# Internship: LLM adaptation to the biomedical domain

Advisors: Richard Dufour, Benoit Favre

November 2023

## 1 Description

**Context** The recent development of Large Language Models (LLMs) has opened a range of opportunities for natural language processing (NLP) in the biomedical field: processing of electronic medical records, generation of reports, support for research... Due to the sensitive nature of the personal data handled and the risks associated with decision support tools, work in NLP will have to address the issues inherent to biases, applicability, and quality control. The goal of the ANR MALADES project is to propose innovative approaches for integrating LLMs in health centers, providing LLMs adapted to the medical field while ensuring the sovereignty of models and control over patient-related data. The project will contribute to four research areas: 1) the study of biomedical LLM legal and ethical aspects in France, 2) the integration of spoken interactions, 3) original case studies for evaluation of generative language models, and 4) the integration of dynamic and sovereign LLMs for the biomedical field, deployed on constrained material resources.

This internship proposal focuses on building local LLMs with instruction-prompting capabilities for the target domain. Following the internship, there is the potential for continuing with a Ph.D. thesis as part of the ongoing project.

**Problem statement** Thanks to their generalization properties and their versatility, LLMs are the go-to method for processing text. After proper fine-tuning on task descriptions, inputs, and expected output triplets (called instructions), they can handle a variety of tasks beyond the examples on which they have been trained [4]. However, the size of LLMs is still a strong predictor of their generalization properties [5], and the hardware infrastructure required for training, fine-tuning, and running them remains costly.

In the context of health centers, relying on third parties for handling patient data is strictly regulated, disallowing their distribution to LLM providers and advocating for running LLMs locally in a privacy-first setting.

Although the size of LLMs is crucial to good generalization, recent efforts have shown that models with a reasonable size can generate accurate results if trained through properly filtered data [2], and that parameter-efficient adaptation methods such as LoRA allow fine-grained control of model capabilities [1]. In order to maximize the potential of local LLMs in the biomedical domain, one would have to find a good compromise between general capabilities, and specific capabilities of the models.

**Objective and steps** The goal of this internship is to propose low-cost methods for adapting general-purpose LLMs that are able to sacrifice general knowledge while acquiring medical knowledge. To this end, the method will have to locate general knowledge in an LLM and overwrite it with domain-specific information, with methods similar to [3]. In addition, it will have to do so under the constraint that patient-related information used in training shall not be recoverable. The following steps will allow us to address the problem:

1. Collect and generate medical instruction triplets for LLM specialization.
2. Propose methods for adapting a relatively small LLM with biomedical knowledge.
3. Evaluate the method with respect to regular fine-tuning and LORA adaptation.

## 2 Profile

The intern will survey existing work, and leverage it to propose, implement and analyze LLM adaptation methods. The work will be implemented using Pytorch and relevant libraries. The candidate should have the following qualities:

- Excellent knowledge of deep learning methods (transformers...)

- Extensive experience with implementing Pytorch models, and handling research code bases
- Great scientific writing skills
- A hunch for the challenges of doing exciting research

The 6-month internship will take place at LIS/CNRS in Marseille during spring 2024. GPUs from the Jean-Zay super-computer will be available for training larger models.

# 3 Contact

Please send a CV, transcripts, and letter of application to benoit.favre@lis-lab.fr and richard.dufour@univ-nantes.fr. Do not hesitate to contact us if you have any questions.

# References


[1] Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models, 2021.

[2] Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. Mistral 7b, 2023.

[3] Kevin Meng, David Bau, Alex Andonian, and Yonatan Belinkov. Locating and editing factual associations in gpt. *Advances in Neural Information Processing Systems*, 35:17359–17372, 2022.

[4] Rafael Rafailov, Archit Sharma, Eric Mitchell, Stefano Ermon, Christopher D. Manning, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model, 2023.

[5] Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, Ed H. Chi, Tatsunori Hashimoto, Oriol Vinyals, Percy Liang, Jeff Dean, and William Fedus. Emergent abilities of large language models, 2022.