Abstract - Automatic Relation Extraction from Clinical Documents: A Study of Fine-Tuned Transformer Models and LLMs

Clinical documents are rich sources of patient information, notably about laboratory tests that inform medical decisions. However, as the volume of such documents grows, there's a pressing need for effective methods to interpret them. This thesis, titled "Automatic Relation Extraction from Clinical Documents: A Study of Fine-Tuned Transformer Models and LLMs", dives into this challenge of pinpointing test results and measurements in clinical documents and associating them with the respective laboratory tests they originate from. We've evaluated several models, such as Multilingual BERT, XLM-RoBERTa, and BioBERT, adapting them for our task. We also explored the potential of advanced large language models like GPT-3.5 and GPT-4 without any fine-tuning. An added dimension to our study is the multilingual nature of the clinical records, spanning Italian, Spanish, and Basque. These languages are often sidelined in research, which mostly centers on English. By focusing on them, we hope to fill a notable research gap. The thesis offers a journey starting with a review of relevant literature, a deep dive into the data and its nuances, a detailed look into our methodology, a discussion on our findings, and ends with insights for future investigations in this sphere.