Zheyuan (Jason) Zhang

✉ zzhang42@nd.edu   📞 +1 347 405 2093   🔗 Homepage   in Linkedin   ○ GitHub   🎓 Scholar

Research Directions

My research focuses on post-training strategies for large language models, with a focus on optimizing LLM agent behaviors, such as improving performance, reducing manual intervention, and enhancing alignment. Specifically, the topics cover the following three aspects:

- **Dynamic Multi-agent System Design:** Multi-agent prompt optimization [4]; Dynamic multi-agent collaboration: [5], [6]. Online multi-agent system [10].
- **Retrieval Augmented Generation:** Domain-specific GraphQA [1]; Retrieval-enhanced recommendation system [2]; Multi-agent KBQA [6].
- **Reinforcement Learning:** Inference time policy optimization [4]; Multi-step RL [9]; Uncertainty Quantification.

My research application fields include: LLM for coding & data science [4], Health intelligence [1-3], LLM4Science [7].

Professional Experience

Amazon, Inc.   *Seattle, WA*
Applied Scientist (Part-time)   *Aug. 2025 – Dec. 2025*
- Research on uncertainty-aware reinforcement learning and LLM agents for coding and data analytical tasks.

Applied Scientist (Intern)   *May 2025 – Aug. 2025*
- Research on LLM agents for code generation. Develop MAPRO, a self-evolving prompt optimization framework for multi-agent system. (Paper Under Internal Review)
- Both the coding agents and the MAPRO frameworks are deployed into the *AIDE (Assistant for Intelligent Deep divE)* project, which aims to build an expert system for deep-dive investigations.

IBM Research   *Yorktown Heights, NY*
Student Researcher (Part-time)   *Jun. 2024 – May. 2025*
- Research on interpretable and explainable foundation models using graph-based retrieval augmented generation on health-care domains, which leads to two publications to ACL 2025 and KDD 2025. Check this website 🗗 for more details of the collaboration project.

Coleridge Initiative, Inc.   *New York, NY*
Research Scientist   *Oct. 2021 – Mar. 2023*
- Lead the research project *Food for Thought*: using ML Methods to link food and nutrition databases on a security platform. Responsible for the toolkit and evaluation pipelines and communications between teams. Check this blog 🗗 for the competition. Lead the *Unemployment Insurance* project, developing a real-time system to monitor the trend of unemployment rate for IL, IN, TN states.
- Lead the Data Team of Coleridge, responsible for all incoming data ingestion and ETL pipelines.

Associated Research Scientist   *Aug. 2020 – Oct. 2021*
- Responsible for the *Rich Context* project. This research project aims to identify and extract dataset mentions from full-text social science publications.
- Host the Kaggle Competition *Show US the data*. Responsible for establishing a benchmark for the competition along with technical implementation. A paper is accepted to Harvard Data Science Review.

Junior Research Scientist (Intern)   *Oct. 2019 – Aug. 2020*
- Work on KG completion and text entity extraction tasks for establishing a large-scale KG for *Rich Context* Competition.

US Ignite, Inc.   *New York, NY*
Research Fellow (Part-time)   *Mar. 2020 – Aug. 2020*
- Research on the Smart Monitor for Accelerating Regional Transformation project. Cooperate with senior research fellows to develop a recommendation system to help small business owners find better business locations.

Selected Publications

* To see the full publication list, please visit my google scholar here ↗.

[1] NGQA: A Nutritional Graph Question Answering Benchmark for Personalized Health-aware Nutritional Reasoning [ACL Main'25]

○ **Zheyuan Zhang**, Yiyang Li, Nhi Ha Lan Le, Zehong Wang, Tianyi Ma, Vincent Galassi, Keerthiram Murugesan, Nuno Moniz, Werner Geyer, Nitesh V Chawla, Chuxu Zhang, Yanfang Ye

[2] MOPI-HFRS: A Multi-objective Personalized Health-aware Food Recommendation System with LLM-enhanced Interpretation [KDD'25]

○ **Zheyuan Zhang**, Zehong Wang, Tianyi Ma, Varun Sameer Taneja, Sofia Nelson, Nhi Ha Lan Le, Keerthiram Murugesan, Mingxuan Ju, Nitesh V Chawla, Chuxu Zhang, Yanfang Ye

[3] Diet-ODIN: A Novel Framework for Opioid Misuse Detection with Interpretable Dietary Patterns [KDD'24]

○ **Zheyuan Zhang**, Zehong Wang, Shifu Hou, Evan Hall, Landon Bachman, Vincent Galassi, Jasmine White, Nitesh V Chawla, Chuxu Zhang, Yanfang Ye

[4] MAPRO: Recasting Multi-Agent Prompt Optimization as Maximum a Posteriori Inference [Preprint Oct.25]

○ **Zheyuan Zhang**, Lin Ge, Hongjiang Li, Weicheng Zhu, Chuxu Zhang, Yanfang Ye

[5] AgentRouter: A Knowledge-Graph-Guided LLM Router for Collaborative Multi-Agent Question Answering [Preprint Oct.25]

○ **Zheyuan Zhang**, Kaiwen Shi, Zhengqing Yuan, Zehong Wang, Tianyi Ma, Keerthiram Murugesan, Vincent Galassi, Chuxu Zhang, Yanfang Ye

[6] NG-Router: Graph-Supervised Multi-Agent Collaboration for Nutrition Question Answering [Preprint Oct.25]

○ Kaiwen Shi, **Zheyuan Zhang**, Zhengqing Yuan, Keerthiram Murugesan, Vincent Galassi, Chuxu Zhang, Yanfang Ye

[7] LLMs4All: A Review on Large Language Models for Research and Applications in Academic Disciplines [Preprint Sept.25]

○ Yanfang Ye, **Zheyuan Zhang**, Tianyi Ma, Zehong Wang, Yiyang Li, Shifu Hou, Weixiang Sun, Kaiwen Shi, Yijun Ma, Wei Song, Ahmed Abbasi, Ying Cheng, Jane Cleland-Huang, Steven Corcelli, Patricia Culligan, Robert Goulding, Ming Hu, Ting Hua, John Lalor, Fang Liu, Tengfei Luo, Ed Maginn, Nuno Moniz, Jason Rohr, Brett Savoie, Daniel Slate, Tom Stapleford, Matthew Webber, Olaf Wiest, Johnny Zhang, Nitesh Chawla

[8] Interpretable Graph-Language Modeling for Detecting Youth Illicit Drug Use [Preprint Oct.25]

○ Yiyang Li, Zehong Wang, Zhengqing Yuan, **Zheyuan Zhang**, Keerthiram Murugesan, Chuxu Zhang, Yanfang Ye

[9] Food4All: A Multi-Agent Framework for Real-time Free Food Discovery with Integrated Nutritional Metadata [Preprint Oct.25]

○ Zhengqing Yuan, Yiyang Li, Weixiang Sun, **Zheyuan Zhang**, Kaiwen Shi, Keerthiram Murugesan and Yanfang Ye

[10] AutoData: A Multi-Agent System for Open Web Data Collection [NeurIPS'25]

○ Tianyi Ma, Yiyue Qian, **Zheyuan Zhang**, Zehong Wang, Xiaoye Qian, Feifan Bai, Yifan Ding, Xuwei Luo, Shinan Zhang, Keerthiram Murugesan, Chuxu Zhang, Yanfang Ye

[11] Scalable Graph Generative Modeling via Substructure Sequences [NeurIPS'25]

○ Zehong Wang, **Zheyuan Zhang**, Tianyi Ma, Chuxu Zhang, Yanfang Ye

[12] LLM-Empowered Class Imbalanced Graph Prompt Learning for Online Drug Trafficking Detection [ACL Findings'25]

○ Tianyi Ma, Yiyue Qian, Zehong Wang, **Zheyuan Zhang**, Chuxu Zhang, Yanfang Ye

[13] Neural Graph Pattern Machine [ICML'25]

○ Zehong Wang, **Zheyuan Zhang**, Tianyi Ma, Nitesh V Chawla, Chuxu Zhang, Yanfang Ye

[14] Towards Graph Foundation Models: Learning Generalities Across Graphs via Task-trees [ICML'25]

○ Zehong Wang, **Zheyuan Zhang**, Tianyi Ma, Nitesh V Chawla, Chuxu Zhang, Yanfang Ye

[15] Can LLMs Convert Graphs to Text-Attributed Graphs? [NAACL Main'25]

○ Zehong Wang, Sidney Liu, **Zheyuan Zhang**, Tianyi Ma, Chuxu Zhang, Yanfang Ye

[16] Computation Tree: A Transferable Pattern Towards Graph Foundation Models [NeuralIPS'24]

○ Zehong Wang, **Zheyuan Zhang**, Chuxu Zhang, Yanfang Ye

[17] Subgraph Pooling: Tackling Negative Transfer on Graphs [IJCAI'24]

○ Zehong Wang, **Zheyuan Zhang**, Chuxu Zhang, Yanfang Ye

[18] SimMLP: Training MLPs on Graphs without Supervision [WSDM'25]

○ Zehong Wang, **Zheyuan Zhang**, Chuxu Zhang, Yanfang Ye

[19] EfficientLLM: Efficiency in Large Language Models [Preprint May.25]

○ Zhengqing Yuan, Weixiang Sun, Yixin Liu, Huichi Zhou, Rong Zhou, Yiyang Li, **Zheyuan Zhang**, Wei Song, Yue Huang, Haolong Jia, Keerthiram Murugesan, Yu Wang, Lifang He, Jianfeng Gao, Lichao Sun, Yanfang Ye

[20] Graph Foundation Models: A Comprehensive Survey [Preprint May.25]

○ Zehong Wang, Zheyuan Liu, Tianyi Ma, Jiazheng Li, **Zheyuan Zhang**, Xingbo Fu, Yiyang Li, Zhengqing Yuan, Wei Song, Yijun Ma, Qingkai Zeng, Xiusi Chen, Jianan Zhao, Jundong Li, Meng Jiang, Pietro Lio, Nitesh Chawla, Chuxu Zhang, Yanfang Ye

Education

University of Notre Dame *South Bend, IN*
Ph.D. in Computer Science *Jul. 2023 – Present*

○ **Research Applications:** Recommendation System, Fraud Detection, Health Intellengence.

New York University, M.S. *Aug. 2019 – Aug. 2020*

Wuhan University, B.E. *Sept. 2015 – Jun. 2019*

○ **Awards:** 1st Class Scholarship 2016 - 2018, Leijun Scholarship 2018

Service

○ **Conference Senior Area Chair:** KDD'26

○ **Conference Reviewer:** ICLR'26, KDD'26, KDD'25, WWW'25, RelKDD'24, SDM'25

○ **Journal Reviewer:** Transactions on Big Data, Data Mining and Knowledge Discovery, TIST, TMLR, TWEB, PLOS Digital Health.

○ **Teaching Assistant:** FA23-CSE-40113, SP24-CSE-40113: Design/Analysis of Algorithms

Invited Talks

○ MOPI-HFRS (Sept. 2024) @ IBM Research, Thomas J. Watson Research Center, Yorktown Heights.

○ LLM Agents for Code Generation (Jun. 2025) @ Amazon, Seattle.