STEVENS
INSTITUTE OF TECHNOLOGY
1870

Ph.D. DISSERTATION DEFENSE

Candidate:	Yupeng Cao
Degree:	Doctor of Philosophy
School/Department:	Charles V. Schaefer, Jr. School of Engineering and Science / Department of Electrical and Computer Engineering
Date:	Tuesday, April 21st, 2026
Time:	2:30 p.m. to 4:00 p.m. (Eastern)
Location:	Burchard 714
Title:	Designing Reliable Large Language Model-Powered Agentic Workflows and Their Applications
Chairperson:	Dr. Koduvayur Subbalakshmi, Department of Electrical and Computer Engineering
Committee Members:	Dr. Hong Man, Department of Electrical and Computer Engineering Dr. Hao Wang, Department of Electrical and Computer Engineering Dr. Ping Wang, Department of Computer Science Dr. Zining Zhu, Department of Computer Science

ABSTRACT

Large language models (LLMs) have demonstrated remarkable capabilities in text comprehension, reasoning, and interactive tasks. However, the errors they produce, including fabricated content, unsubstantiated claims and hallucination, pose significant threats to information integrity in high-stakes domains such as scientific communication, knowledge verification, and financial decision-making. This dissertation addresses a central question: *How can we design and implement more reliable, verifiable, and explainable LLM-based agent workflows for deployment in high-risk applications?* To answer this, we propose a systematic methodology along with empirical validation across multiple application domains.

We begin with two empirical studies on scientific misinformation detection and hallucination detection, which together reveal the information security risks introduced by LLMs and establish a practical foundation for studying misinformation in real-world settings. Building on these findings, we investigate how internal signals from LLMs can be leveraged to enhance hallucination detection in challenging retrieval-augmented generation (RAG) scenarios. We then extend this line of work by constructing agentic workflows that exploit the tool-calling capabilities of LLMs, enabling models to interact with external environments for more challenge realistic veracity assessment task. A memory mechanism is further integrated into the agent framework to improve efficiency.

Finally, we extend these reliability-oriented agentic workflow principles to the high-stakes financial domain. For multimodal financial risk prediction, we design a hierarchical information extraction agent that generates actionable signals from reliable data source while preserving full reasoning trajectories and improving predictive performance. For sequential financial decision-making tasks, we develop a multi-agent framework that exhibits stronger risk-adjusted behavior compared to single-agent baselines. We further benchmark LLM-based agents on diverse financial tasks to reveal the boundaries of LLMs' capabilities.