MINI PROJECT 3 (OLA): NLP, NLU AND GENAI 2024

A great part of the information about the world comes to us as text in a language we can read, write, and understand - natural language. To be able to process, analyse and generate text automatically by means of a computer program, we need to apply certain preparation and transformation of it.

Natural Language Processing (NLP), Natural Language Understanding (NLU), and Generative AI (GenAI) are the among the most advanced and most popular branches of AI nowadays, which accumulate the numerous and years long achievements of researchers' and developers' in creating software that can communicate like a human.

NLP, NLU, and GenAI consists of multiple methods, algorithms, and tools in support of building modern AI applications.

The **objectives** of this project are:

- understanding the fundamental AI concepts of working with text
- gaining experience in implementation of relevant methods, algorithms, and libraries for creating usable AI applications

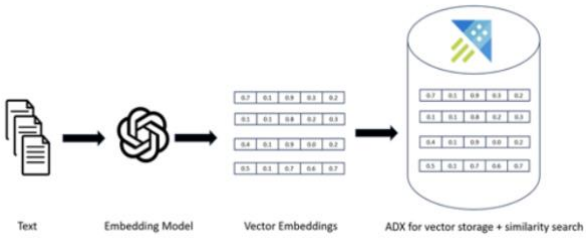
Your **task** is to create an interactive application that can answer questions, related to a specified domain area, in natural language.

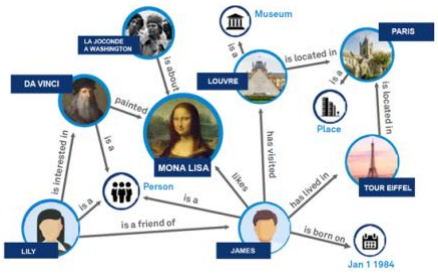
The solution includes several stages/subtasks, such as:

1. Collect and load text documents from various sources and formats in the context of the chosen domain – e.g. txt, doc, csv, json, pdf files, web pages, or APIs in data frames (choose a few).

2. Extract, clean, and transform into vectors the text from the original sources, to prepare it for AI processing, analysis, and generation.

3. Store the transformed documents in an appropriate software structure, such as vector database and/or knowledge graph.

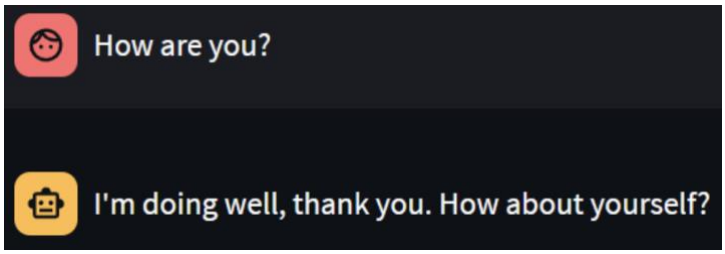




4. Apply the stored content for augmenting a selected pretrained Large Language Model (LLM) in order to improve its ability to understand and participate in communication within your domain of choice.

5. Create a simple interactive application prototype, which can input text documents and a question from a human user, and then output an answer of that question, considering the input documents and applying **GenAI** approach.

6. Suggest some business implementations of such an application.



Notes

- You will be able to learn how to solve the task in the coming AI sessions.
- You may work on the solution during the sessions.
- You can re-apply the solution or components of it in the AI exam project, as well as in the DB exam project as appropriate.

This OLA project brings an amount of 30 study points to your quantified pre-requisite of exam preparation.

Enjoy it!

the instructor