

Dissertation Defense
Doctor of Philosophy in Computer Science

“Automated Assessment of Classroom Discussion Quality with Large Language Models” by Nhat Tran

Date: December 8, 2025

Time: 2 – 4 p.m.

Place: Room 6106, Sennott Square, 3810 Forbes Ave.
Pittsburgh, PA 15213

Committee:

- Diane Litman, Professor, Department of Computer Science, SCI
- Erin Walker, Associate Professor, Department of Computer Science, SCI
- Milos Hauskrecht, Professor, Department of Computer Science, SCI
- Richard Correnti, Associate Professor, Learning Research and Development Center, School of Education

Abstract:

Rigorous and interactive classroom discussions that engage students in high-level thinking and reasoning are central to effective learning and a cornerstone of many teaching interventions. However, assessing discussion quality at scale is costly and impractical for most researchers. Automated assessment of classroom discussion quality can enable large-scale data collection to study how discussions foster student reasoning and provide formative feedback to help teachers improve instructional practices.

Recent advances in Large Language Models (LLMs) have transformed Natural Language Processing (NLP), offering new opportunities for automating educational assessment. This dissertation investigates how to leverage LLMs to automatically evaluate multiple dimensions of classroom discussion quality through three key directions: **(i)** prompting strategies, **(ii)** improved in-context learning (ICL) example selection, and **(iii)** synthetic data generation for fine-tuning. First, the work systematically examines how three prompt-based factors (task formulation, context length, and the inclusion of few-shot examples) affect both the performance and computational efficiency of LLMs in assessing classroom discussion quality. The findings highlight that ICL examples are critical and that formulating the task as multiple binary classification problems yields the strongest performance.

Building on these insights, the dissertation introduces automated ICL example retrieval to further enhance performance. After identifying the limitations of standard retrievers, a new approach that combines re-ranking and label ratio regulation is proposed to enable dynamic and balanced example selection. Finally, to address the shortage of annotated discussion data, the dissertation explores synthetic data generation for fine-tuning LLMs. A novel data curation method is developed to identify high-quality synthetic examples by testing whether a generated instance, when used as an ICL example, helps the model correct a previous error.

Overall, this work presents a holistic approach to harnessing LLMs for automated assessment of classroom discussion quality, paving the way for scalable, data-driven insights into effective classroom discourse.