University of Pittsburgh | School of Computing and Information

Dissertation Defense

Doctor of Philosophy in computer science

"**On The Role of Argumentation in Long-form Abstractive Summarization**" by
Mohamed Elaraby

Date: August 24, 2026
Time: Noon to 2 p.m.
Place: 6106 Sennott Square, 210 S. Bouquet Street, Pittsburgh, PA 15213

Committee:

- Diane Litman, Professor, Computer Science, School of Computing and Information, University of Pittsburgh
- Kathleen McKeown, Professor, Computer Science, School of Computer Science, Columbia University
- Adriana Kovashka, Associate Professor, Computer Science, School of Computing and Information, University of Pittsburgh
- Xiang Lorraine Li, Assistant Professor, Computer Science, School of Computing and Information, University of Pittsburgh

Abstract:
Abstractive summarization aims to produce a concise, coherent, and informative representation of the information contained in a longer source. Recent progress in Large Language Models (LLMs) and large-scale training corpora has substantially improved the quality of generated summaries, particularly in terms of fluency and coherence. Despite this progress, summarization still faces practical challenges, especially in high-stakes long-context domains where fluent output can mask persistent failures in factuality and saliency coverage. This thesis focuses on such domains—law and science—where the salient information typically follows a structured argumentative scheme (issues, reasons, and conclusions in legal opinions), is sparsely distributed, and extends across very long contexts, posing an additional challenge to representing this information effectively in summaries. To address these challenges, we develop *argument-aware* techniques for both improving and evaluating abstractive summarization in high-stakes argumentative domains. We first show that explicitly integrating argument role information into the summarization process improves the coverage of salient argumentative content in long legal opinions, and that a second-stage reranking method leveraging argument structure yields further gains. Recognizing that standard automatic metrics inadequately capture argumentative quality, we introduce human evaluation guidelines centered on argument coverage, and subsequently propose ARC, a hierarchical, interpretable evaluation framework that decomposes argument roles into atomic facts to measure coverage while distinguishing omissions from factual errors. Using ARC, we uncover systematic limitations of instruction-following LLMs, including incomplete coverage of salient content (especially in smaller models), positional bias, and

role-specific coverage preferences. Finally, to tackle these limitations, we demonstrate that sequence-level distillation from a capable long-context teacher is a simple, annotation-free, and data-efficient strategy for improving argument saliency coverage in smaller models, and further explore preference-based optimization through argument-aware DPO distillation. Across these contributions, we span the progression from long-context pretrained language models to long-context instruction-tuned models and, most recently, long-context reasoning models, offering a unified argument-aware perspective on generating and evaluating summaries in high-stakes domains.