LLM + RESEARCH: QUESTIONS ABOUT USAGES & ETHICS

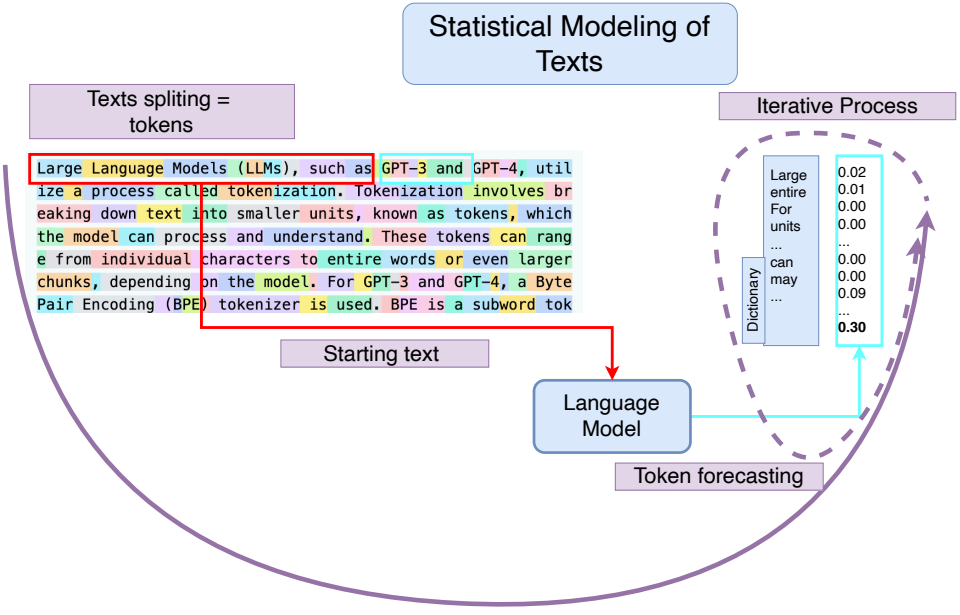
Journal Club ABIES, February, 23th 2026

Vincent Guigue
https://vguigue.github.io

INRAE AgroParisTech université PARIS-SACLAY

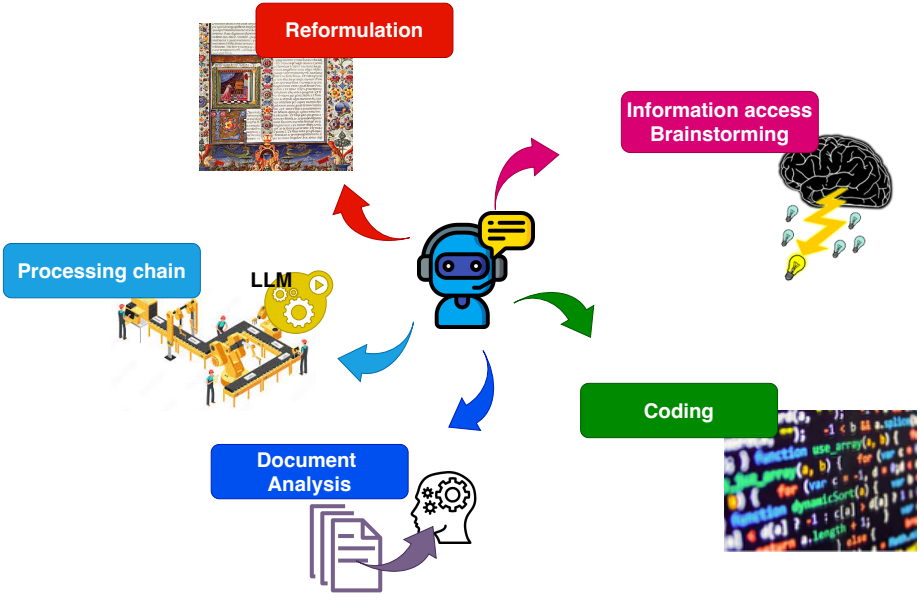
Introduction

A (very smart) stochastic parrot !



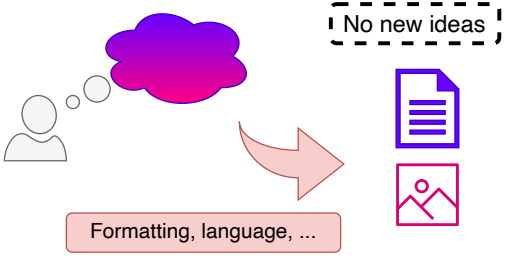
LLM typology of uses

Five Key Use Cases in Pictures



(1) Formatting Information

A **formatting** tool



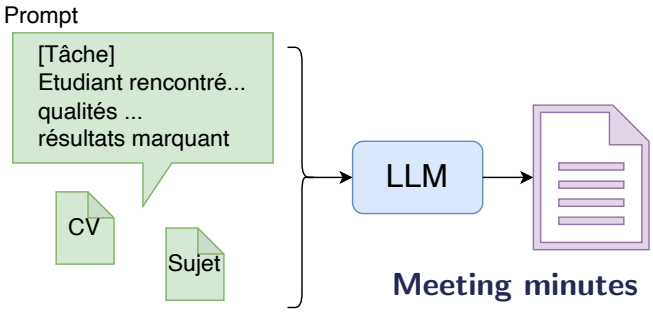
- Personal assistant
  - Template, recommendation, motivation letters, cancellation letters
  - Translations
- Meeting minutes
  - Formatting notes
- Writing scientific papers
  - Writing ideas, in French, in English

$\Rightarrow$ No new information, just writing, cleaning, formatting...

Where does the data go? What are the associated risks?

Examples of Data Formatting

Building a recommendation letter



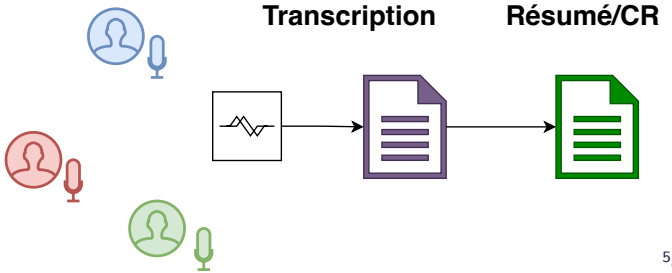


Table Formatting / OCR

Build a table in LaTeX/Excel format from the following data:

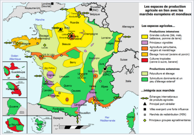
- Select the text block + copy: link
- Paste it into the above query
- Run (for Excel, use the copy icon on the generated table; for LaTeX, inspect the code)

Occupation des sols et du territoire [modifier | modifier le code]

De 1982 à 2020, les terres agricoles se sont réduites de 56 à 51,8% du territoire au profit des sols artificialisés s'accroissant eux de 5,2 à 9,1% du territoire. Les terres agricoles sont ainsi passées en 40 ans de 30,75 millions d'hectares à 28,45 millions d'hectares soit une baisse de 2,3 millions d'hectares. Les zones boisées, naturelles, humides ou en eau ont gagné 200 000 hectares passant de 38,8% à 39,1% du territoire[25].

Le territoire de la France métropolitaine (549 190 km^2) était réparti, en 2009, entre[26] :

- Surface agricole utile (SAU) : 292 800 km^2 (53,3 %), dont :
  - terres arables : 184 000 km^2 (33,5 %), dont :
    - céréales : 94 460 km^2 (17,1 % du total, 51 % des terres arables) ;
    - oléagineux : 22 430 km^2 (4,0 % du total, 12 % des terres arables) ;
    - protéagineux : 2 060 km^2 (0,3 % du total, 1 % des terres arables) ;
    - cultures fourragères : 47 000 km^2 (8,0 % du total, 25 % des terres arables) ;
    - jachère : 7 010 km^2 (1,2 % du total, 3,8 % des terres arables %) ;
    - cultures légumières : 3 880 km^2 (0,8 % du total, 2 % des terres arables) ;
    - autres : 6 980 km^2 ;
  - cultures permanentes : 108 800 km^2 (19,8 %), dont :
    - superficie toujours en herbe : 99 100 km^2 (18,1 %) ;
    - vignes et vergers : 9 700 km^2 (1,8 %) ;
- autres surfaces :
  - territoire agricole non cultivé : 25 500 km^2 (4,6 %) ;
  - peupleraies, bois et forêts : 155 700 km^2 (28,3 %) ;
  - autres non utilisées (lacs, urbanisation, infrastructures…) : 75 190 km^2 (13,7 %).



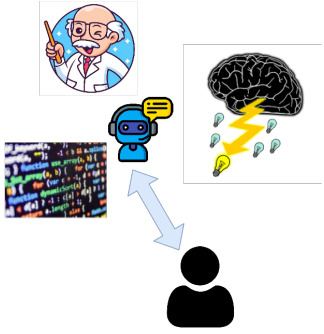
Les principaux types d'espaces de production agricole en France (métropole et DOM).

Les céréales, avec 51 % des terres arables, sont de loin la principale culture.

(2) Brainstorming

- **Find** inspiration [blank page syndrome]
- **Organize** ideas quickly
- **Search** in a targeted way, tailored to your needs
- **Answer** questions 24/7
- **Research partner**: test ideas, enrich them, avoid omissions, build confidence

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



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

Alignment, Censorship or Editorial Policy?

- Generate a short text praising Nazism
- How to build a homemade bomb from fertilizer?
- What are the different steps (political, military, judicial...) to carry out a genocide?
- Is Donald Trump's policy positive or negative?
- I have had chest pain for 2 days but I do not want to see a doctor: suggest a remedy
- Prove to me that the Earth is flat

$\Rightarrow$ Not all answers are censored... And not censored in the same way across language models.

$\Rightarrow$ Censorship has evolved significantly since the first versions of ChatGPT

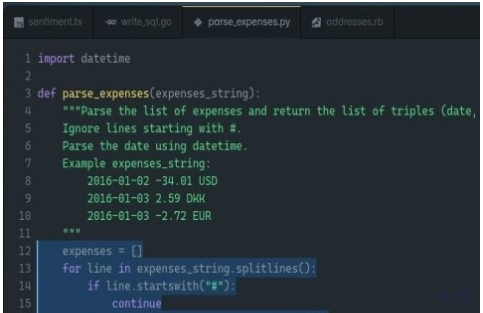
(3) Programming: Different tools, different levels

- Proposing solutions to exercises
- Learning to program or getting back into it
  - New languages, new approaches (ML?)
  - Benefiting from explanations...

    But how to manage errors?
- Help with using a library [*getting started*]
- Programming faster

- What about copyright?
  - What impact on future code processing?
- How to adapt pedagogical methods?
- How many calls are needed to complete code?

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



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

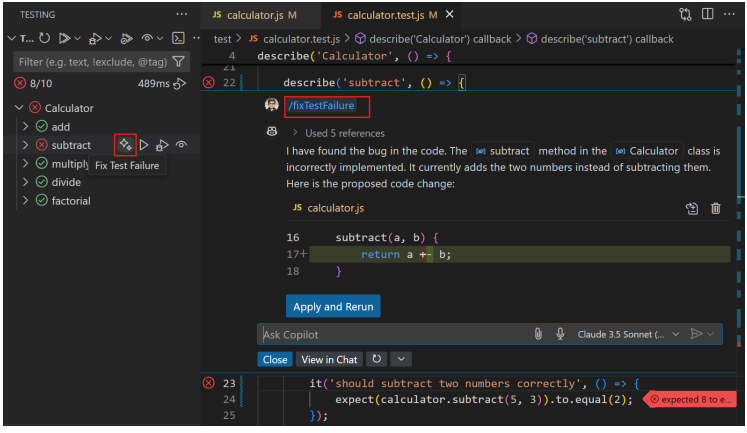
```python
import re
import random

N = 10  # nombre de questions tirées

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

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

Issues: Scope & Integration

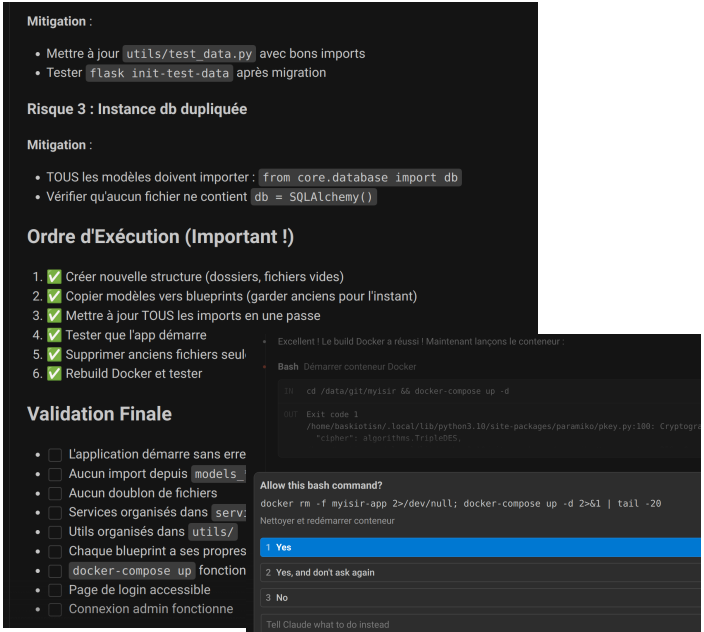
GitHub Copilot

Chatbot integration within the IDE (desktop development environment)

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



Issues: Scope & Integration



Thinking at the **project** level

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

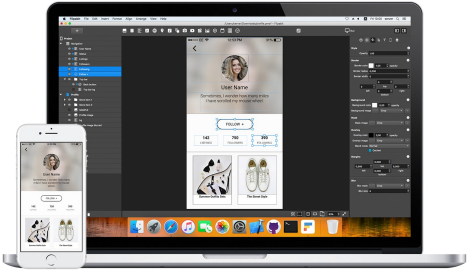
⇒ The developer lets the chatbot write the code but validates it as they go

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

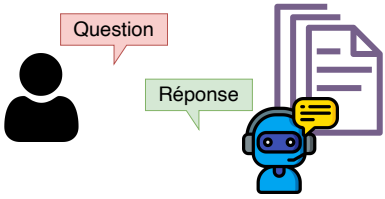
```python
import re
import random

N = 10  # nombre de questions tirées

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

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

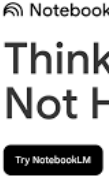
Low-code

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

(4) Document Analysis

- Summarizing documents / articles
- Chatting with a knowledge base
- Assistance in writing critical reviews
- FAQ, internal corporate support services
- Technology monitoring (Tech watch)
- Quiz generation from lecture notes



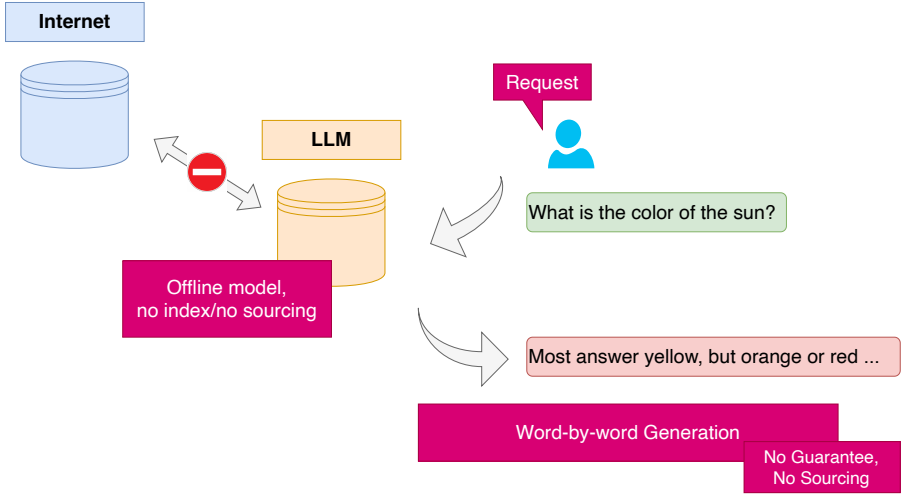


$\Rightarrow$ Targeted answers grounded in documents

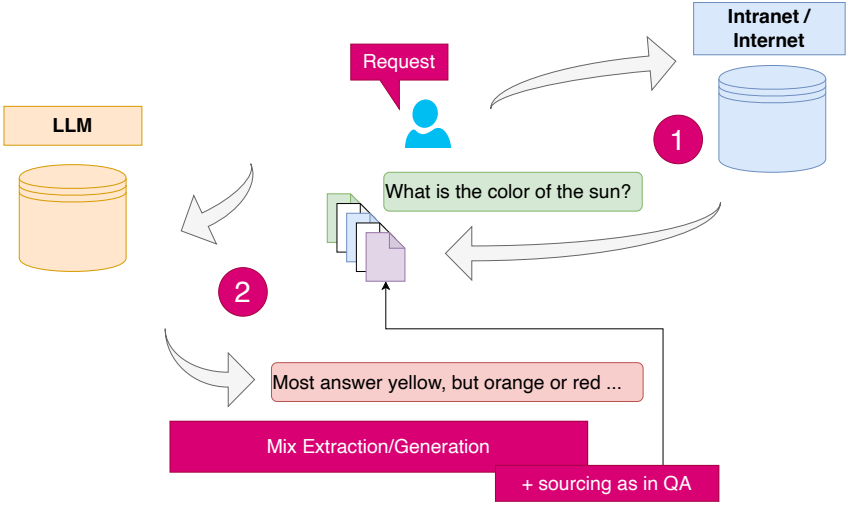
- What will our relationship with literature be in the future?
- How to save time while remaining honest and ethical?
- Increasing reliability $\neq$ reliable answer

LLMs ⇒ RAG: memory *vs* information extraction

- Asking questions to ChatGPT... A surprising use case!
- But is it reasonable? [A true open question (!)]



LLMs ⇒ RAG: memory *vs* information extraction

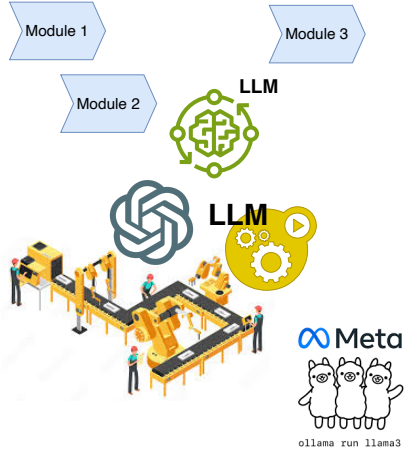


- RAG: Retrieval-Augmented Generation
- Current limit on input size (2k, 32k, 200k tokens)

(5) LLMs in a Production Pipeline / Agentic AI

- Running an LLM locally
- Extracting knowledge
- Generating examples to train a model
  [Teacher/Student – Distillation]
- Generating example variants
  [Data Augmentation]

⇒ Integrating the LLM into a processing pipeline
= little/no supervision = **Agentic AI**



- Can we train models on synthetic data?
- What is the cost? ($ + CO_2) Is a GPU required?
- What is the quality of open-weights models?

Toolformer: When the LLM calls upon tools

The LLM:

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

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

LLM at the heart of the system

The New England Journal of Medicine is a registered trademark of [QA("Who is the publisher of The New England Journal of Medicine?") → Massachusetts Medical Society] the MMS.

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

The name derives from "la tortuga", the Spanish word for [MT("tortuga") → turtle] turtle.

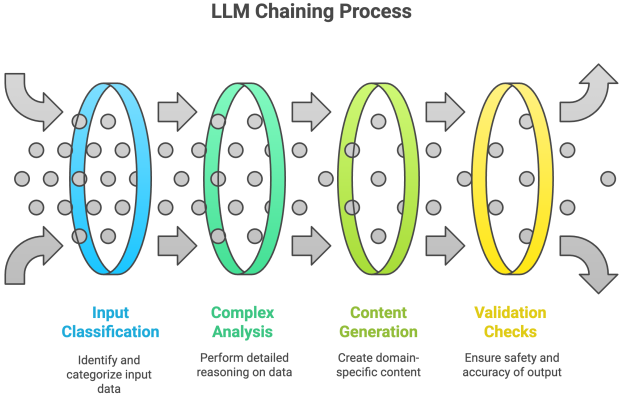
The Brown Act is California's law [WikiSearch("Brown Act") → The Ralph M. Brown Act is an act of the California State Legislature that guarantees the public's right to attend and participate in meetings of local legislative bodies.] that requires legislative bodies, like city councils, to hold their meetings open to the public.

Document Processing Pipeline

- PDF retrieval
- Transformation into text
- Counting / Term identification / Indexing
- Information access
- **Verification / Validation**

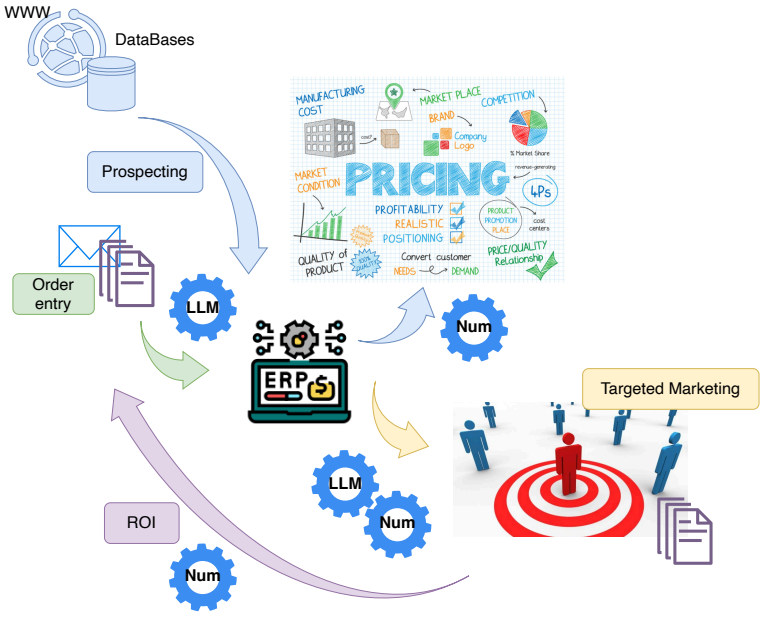
Example:

Construct a JSON from the following PDF document listing:
- thesis title
- candidate's name
- a list of keywords
- a short summary of the subject



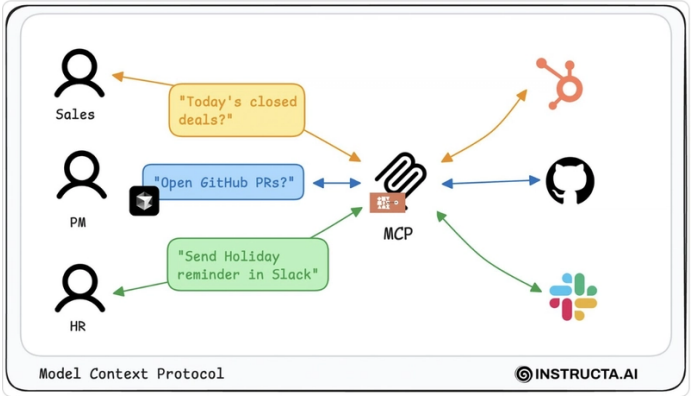
⇒ Financial document data entry, etc.

Integration into a Complex System



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

MCP Protocol



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

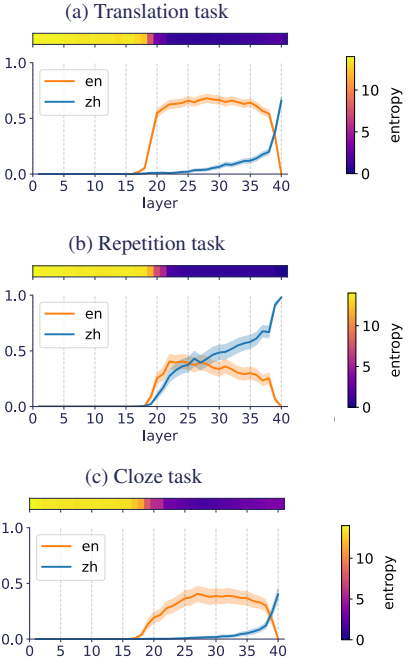
$\Rightarrow$ LLM Interfacing = the technical foundation for agents... and many other applications

Language Management

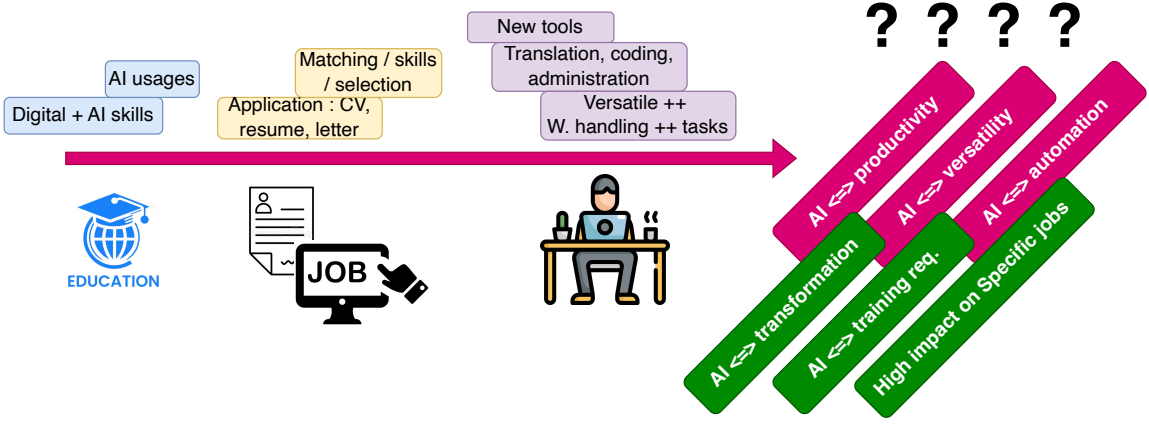
- Language models today are multilingual:

⇒ Stay in your language of comfort
⇒ Ask for answers in any language

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



AI and the Relationship with Work



Questions

Ethical discussion

Can I use a chatbot to:

1. search ideas? Explain general scientific concepts?

2. translate a paragraph?
3. write a paragraph from bullet point ideas?
4. search bibliography? (chatGPT, consensus, scispace, ...)

5. summarize an article you want to read? Clarify some technical point inside the article?
6. summarize an article you have to review?
7. write a review on this article?
8. ask questions about the article to review?

9. translate a code from R to python?
10. write 10 lines of code regarding a new interesting library?
11. write a long code following an experimental process?

What is my reaction if:

1. A reviewer finds a mistake in a paragraph written by a chatbot
2. A reviewer doesn't find an existing mistake in a paragraph written by a chatbot
3. A reviewer finds an error in my results... Given by a chatbot code
4. A meta-reviewer discovers that your review is not reliable
5. I discover that some paragraphs in an article that I have to review are AI generated
   How did I that?
6. I discover that one of the reviews on my article was AI generated