General context of the PhD position

The recent arrival of Large Language Models (LLMs) and their associated tools for the general public reveals major challenges for society. Among the many fields that are, or will be, impacted by these generative models, the biomedical field is one of those that currently attract the attention of industrialists, researchers, but also the general public. Indeed, the need for tools and potential applications seems immense, whether at the level of the processing of textual documents, medical imaging, or even voice interaction. Due to the sensitive nature of the personal data handled and the fears of society associated with decision support tools, work in natural language processing (NLP) must innovate by addressing the issues inherent in this field. The PhD position is funded by the ANR project MALADES, in which we present innovative approaches for the integration of LLM in health centers. The aim of the MALADES project is to equip these centers with NLP tools derived from LLMs and adapted for the biomedical field while maintaining sovereignty of the models and complete control of their health data.

Objectives of the PhD position

Biomedical oriented textual instruction generation for LLMs training: One of the ways to train a chat-oriented LLM is to use a set of instructions, in the form of questions/answers, through a reward model based on reinforcement learning. We will explore two strategies to obtain these instructions, while having a restricted stroke in terms of human annotation: 1) Use of medical examination records, where we can obtain the associated questions and answers (both from the teacher and the students), 2) Self-generated instructions with approaches to artificially generate data for model training. The first part, which involves using exam data, will require manual annotation of students' answers to some extent, in order to set up the reward model based on the quality of the answers. The second, which involves generating artificial instructions, will also rely on annotated data to generate more robust instructions.

Study of LLM biases regarding social determinants: While biomedical LLMs will inevitably be used more and more by society, whether at the level of healthcare professionals or the general public, it is clear that the LMs contain inherent biases in the data (bias on gender, ethnicity, etc.) used for their training, having an impact on the results provided by these models. We seek to complete the evaluation by studying the potential biases contained in LLMs, focusing here on the social determinants of patient records. Our objective is not to conduct a comprehensive study of the biases present in the models, but rather to establish the necessary milestones for controlling language models, especially with regard to patient records that include social determinants (e.g. living, marital status or education).

Construction and annotation of original medical tasks: Current models mainly focus on classical NLP tasks in the medical field. However, these tasks are poorly designed to evaluate generative models. In order to fill this gap, we propose three original use cases: 1) Exams of health students on various medical fields with open questions and answers, 2) The evaluation and the potential production of pre-discharge reports of hospitalized patients from their records, and 3) Simulated real doctor / virtual patient scenario.

Cross-evaluation between data warehouses: Recent evaluations of language models in medical areas have shown disparities in the results obtained according to the data on which these models are trained and evaluated. This difference on the origin of the data is rarely evaluated, only the chosen tasks ultimately account for the intrinsic performance of the models. We want to conduct a cross-evaluation of the biomedical models across healthcare centers to assess the model suitability for deployment to other healthcare providers. This will enable us to determine whether or not the models (in whole or in part) need to be retrained at each healthcare center. This more prospective aspect will depend on collaborations with other data warehouses of different sizes which have expressed their willingness to participate in the evaluation process.