Gorokhovatskyi O. Perplexity-based AI-generated text classification in Ukrainian using small language models / O. Gorokhovatskyi // Advanced Information Systems. – 2026. – Vol. 10 (№ 3). – Pp. 5-12.

Perplexity-based AI-generated text classification in Ukrainian using small language models

The aim of the research. The rapid advancement of generative artificial intelligence language models has introduced new complexities in discerning the authorship and quality of textual content. In this paper, we explored the feasibility of using perplexity – a measure of token predictability – as the only discriminative feature for classifying AI-generated versus human-written texts in Ukrainian within the IT domain. Our approach employed small language models to calculate perplexity and detect content generated by state-of-the-art models, evaluating the potential for lightweight solutions. Research results. Initial experiments using a single perplexity threshold across Gemma 3 / Llama 3.2 1B models yielded classification accuracies around 0.70. The full token-level probability sequences were proposed as feature vectors, enabling us to achieve an accuracy of 0.68 via simple KNN classification. Finally, the convolutional neural network architectures trained on these features allowed us to obtain 0.82–0.87 accuracy. Conclusions. The comparative analysis with a traditional NLP-based discriminative neural network model revealed that direct text piece classification outperforms perplexity-based methods, although the latter still demonstrate practical utility.

Keywords: perplexity, small language models, Gemma, Llama, AI, AIG, human-written text, CNN, text classification, accuracy, token-level probability vector, KNN