Large Language Models in Structure Optimizing of Machine and Deep Learning Models

Supervisor: Prof. Ing. Komínková Oplatková Zuzana, Ph.D.

Consultant: Ing. Turečková Alžběta, Ph.D., ---

Department: Department of Informatics and Artificial Intelligence

Programme: Information Technologies

Abstract:

This research aims to advance the use of Large Language Models (LLMs) in structurally optimizing machine learning and deep learning models. Central to this approach is the integration of LLM-driven insights into the model design process, with a specific focus on exploring LLMs' ability to identify and propose innovative architectural patterns. This exploration aims to challenge and extend the limits of conventional model design, including the development of novel structures and functional blocks, as well as optimizing hyperparameters, tailored to specific dataset characteristics and performance goals. The expected outcome is a more efficient methodology, characterized by significant enhancements in model efficiency, accuracy, and adaptability across various domains, while also reducing energy consumption during training. The research will propose and validate this methodology through comprehensive experiments. The anticipated results promise to enrich the scientific community with new approaches and methodologies, offering valuable insights and practical prompts for LLM applications in this field.

Literature:

[1]   KHAN, Wahab, et al. Exploring the frontiers of deep learning and natural language processing: A comprehensive overview of key challenges and emerging trends. Natural Language Processing Journal, 2023, 100026.

[2]   BU, Kun; LIU, Yuanchao; JU, Xiaolong. Efficient Utilization of Pre-trained Models: A Review of Sentiment Analysis via Prompt Learning. Knowledge-Based Systems, 2023, 111148.

[3]   ESPEJEL, Jessica López, et al. GPT-3.5, GPT-4, or BARD? Evaluating LLMs reasoning ability in zero-shot setting and performance boosting through prompts. Natural Language Processing Journal, 2023, 5: 100032.

[4]   YANG, Zhen, et al. Improving domain-specific neural code generation with few-shot meta-learning. Information and Software Technology, 2023, 107365.

[5]   HE, Xin; ZHAO, Kaiyong; CHU, Xiaowen. AutoML: A survey of the state-of-the-art. Knowledge-Based Systems, 2021, 212: 106622.

[6]   NIKITIN, Nikolay O., et al. Automated evolutionary approach for the design of composite machine learning pipelines. Future Generation Computer Systems, 2022, 127: 109-125.