

Xuyuan Liu

✉ xuyuan.liu@dartmouth.edu · 🏠 xuyuano204.github.io · 🌐 Github ·

RESEARCH INTERESTS

My research focuses on **the science of Language Models**, especially on understanding their behavior during both training and inference through methods that probe their internal mechanisms. I aim to **understand how LLMs work internally** and leverage that understanding to develop technical methods that enhance their performance.

EDUCATION

- **Dartmouth College** Sept. 2023 - Present
Hanover, NH
PhD in Computer Science
 - Advisors: Prof. [Yujun Yan](#)
- **Nankai University** Sept. 2019- June 2023
Tianjin, China
Bachelor of Science, Computer Science
 - Thesis:*Few-shot Knowledge Retrieval Framework in multi-turn Dialogue System*

PUBLICATIONS & PREPRINTS

* EQUAL CONTRIBUTION; S=IN SUBMISSION, CP=CONFERENCE PROCEEDINGS

- [CP.5] Xuyuan Liu, Zhengzhang Chen, Xinshuai Dong ... Yujun Yan, Haifeng Chen **Representation Interventions Enable Lifelong Unstructured Knowledge Control**. In *Submission*
◆ LLM Personalization ◆ Efficient LLM
- [CP.4] Xuyuan Liu, Lei Hsiung, Yaoqing Yang, Yujun Yan. **Spectral Insights into Data-Oblivious Critical Layers in Large Language Models**. In *Findings of the Association for Computational Linguistics: ACL 2025/Also Presented in UniReps Workshop@NeurIPS*
◆ LLM Training Behavior ◆ Efficient LLM
- [CP.3] Xuyuan Liu, Yinghao Cai, Qihui Yang, and Yujun Yan. **Exploring Consistency in Graph Representations: from Graph Kernels to Graph Neural Networks**. In *Proceedings of the Thirty-Eighth Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2024.
◆ Representation Learning ◆ Graph Learning
- [CP.2] Xinshuai Dong, Zhengzhang Chen, **Xuyuan Liu**, Shaoan Xie, Kun Zhang, Haifeng Chen **The Power of Order: Fooling LLMs with Adversarial Table Permutation**. In *Submission*
◆ LLM Safety
- [CP.1] Zichen Liu, **Xuyuan Liu**, Yanlong Wen, Guoqing Zhao, Fen Xia, and Xiaojie Yuan. **TreeMAN: Tree- enhanced multimodal attention network for ICD coding**. In *Proceedings of the 29th International Conference on Computational Linguistics (COLING)*, 2022. Oral
◆ Multi-Modality Learning

INTERNSHIP

- **NEC Laboratories America** June. 2025 - Sep. 2025
Princeton, NJ
Research Scientist Intern
Advisor: Dr. Zhengzhang Chen and Dr. Haifeng Chen
Project: *LLM Knowledge Customization*
* Lifelong LLM behavior control through interventions on hidden states, enabling precise, localized, and lightweight knowledge management. [CP.5]

HONORS, AWARDS, AND GRANTS

- **Guarini Travel Grant (2024)**. Guarini School of Graduate and Advanced Studies, Dartmouth College
- **Dartmouth Fellowship (2023)**. Guarini School of Graduate and Advanced Studies, Dartmouth College
- **Academic Excellence Scholarship**. Nankai University (7/130)
- **Innovation Award of Technology and Research Scholarship**. Nankai University (3/130)

TEACHING EXPERIENCES

Deep Learning, Head TA (Spring2024); Machine Learning & Statistical Data Analysis, Head TA (Fall2024);

PROFESSIONAL SERVICES

- **Reviewers:**(Conf.) ICLR2025-2026, NeurIPS 2024-2025, ICML2025,

SKILLS

- **Programming Languages:** C, C++, Python, Java, SQL, Shell Scripting
- **ML/LLM Frameworks and Tools:** PyTorch, vLLM, HuggingFace, PyTorch FSDP, VeRL
- **Tools:** Git, L^AT_EX