Bureaucracy AI

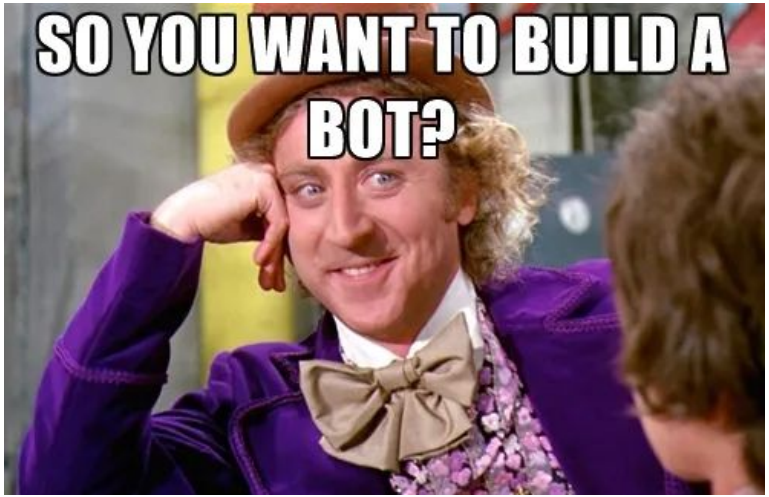
Cappellini Alessio **1693930**

Cataldi Bruno **2026604**

Sirico Giuseppe Gabriele **1810153**

Idea

- New concept of smart law assistant
- Support in everyday life use dealing with EU bureaucracy
- Find a way out of the existing tangle of laws



PROs

- Grant easy access to a complex domain to everyone
- Overcome cultural and legal barriers
- User-friendliness of the chatbot

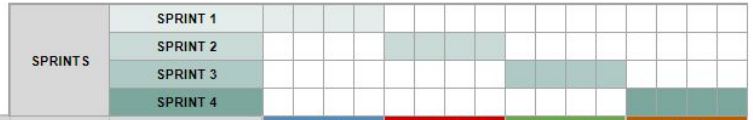
CONs

- Very high complexity of the overall ecosystem
- High level of sparse knowledge
- Very hard to achieve high quality data and accuracy

Development

GANTT chart and architecture

GANTT chart



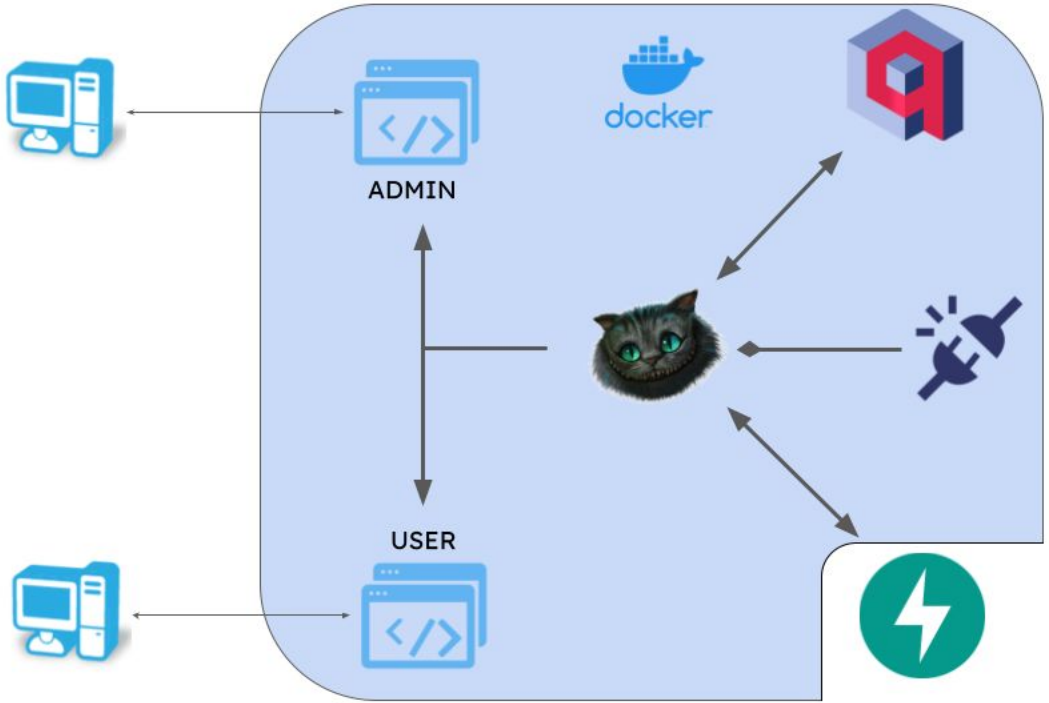
WORK BREAKDOWN STRUCTURE	TASK TITLE	TASK OWNER	AMOUNT OF WORK IN HOURS			SPRINT	START DATE	DUE DATE	DURATION	PCT OF TASK COMPLETE
			ESTIMATE	COMPLETED	REMAINING					
1	General Infrastructure		11	11	0					100%
1.1	Docker setup	Gabriele S. - Bruno C.	6	6	0	1	9/2/2023	9/3/2023	2	100%
1.2	Cheshire setup	Alessio C.	5	5	0	1	9/4/2023	9/6/2023	3	100%
2	Plugin		19	19	0					100%
2.1	Plugin set-up	Bruno C.	5	5	0	2	9/9/2023	9/10/23	2	100%
2.2	Plugin creation	Gabriele S.	8	8	0	2	9/10/2023	9/11/2023	2	100%
2.3	Plugin integration	ALL	6	6	0	2	9/11/2023	9/13/2023	3	100%
3	Backend		38	38	0					100%
3.1	Tesseract	Alessio C.	8	8	0	3	9/16/2023	9/16/2023	1	100%
3.2	Backend set up	Gabriele S. - Bruno C.	12	12	0	3	9/17/2023	9/18/2023	2	100%
3.2.1	Backend configuration	Gabriele S. - Bruno C.	18	18	0	3	9/18/2023	9/20/2023	3	100%
4	Testing		15	15	0					100%
4.1	OCR plugin test	Alessio C.	9	9	0	4	9/23/2023	9/24/2023	2	100%
4.2	Updates	Alessio C.- Gabriele S.	2	2	0	4	9/25/2023	9/25/2023	1	100%
4.3	General tests	ALL	4	4	0	4	9/25/2023	9/27/2023	3	100%
			ESTIMATE	COMPLETED	REMAINING	DAYS	EST/DAYS			
BURNDOWN D		TOTAL HOURS	83	83	0	60	1.38333333			

DAY	1	2	3	5	6	7	8	10	11	12	13	15	16	17	18	20	TOTAL HOURS
PLAN	83	82	80	77	76	75	73	71	69	68	66	64	62	61	59	57	
ESTIMATE	83	80	77	73	70	66	60	54	49	44	34	22	16	11	7	4	903
HRS COMPLETED	3	3	4	3	4	6	6	5	5	10	12	6	5	4	3	4	83
HRS REMAINING	80	77	73	70	66	60	54	49	44	34	22	16	11	7	4	0	820

Architecture

The system can be divided in 2 macro components:

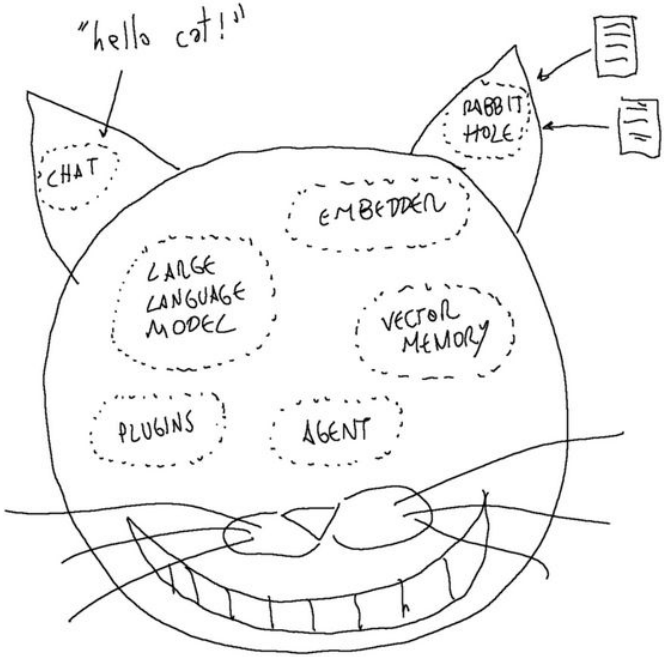
- Cheshire Cat Environment
- Fast API + LLAMA2 instance



Technologies

Cheshire Cat

- Framework to build custom AI
- Long term memory embedder
- Extensible via plugin and able to digest documents



Rabbit hole

This module takes care of ingesting resources and storing them as local memory in the form of vectors and loaded as needed.

- *Episodic Memory*, contains user conversations with the AI
- *Declarative Memory*, contains uploaded resources content, such as documents
- *Procedural Memory*, contains the information related to tools and hooks in order to make the Cat conscious of them.

Vector Database

- It’s a database that stores information as vector embeddings.
- The distance between each vector embedding is what enables the determination of similarity between vectors
- We used Qdrant to implement this functionality

Optical Character Recognition

- This approach is used to convert virtually any kind of image containing written text into machine-readable text data

- Tesseract is an open-source neural net (LSTM) based OCR engine focused on line recognition

- **Best practises:**
  - 300- 600 DPI at least
  - Applying different pre-processing techniques
    - De-noising image
    - Increase image sharpness
    - etc.
  - Grayscaling images
  - Cropping images borders

- **Alternatives:**
  - docTR
  - Attention-based OCR
  - kraken

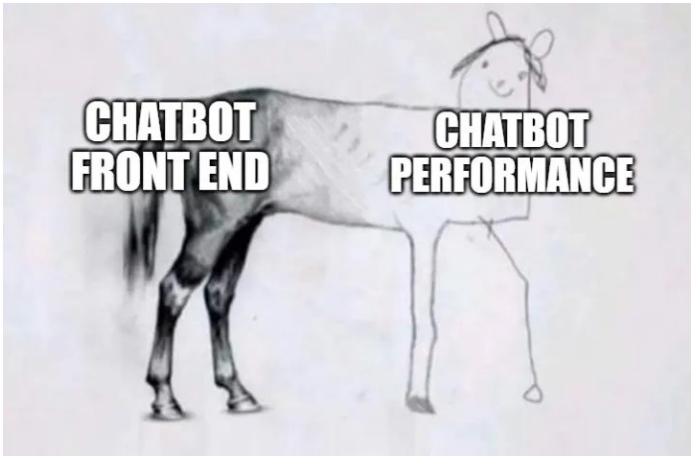
Fast API + Llama 2

- High-level Python micro web framework wrapping Llama 2 model
- Using CTransformers library, it makes the pre-trained model available to the Cheshire Cat
- Transformers models implemented using GGML library

Conclusions

All that glitters is not gold

- AI fine tuning on documents
- Data quality given by OCR approach
- Setup server for custom LLM
- Lack of authentication management
- Slow response speed



Future work

- OCR improvement
- Replace Cheshire Cat framework
- Train AI on documents
- Improve response speed

Demo