
Application of Large Language Models to enhance animal ethics evaluation in research

Emeline Perthame^{*1}, Marion Bérard¹, and Agash Uthayakumar^{1,2}

¹Institut Pasteur – Institut Pasteur de Paris – France

²Université Paris-Saclay – Université Paris-Saclay,Sorbonne Universités – France

Résumé

I hope my presentation may interest members of the Statomique group even if this work is slightly different from the usual topics addressed in the group. Indeed, it is not about analysis of omics-data, but rather about text data analysis. Nevertheless, I think it constitutes an interesting application of large language models (LLMs), as AI is a new tool for statisticians.

In Europe, the framework related to the projects involving animals is strictly regulative (European Directive 2010/63/EU) to protect the animals involved while allowing for progress in biomedical research. In France, research projects involving animals must be authorized by the MESR (Ministère de l'Enseignement Supérieur et de la Recherche). To obtain this license, the researchers have first to submit their project to the ethics committee for animal experiments (CEEA) for an ethics evaluation. The design and ethics evaluation of projects involving animals follow the 3R principles (Replace, Reduce, Refine) and a harm/benefit analysis. In the form they submit, the researchers must explain in particular the aim of their project, why it cannot be achieved without using animals, and how they will reduce the number of animals used and the impact on their welfare. They must describe which animals will be used and in particular they need to justify the age and sex they will chose.

One of the missions of the ethics committee from Institut Pasteur (CETEA) is to improve the culture of care on the campus, which have a great impact on both animal welfare and the quality of the data generated by these experiments. Multiple actions have been deployed to answer this objective, and one of them is to learn from the analysis of the hundreds of forms handled by the ethics committee (since 2018, approx. 350 authorization requests).

This database is a gold mine of information but is hard to analyze by statistical methods because of the nature of the data, which is text data written in an almost unformatted text document. Usual statistics tools based on lemmatization and factorial analysis fail because of the huge heterogeneity of ways of describing and justifying the progress of the projects. For example, even the choice of the sex of the animals is noted, not in a formatted checkbox, but in a paragraph where researchers indicate the sex they will use and why. This is where LLMs appear to be an adapted tool, as these models are designed to analyze natural language data if appropriately mastered with a prompt.

In this presentation, I will present the context of the analysis and how we used LLMs as reading assistants to identify and cluster the choice of using animals of one specific sex or both sexes and summarize the main arguments associated with this choice. We have the

^{*}Intervenant

chance at Institut Pasteur to have LLM models hosted on the campus, which preserves the confidentiality of the data. These models are accessible via a chat box but also via an API. I will therefore show how interactions with an LLM can be made reproducible by a script through this API, the importance of prompt engineering to get relevant results and avoid overloading the model as the tokenization is limited, how the LLM can be relevantly (or not !!) creative to detect new elements in the text, and how we handled its creativity and stochasticity to obtain more robust results. I will also show some of the results to answer the problem. In the end, I will explain how this kind of analysis can help the ethics committee to standardize the submission form and contribute to improve practices in animal experiments. This work is a joint work with Marion Bérard (Institut Pasteur) and Agash Uthayakumar (M1 intern, Université Paris-Saclay, Institut Pasteur).

Mots-Clés: LLM, animal ethics, IA application