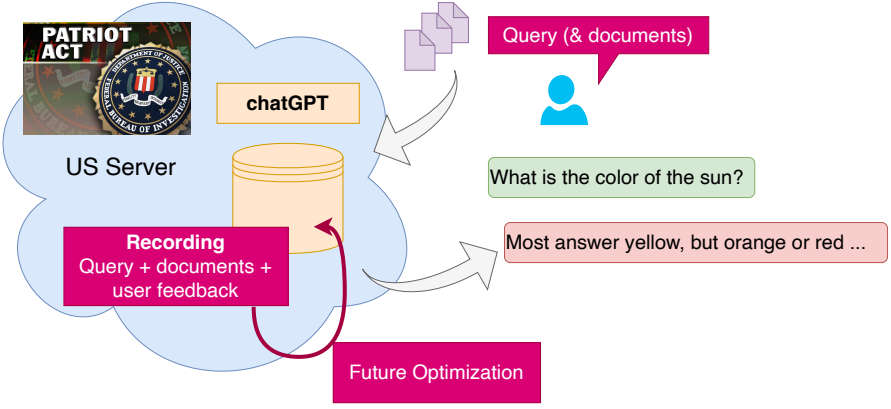
$+$ confidence level in detection

Detection of *texts used by* chatGPT



- Closed corpora $\Rightarrow$ challenge of detection of texts used in training
- Detection of **likelihood/surprise of observed word sequences**

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

Tools

Commercial tools, free to use
Variable licence

Commercial tools,
Paid licence
more guaranties vs patriot

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

Institutional LLMs
deployed within a
controlled perimeter

Local use
pre-trained/finetuned models





Data

Any document

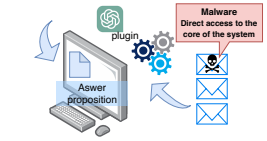
Personal information

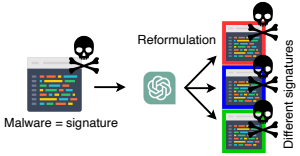
Ongoing project

Medical records

Security Issues : Attacking computers

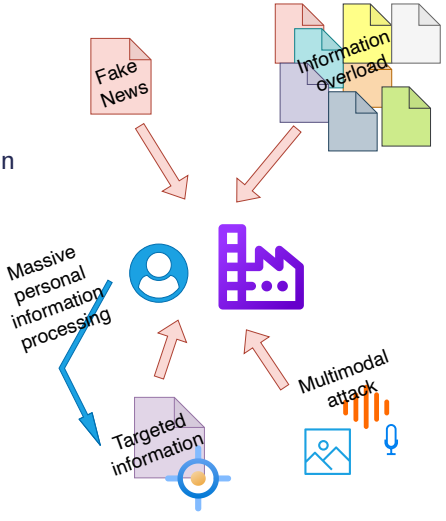
- Plug-ins / openclaw / MCP
  - $\Rightarrow$ Often significant security vulnerabilities for users
    - Email access / transfer of sensitive information etc...
- Management issues for companies
  - Securing (very) large files
- Increased opportunities for malware signatures
  - $\approx$ software rephrasing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





Attacking people & institutions

- Harassment & abuse
  - Cheap massive campaign (Scalable bullying, automated "pile-ons")
  - Non-consensual sexual content generation (deepfake pornography)
  - Doxxing and sensitive data inference + publication
- Identity falsification
  - Voice/text impersonation
  - Fake authority impersonation (boss, teacher, lawyer, doctor)
- Social engineering at scale
  - Phishing messages that are highly personalized
- Informational harms
  - Misinformation & disinformation
  - Decision sabotage (finance, health, safety)
  - Integrity attacks (e.g. scientific, journalist)



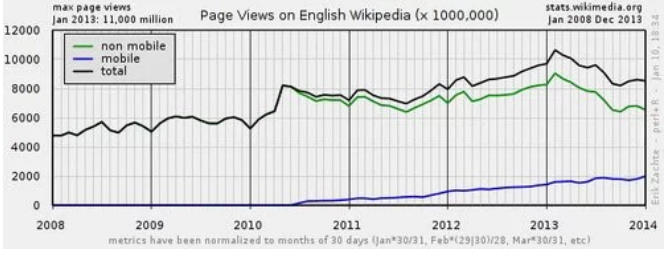
Economic Questions

- 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?

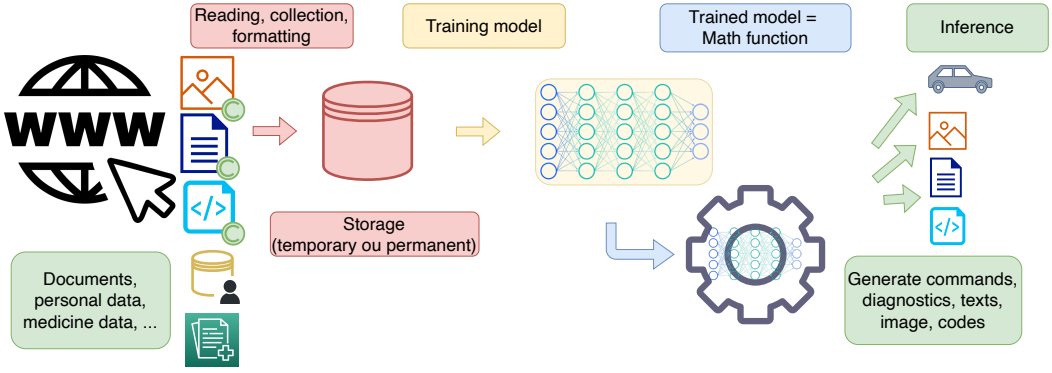
Google's Knowledge Graph Boxes: killing Wikipedia?

by Gregory Kohs



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

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

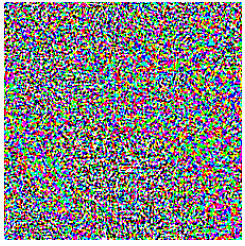
Attacking the algorithm

If an algorithm takes critical decision, it can be attacked !

$\boldsymbol{x}$

"panda"
57.7% confidence

$+ .007 \times$



$\text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$

"nematode"
8.2% confidence

$=$

$\boldsymbol{x} + \epsilon\text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$

"gibbon"
99.3 % confidence

Explaining and Harnessing Adversarial Examples, ICLR 2015, Goodfellow et al.

Attacking the algorithm

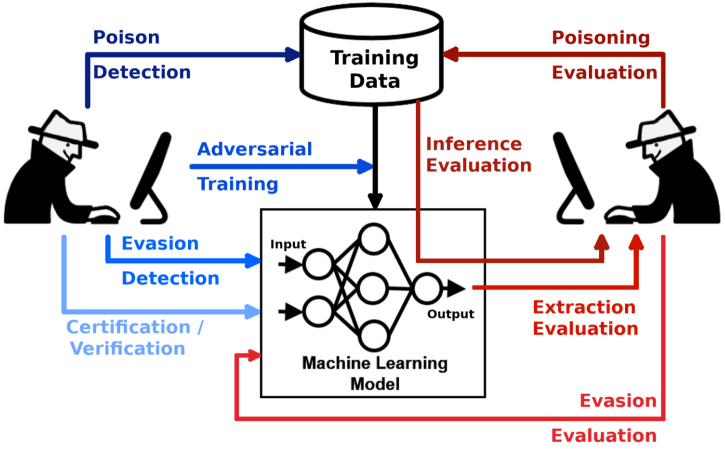
If an algorithm takes critical decision, it can be attacked !

Justin Johnson, Stanford CS231n

Attacking the algorithm

If an algorithm takes critical decision, it can be attacked !

A typology to attack ML algorithms



Attacking data / diag

Knowing the model / gradient / nothing

How to protect?

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

Risks of AI Generalization

AI everywhere =
loss of meaning?

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