# TAGGING THROUGH LLM

## 1. Folder structure

```
tagging/
├── final_deliverable/
│   ├── Final deliverables.zip
│   ├── Tagging_1.ipynb
│   ├── Tagging_2.ipynb
│
├── Document_Tags_2.xlsx
├── Tagging_confidence_310525.ipynb
├── tagging.ipynb
├── tagging_oop_230425.ipynb
├── tagging_sub2_medidedalia_170425 (1).ipynb
├── tagging_sub_medidedalia_170425.ipynb
├── tagging_sublabels_demo.ipynb
```

## 2. Documents included

Relevant documents are all in the final_deliverable folder:

- **Final deliverables.zip** : packaged folder with all files for easy download, including the files used for tagging
- **Tagging_1.ipynb** : Jupyter notebook containing the two tagging functions but with a wider range of secondary tags
- **Tagging_2.ipynb** : Jupyter notebook containing the two tagging functions but with the 10 most important secondary tags

All documents below are iterations that might not be relevant anymore but that are stored there as it was our workspace:

- Document_Tags_2.xlsx
- tagging.ipynb
- tagging_oop_230425.ipynb
- Tagging_confidence_310525.ipynb
- tagging_sub_medidedalia_170425.ipynb
- tagging_sub2_medidedalia_170425 (1).ipynb
- tagging_sublabels_demo.ipynb

# 3. How to run Tagging_1.ipynb

**Step 1: Install requirements**

pip install requests urllib3 pydantic

**Step 2: fix the filepaths**

The filepaths included in the document are not relative, but local, so most will have the following format:

- C:\Users\Sophie\Final deliverables\documents_simulated\HISTORIAL_MEDIC.pdf
- folder_path = r"c:\Users\Sophie\Final deliverables\documents_simulated"

If you download the zip files, the folders structure will be the same within 'Final deliverables', so the part to change will be the beginning: c:\Users\Sophie\

This must be adapted depending on who and where the 'Final deliverables' folder is saved.

Using os.path.join can be a solution. For example, for the folders:

- base_path = os.getcwd()
- folder_path = os.path.join(base_path, "Final deliverables", "documents_simulated")

! But this means you must add 'import os' when importing the libraries in the first cell

**Step 3: run all cells**

# 4. How to run Tagging_2.ipynb

**Step 1: Install requirements**

pip install requests urllib3 pydantic pandas

**Step 2: fix the filepaths**

The filepaths included in the document are not relative, but local, so most will have the following format:

- C:\Users\Sophie\Final deliverables\documents_simulated\HISTORIAL_MEDIC.pdf
- folder_path = r"c:\Users\Sophie\Final deliverables\documents_simulated"

If you download the zip files, the folders structure will be the same within 'Final deliverables', so the part to change will be the beginning: c:\Users\Sophie\

This must be adapted depending on who and where the ‘Final deliverables’ folder is saved.

Using os.path.join can be a solution. For example, for the folders:

- base_path = os.getcwd()
- folder_path = os.path.join(base_path, "Final deliverables", "documents_simulated")

! But this means you must add ‘import os’ when importing the libraries in the first cell

**Step 3: run all cells**