

NLP SEMINAR

Toward A Foundation for Reprogramming Large Language Models



Tuesday, March 10, 2026, 1:30 pm – 2:30 pm
CSL Building, E 144 Colloquium Room

Chi Han
University of Illinois
Urbana-Champaign

Abstract: Large language models (LLMs) have achieved strong performance in domains covered by pre-training corpora. However, once pre-trained, their internal computations are largely treated as fixed black-boxes, and systematically modifying them remains an open challenge. As a result, when LLMs exhibit limitations in open-world settings, such as extreme-scale contexts, novel scenario generation, and scientific discovery, we lack principled methods to intervene at the internal mechanistic level where these failures originate. In this talk, I present my research toward a principled foundation for LLM reprogramming by uncovering governing principles of their internal computations and developing methods for post-hoc reprogramming of LLMs.

My approach integrates two complementary components: theoretical analyses of how LLMs process context and generate decisions, and empirical methods that reprogram internal architectures with minimal additional resources. First, I present LM-Infinite, which provides a mechanistic explanation of context-length failure and enables generalization to contexts exceeding 200 million tokens without retraining. Second, in LM-Steer, I show that word embeddings act as controllable steers for generation, enabling efficient, interpretable, and transferable control over LLM outputs. Third, extending beyond language, I introduce a modular chemical language model that incorporates domain-grounded representations to support synthesis-aware molecular reasoning and drug discovery. Together, my research sheds light on principled mechanisms for making LLMs more interpretable, adaptable, and reliable across scientific disciplines and real-world settings.

Bio: Chi Han is a final-year Ph.D. candidate in Computer Science at the University of Illinois Urbana-Champaign (UIUC), where he works in the NLP group under the supervision of Prof. Heng Ji. He received his undergraduate degree from Tsinghua University, China, through the Yao Class program. His research has led to first-author publications at top venues including NeurIPS, ICLR, TMLR, ACL, and NAACL, which received Outstanding Paper Awards at NAACL 2024 and ACL 2024, as well as the Best Demo Award (1st place) at the NSF Summit for AI Institutes Leadership (SAIL). His research has been supported by the IBM PhD Fellowship, Amazon AICE PhD Fellowship, the Mavis Future Faculty Fellowship, and Capital One seeding grant. His research interests focus on developing theoretical foundations and empirical methods for post-hoc reprogramming of large language models (LLMs). His work addresses intrinsic limitations of LLMs across multiple domains, including context length, decision-making transparency, generation control, and scientific discovery, and achieves strong performance across downstream tasks.

Faculty host: [Chuang Gan](#)