CONTRIBUTIONS.md 2025-06-06

## Contributions

CONTRIBUTIONS.md 2025-06-06

data\_access\_call\_sequence\_diagram.wsd |---- llm\_retrieval.wsd |---- services\_diagram.wsd |---- tool\_calling.wsd |---- use\_case.wsd

## **Notebooks**

Additionally, the project/developer chat's retrieval mechanism requires some amount of parsing/cleaning/loading (all project code and documents), which is done offline using scripts and document\_loading.ipynb inside Eprice/Notebooks/. These are at least in some parts based on Paavo's (heh, me in 3rd person...) other previous and contemporary projects/works. Not everything is included in this repo, but will be later made available in https://github.com/PaavoReinikka.

To aid in the documentation of the project, there is also a UML diagram generator dashboard in Eprice/Notebooks/ (using PlantUML syntax). It was written with Python (see Notebooks README for details and instructions).

## Developer chat

The developer chat is based on OpenAI's gtp-4o-mini, which is basically free for research (and such) purposes. There are also a lot of open-source generative text models (chat and completion) in HF, and also good datasets for fine-tuning. Some of the experiments, with various open models, during this project will eventually be published in https://github.com/PaavoReinikka.

The embedding model used here, is open-source from HuggingFace. There are many good models available at HF, and a there is a leader board for embedding models, which is a good place to start looking (https://huggingface.co/spaces/mteb/leaderboard).

## Documentation

Most of the documentation was written by Paavo Reinikka, with some help from Juho Ahopelto, and with a lot of help from Copilot.