Using Generative AI and Content Service Platforms together

Angel Borroy
Developer Evangelist
Agenda

- GenAI Stack
- Features
- Integration
  - Existing Content
  - New Content
- What Else?
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
Ethical AI Rating

Mistral 7B

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

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?
Features

MacBook Pro 2021
Apple M1 Pro
32 GiB RAM
GPU
Document Centric Features

- Classify
- Data Extraction
- Text Recognition
- Content Analysis
- Translate
- Summarize
- Describe
- Prompt
- Redact
- Automation

GenAI

AI

NLP
GenAI App

Python application that provides REST API endpoints

Document Loader ➔ Split document into smaller chunks

Store in VectorDB ➔ Embed and save chunks

LLM Embeddings ➔ Create QA Chain

With the ability to lookup Vector DB

LLM ➔ Response

Handle picture ➔ Response

LLM=mistral
LLM_VISION=llava
SUMMARY_LANGUAGE=English
SUMMARY_SIZE=120
TAGS_NUMBER=3
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"}
curl --location \\n'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

ALFRESCO
- model-repo
- model-share

GENAI STACK
- llm
- neo4j
- langchain

Docker Network

http://alfresco:8080
http://genai:8506/summary
http://genai:8506/classify
http://genai:8506/prompt
http://genai:8506/describe

tcp://activemq:61616

https://github.com/aborroy/alfresco-genai
Content Service Platform Integration

summary

prompt

classify

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-sdk-docker-first-demo
Thanks!