

Kevin Chovanec

Abstract: “LLM and SLM in Higher Ed Administration”

This dissertation explores the use of pre-trained language models (PLM) in the domain of higher education, focusing especially on two common tasks: Multi-label text classification (MLTC) of student survey data, and Automated Essay Scoring (AES) of admissions essays. For each task, we investigate open critical questions, sensitive to the higher ed context. With MLTC, we explore novel architectures using small language models (SLM), including both ensemble and model blending approaches and the fusion of tabular and text data, finding that additional tabular data does indeed improve classification accuracy. As data can be quite limited and generally cannot be shared across institutions in higher education, we also explore large language models (LLM) for synthetic data generation for MLTC, evaluating uses and contexts in which synthetic data is most effective. Finally, to encourage adoption within higher ed, a domain that prioritizes fairness and transparency, we benchmark common eXplainable Artificial Intelligence (XAI) methods when explaining the results of our MLTC task. For automated essay scoring, we benchmark various SLM on common admissions-essay-evaluation tasks, and we develop novel approaches to AES based on feature engineering. We also offer a detailed discourse and stylometric analysis of both student-written and AI generated essays and combine discriminative and generative approaches to automated essay scoring, suggesting that a blend of traditional machine learning and small, local generative models can produce highly accurate, task-independent essay evaluation pipelines.