Zgłoszenie tematu pracy dyplomowej **:: STUDIA II STOPNIA ::**

rok akademicki 2025/26

Promotor:	**dr hab. Jozef Kapusta, prof. UKEN**
Temat pracy magisterskiej (j. polski, j.angielski):	Analysis of the effectiveness of various language models in the classification tasks *Analiza skuteczności różnych modeli językowych w zadaniach klasyfikacyjnych*
Zakres pracy i oczekiwane rezultaty praktyczne:	In recent years, we have witnessed a revolution in human–computer communication driven by progress in large language models (LLMs). LLMs are trained for a wide range of tasks such as text generation, question answering (QA), machine translation, text summarization, code completion, grammar correction, and others. In practice, these models are now primarily accessed through the Hugging Face platform, which provides a broad ecosystem of open models, datasets, and tools for training, fine-tuning, and deployment. The aim of this master's thesis is to examine language models available on the Hugging Face platform, with a particular focus on models suitable for the Polish language and related languages. The student will analyze these models mainly in terms of their capabilities with respect to the tasks for which they were trained or fine-tuned (e.g., text generation, QA, summarization, classification, translation). The thesis will produce a structured database/overview of selected Hugging Face models, complemented by usage examples and source-code snippets. The analysis and comparison of the models will be carried out on large and thematically diverse Polish datasets, such as WiktorS/polish-news and the COVID-19 Polish Media Dataset, which include news content from domains like politics, sports, and economics. The thesis will provide a standardized experimental evaluation of the models using appropriate performance metrics. Special emphasis will be placed on comparing the efficiency of classical and modern architectures in the context of the Polish language and on assessing their applicability in practical scenarios, such as automatic categorization of media content. The thesis will also include a review of scientific literature on text classification and deep-learning applications in NLP, with a focus on Polish-language research and resources. Overall, the work will concentrate on LLM and NLP models for Polish, highlighting their real-world usability and providing a systematic comparison of available solutions within the Hugging Face ecosystem.
Aspekt naukowy, problemowy, innowacyjny pracy:	Implementation neural networks methods, implementation machine learning methods.

*Oprogramowanie, język programowania, środowisko systemowe:	Python 3.10–3.12, biblioteki: transformers, datasets, accelerate, peft, scikit-learn, pandas, numpy, matplotlib/seaborn; frameworki: PyTorch (2.0+), Hugging Face Transformers, lokalne środowisko Linux (Ubuntu 24.04) z GPU NVIDIA RTX 5080
*Środowisko uruchomieniowe	
Dodatkowe wymagania i uwagi:	English language
*Literatura:	• Bator, M., & Kocon, J. (2022). Polish court ruling classification using deep neural networks. Sensors, 22(6), 2137. • Kuleshov, V., et al. (2024). Text classification: Neural networks vs machine learning models vs pre-trained models. arXiv:2412.21022. • Rybak, P., et al. (2024). Towards explainable fake news detection: Polish media use-case. Neurocomputing, 608, 128450.

*pola opcjonalne