How Good are Multilingual Small Language Models at Explaining Medical Knowledge?

Ioana Buhnila*[1]

[1]ATILF UMR 7118 (CNRS - Université de Lorraine) – Université de Lorraine, Centre National de la Recherche Scientifique – France

Résumé

Large Language Models (LLMs) became easily accessible to the general public through chatbots like ChatGPT (OpenAI, 2022). However, full access remains costly due to the very big size of LLMs: dozens or even one hundred billion parameters. Recent research shows that Small Language Models (SLM) (Schick and Schütze, 2021) can be as effective on specific tasks and faster on small devices such as mobile phones (Abdin et al., 2024; Lepagnol et al., 2024). The general public use language models for many tasks, including for health related questions (Singhal et al., 2023). Nevertheless, medical terms, which are specialized lexical units (Condamines and Rebeyrolle, 1997) can be difficult to understand for laypeople (LeBlanc et al., 2014; Tavakoly Sany et al., 2020). Thus, medical concepts have to be simplified through paraphrases or definitions, such as "arthrosis of the thumb" for *rhizarthrosis*. In this sense, we tested an open-access multilingual medical SLM, BioMistral (Labrak et al., 2024), to assess the quality of its answers in a downstream task: people asking medical-related questions to a language model. We evaluated the zero-shot performance of BioMistral in two languages: French and Romanian, a low resource-language. We used a quantized version of the model, BioMistral-7B-SLERP-GPTQ, that has *only* 1,38 billion parameters, compared to the original version of 7 billion parameters. BioMistral's pre-training corpus covers nine languages, including French and Romanian, but English represents 98,75% of the data.

Our experiences were conducted on 100 pairs of medical terms and their corresponding paraphrases in French and Romanian extracted from the RefoMed dataset (Buhnila, 2023). We asked (*prompted*) BioMistral to generate a concise answer containing a paraphrase or short definition for the given medical term. We evaluated 400 generated answers through a quantitative and qualitative analysis to assess: *correctness* (semantic equivalence), *completeness* (full syntactic form), and *readability* (easy to understand for laypeople) (Buhnila et al., 2024).

French results show that BioMistral's answers in French are 94% correct using an English prompt, while in Romanian only 70% are correct, as the language might be rare in the corpus of the language model. However, when using a Romanian prompt, the model's performance rises to 94%.

> Asthme: maladie où les *airways* se ferment et se contractent. Tumora malignă *(malignant tumor) is a type of tumor that has the potential to grow and spread to other parts of the body.*

*Intervenant

We noticed that BioMistral's answers in both languages can contain English words such as *airways* (first example). In the second example, an English definition follows the Romanian medical term "tumora malignă" *(malignant tumor)*. While the answers generated are grammatically correct in French, in Romanian we observed a higher percentage of invented words or bad grammar, such as "batemente", correct form "bătăi" (*beats*) or "mișcute" from "mișcate" (*moved*).