Master Thesis:
Natural Language Processing for Automated Data Analysis

Description:

This master thesis aims to leverage natural language processing (NLP) technologies, specifically large language models like GPT-4 or Med-Gemini, to develop a neural network that can interpret and execute data analysis tasks directly from natural language prompts provided by clinicians. Traditionally, clinicians must use tedious workflows for database querying and to manually extract and analyze data using software such as R or SPSS. The project seeks to simplify this process, enabling clinicians to input their analytical needs in natural language. The trained model will interpret these prompts to perform the necessary database operations, retrieve data, conduct analyses, and present findings in an interpretable and clinically relevant format. This system will transform the current complex data interaction into a streamlined, efficient, and user-friendly process, enhancing data-driven decision-making in healthcare.

Relevant knowledge/experience:

- **Computer Science and Mathematics:** Solid foundation in algorithms, data structures, and mathematical principles for machine learning and NLP.
- **Natural Language Processing:** Experience with NLP and frameworks, particularly in training large language models.
- **Machine Learning Frameworks:** Familiarity with libraries such as TensorFlow or PyTorch for neural network development.
- **Data Analysis:** Solid foundation in statistical analysis and visualization, with experience in tools like SPSS or R
- **Clinical Knowledge:** Basic understanding of clinical processes and medical terminologies to ensure the relevance and accuracy of data analysis.

Start:

August 2024

How to apply:

Interested candidates are requested to send their CV, certificates and a short description of themselves to **Assoc. Prof. Francesco Moscato** (francesco.moscato@meduniwien.ac.at) and **DI Laurenz Berger** (laurenz.berger@meduniwien.ac.at)