Hyland

FOSDEM'24

Using Generative AI and Content Service Platforms together

Angel Borroy

Developer Evangelist

LangChain

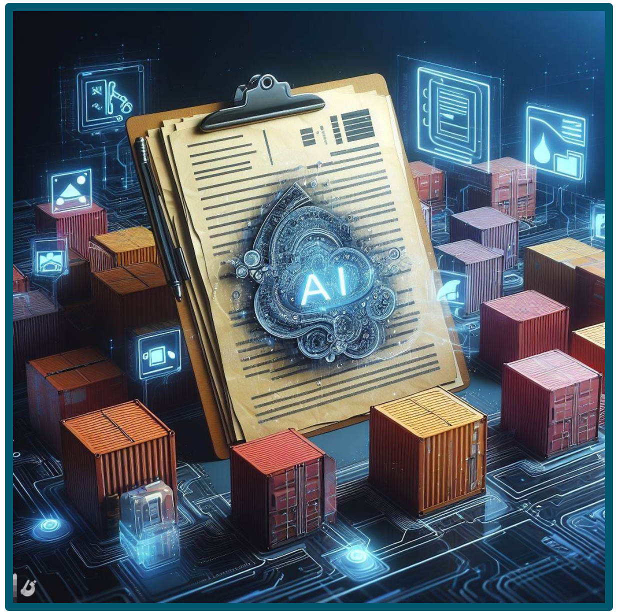
neo4j

docker

February 1, 2024

Agenda

- GenAI Stack
- Features
- Integration
  - Existing Content
  - New Content
- What Else?



Content credentials
Generated with AI · 25 January 2024 at 10:09 am

GenAI Stack

Docker GenAI

Components

[ollama]

- Local management of *open source* LLMs
- Catalog of preconfigured LLMs, such as Llama2 or Mistral

neo4j

- Graph and native vector search capabilities
- Ground LLMs for more precise GenAI predictions and outcomes

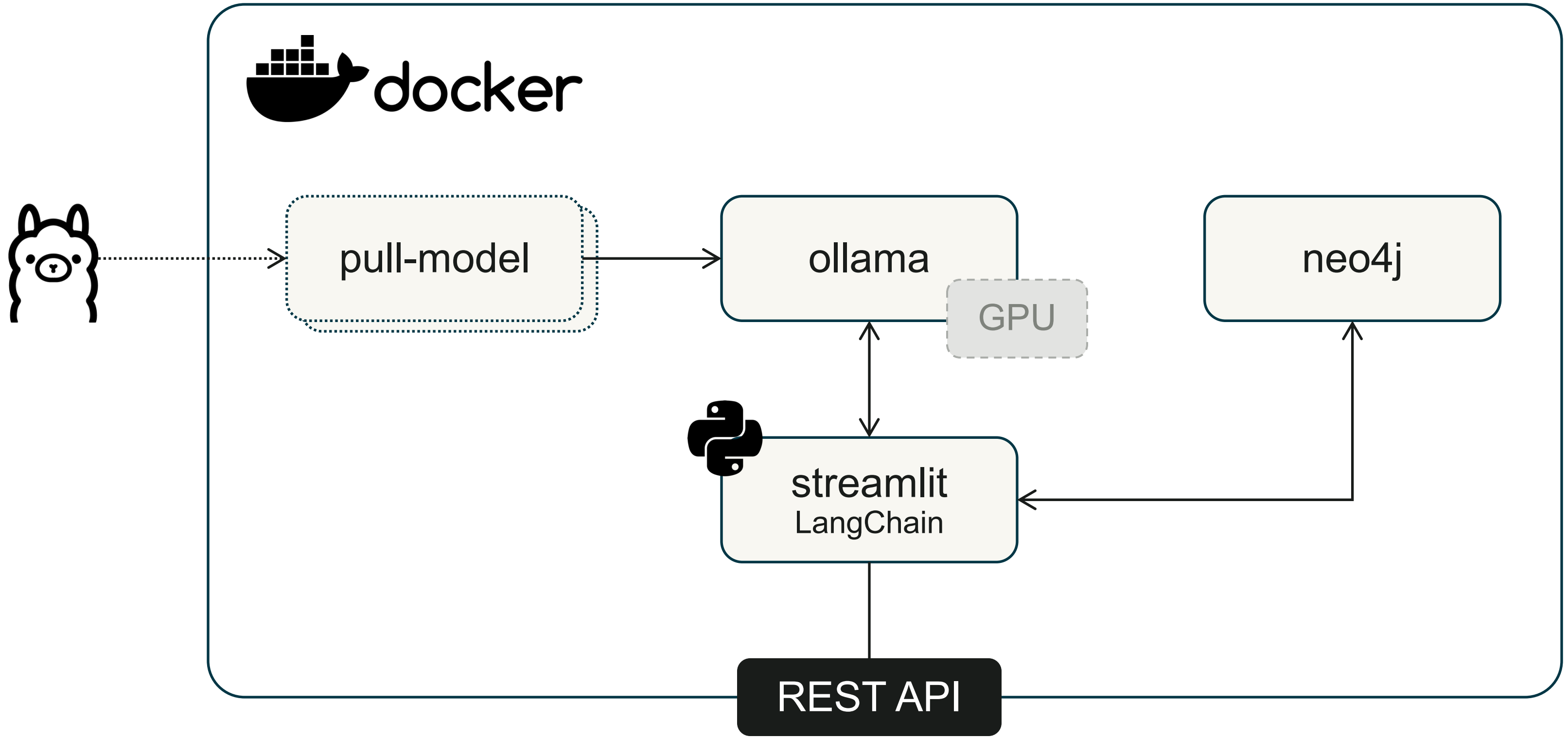
LangChain

- Communication between the LLM, your application, and the database
- Python framework for developing applications powered by LLMs

Docker GenAI

Deployment

https://github.com/docker/genai-stack



Docker GenAI

https://ollama.ai/library

LLM

Meta · MISTRAL AI_ · Technology Innovation Institute · deepseek coder

13⁺B

40+ GB Size · 64+ GB RAM

llama2	Llama2 Community License Agreement
codellama	Llama2 Community License Agreement
mixtral	Apache License 2.0
deepseek-coder	Deepseek License Agreement
vicuna	Llama2 Community License Agreement
falcon	Falcon TII License 1.0

Microsoft · deepseek coder

3B

4 GB RAM · 2 GB Size

phi	MIT License
dolphin-phi	MIT License
orca-mini	cc-by-nc-sa-4.0
deepseek-coder	Deepseek License Agreement

13B

8 GB Size · 16 GB RAM

llama2	Llama2 Community License Agreement
codellama	Llama2 Community License Agreement
vicuna	Llama2 Community License Agreement
orca-mini	cc-by-nc-sa-4.0
llava	Apache License 2.0
orca2	Microsoft Research License

Meta · Microsoft

7B

8 GB RAM · 4 GB Size

llama2	Llama2 Community License Agreement
codellama	Llama2 Community License Agreement
vicuna	Llama2 Community License Agreement
mistral	Apache License 2.0
mistral-openorca	Apache License 2.0
llava	Apache License 2.0
orca-mini	cc-by-nc-sa-4.0
deepseek-coder	Deepseek License Agreement
orca2	Microsoft Research License
falcon	Falcon TII License 1.0

Meta · MISTRAL AI_ · Technology Innovation Institute · deepseek coder

Ethical AI Rating

Mistral 7B

Nextcloud

Is the software (both for inferencing and training) open source?

https://github.com/mistralai/mistral-src

Is the trained model freely available for self-hosting?

https://ollama.ai/library/mistral

Is the training data available and free to use?

No

Ethical AI Rating

LlaVA with Visual Encoder

Nextcloud

- Is the software (both for inferencing and training) open source?
  https://github.com/haotian-liu/LLaVA
- Is the trained model freely available for self-hosting?
  https://ollama.ai/library/llava
- Is the training data available and free to use?
  https://github.com/haotian-liu/LLaVA/blob/main/docs/Data.md

MacBook Pro 2021

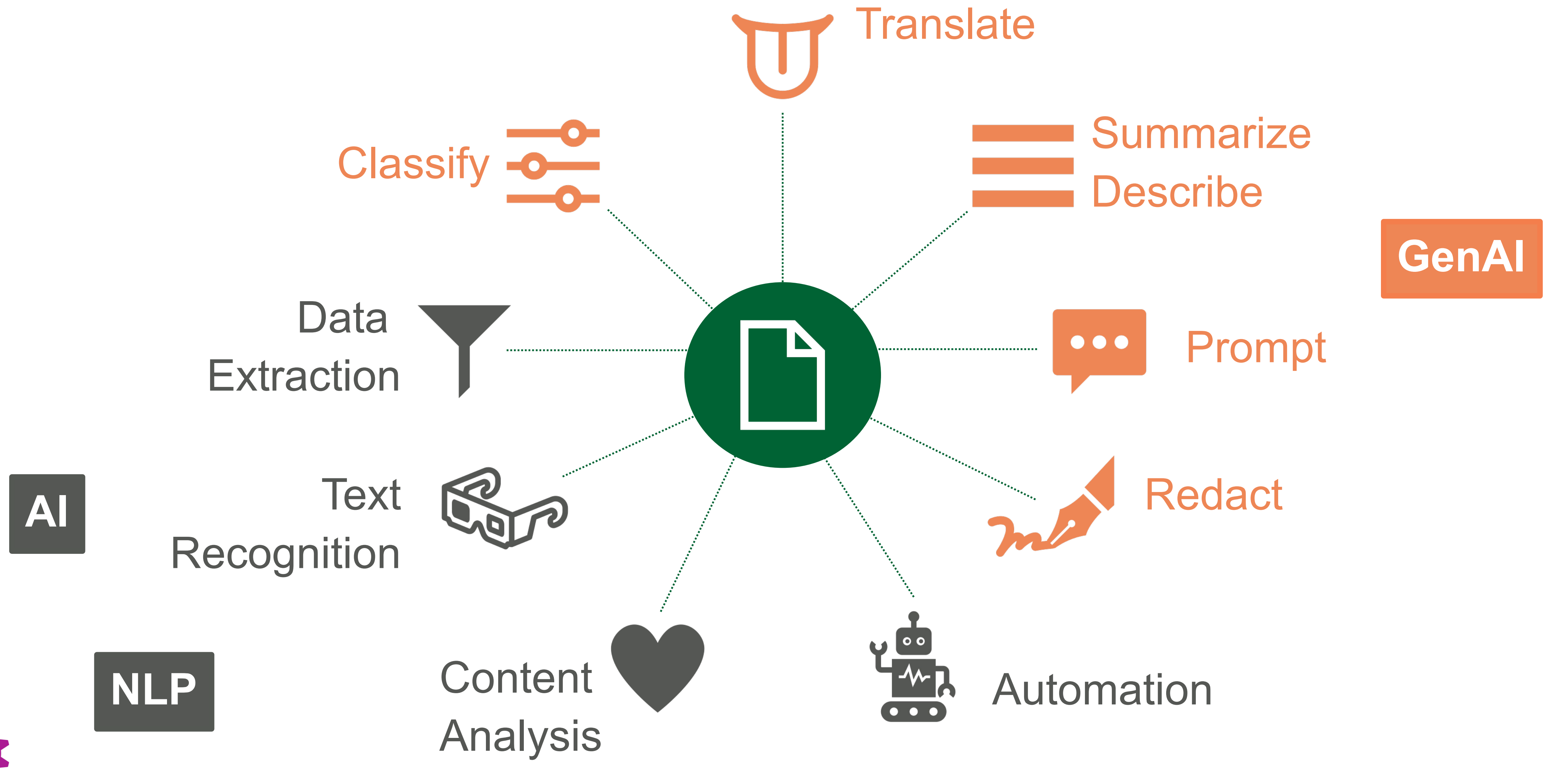
Apple M1 Pro

32 GiB RAM

~~GPU~~

Features

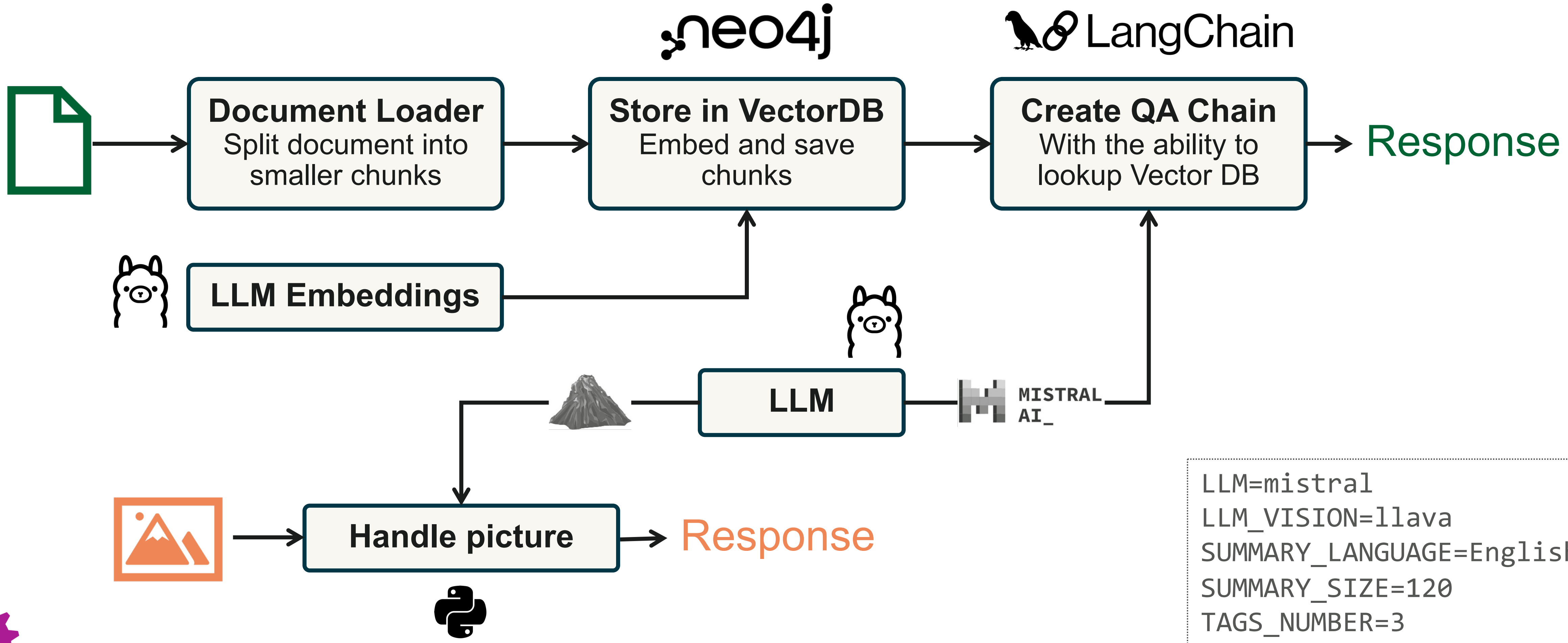
Document Centric Features



GenAI App

GENAI STACK

Python application that provides REST API endpoints




```
LLM=mistral
LLM_VISION=llava
SUMMARY_LANGUAGE=English
SUMMARY_SIZE=120
TAGS_NUMBER=3
```

GenAI REST API

```
curl --location 'http://localhost:8506/summary' --form 'file=@"./file.pdf"'
{
    "summary": " The text discusses...",
    "tags": " Golang, Merkle, Difficulty",
    "model": "mistral"
}
```

```
curl --location \
'http://localhost:8506/classify?termList="Japanese,Spanish,Vietnamese"' \
--form 'file=./file.pdf"'
{
    "term": " English",
    "model": "mistral"
}
```

GenAI REST API

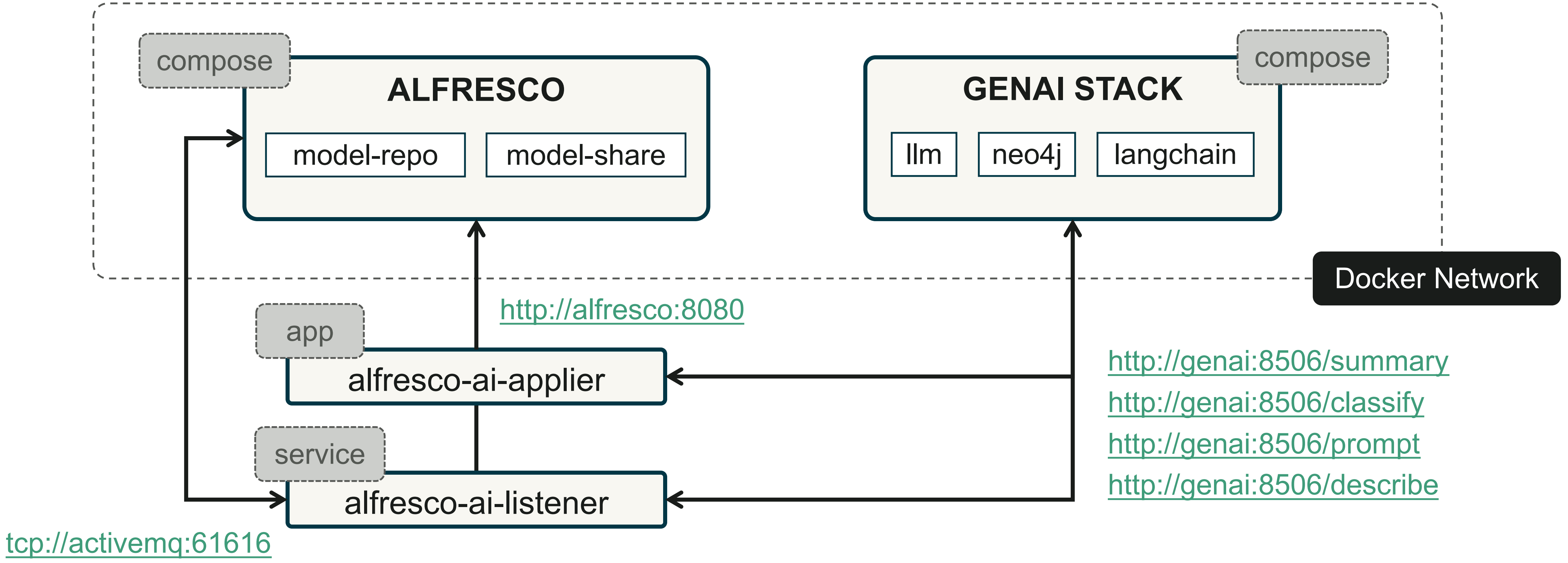
```
curl --location \
'http://localhost:8506/prompt?prompt="What is the name of the son?"' \
--form 'file=./file.pdf"'
{
    "answer": "The name of the son is Musuko.",
    "model": "mistral"
}
```

```
curl --location 'http://localhost:8506/describe' \
--form 'image=@"image.jpg"'
{
    "description": "The image features a man standing confidently. He is wearing
                    glasses, a beanie hat, and a jacket.",
    "model": "llava"
}
```

Integration

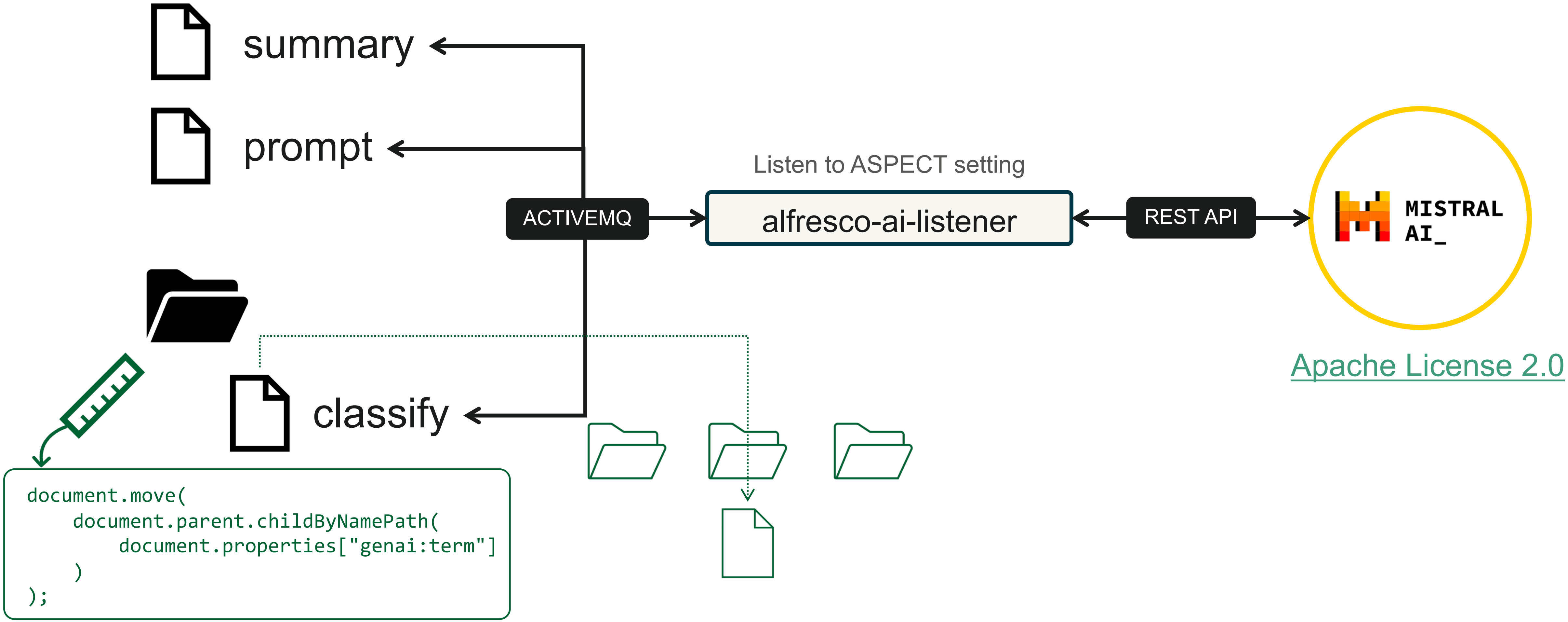
Integration with Content Service Platform

Deployment



https://github.com/aborroy/alfresco-genai

Content Service Platform Integration




```
document.move(
    document.parent.childByNamePath(
        document.properties["genai:term"]
    )
);
```

What Else?

Find your way

[Docker AI/ML Hack-a-thon]

- Readme AI
- Techdocs
- Docker Image Analyzer
- Docker Log Sentiment Analyzer
- GitChats AI

ollama alternatives

- https://gpt4all.io/index.html
- https://localai.io
- https://www.secondstate.io/run-llm
- https://huggingface.co/docs/hub/spaces-sdks-docker-first-demo

Hyland™

Thanks!