ADAPTING LARGE LANGUAGE MODELS FOR BULGARIAN CLINICAL DIETETICS: A MULTI-STAGE TRAINING APPROACH

Simeon Monov, Detelinka Trifonova, Nikolay Pavlov

Abstract. *The application of Large Language Models (LLMs) in specialized domains like medicine remains limited, particularly for low-resource languages such as Bulgarian [3]. Currently, there is a notable lack of specialized models and datasets for clinical dietetics within this linguistic context, despite the growing potential for LLMs to assist in nutritional risk assessment and dietary planning [2]. In this work, we address this gap by developing several specialized Bulgarian datasets, including a corpus of educational materials, a question-answering (Q&A) dataset, and structured clinical cases developed in collaboration with a certified clinical dietitian. Following established protocols for medical domain adaptation [1], we utilized these data to train a Large Language Model using a three-stage methodology: Domain-Adaptive Pre-Training (DAPT), followed by Supervised Fine-Tuning (SFT) for Q&A tasks and diet generation. Finally, we conducted an expert-led evaluation of the model's performance as an assistant in dietetic practice, focusing on its ability to generate personalized recommendations aligned with established clinical guidelines.*

Key words: Large Language Models, LLM, Clinical Dietetics, Low-Resource Languages, DAPT, SFT, NLP, Medical AI

Acknowledgments

This study is supported by the project SP25-FMI-008 "Research and implementation of modern language models and artificial neural networks for automated processing, forecasting, and structuring of data and texts in specific application domains" at the Paisii Hilendarski University of Plovdiv.

References

[1] J. Chen, X. Wang, K. Ji, et al., HuatuoGPT-II, one-stage training for medical adaption of LLMs, *arXiv*, 2023, ISSN: 2331-8422, `https://doi.org/10.48550/arxiv.2311.09774`

[2] A. Belkhouribchia, C. Pen, Large language models in clinical nutrition: An overview of its applications, capabilities, limitations, and potential future

prospects, *Frontiers in Nutrition*, 2025, vol. 12, ISSN: 2296-861X, https://doi.org/10.3389/fnut.2025.1635682

[3] P. Niszczota, I. Rybicka, M. Kaczmarek, Large language models for real-world nutrition assessment: Structured prompts, multi-model validation and expert oversight, *Nutrients*, 2026, vol. 18, no. 1, pp. 23–41, ISSN: 2072-6643, https://doi.org/10.3390/nu18010023

Simeon Monov[1,*], Detelinka Trifonova[1], Nikolay Pavlov[1]
[1] Paisii Hilendarski University of Plovdiv,
Faculty of Mathematics and Informatics,
236 Bulgaria Blvd., 4027 Plovdiv, Bulgaria
Corresponding author: smonov@uni-plovdiv.bg