

Binchi Zhang

Updated September 24, 2026

Department of Electrical and Computer Engineering

University of Virginia

Office: C-308 Thornton Hall

Phone: (434) 257-8089

Email: epb6gw@virginia.edu

Homepage: <https://zhangbinchi.github.io/>

Education

University of Virginia

Charlottesville, VA, USA

Ph.D. in Computer Engineering

Sep 2022 – Present

Advisor: Prof. Jundong Li

Xi'an Jiaotong University

Xi'an, Shaanxi, China

B.E. in Electrical Engineering

Sep 2018 – Jun 2022

GPA: 90.88/100, Special Class for the Gifted Young

Research Interests

Responsible AI, Large Language Models, Parametric Knowledge Management, and Graph Mining

My research focuses on parametric knowledge management for AI systems, where knowledge is encoded in external model parameters, allowing flexible and lifelong knowledge update, removal, and fusion. I study both fundamental problems (certification and verification of knowledge removal, parameter space symmetry) and practical challenges (knowledge conflicts, agentic memory, personalization, generative recommendation, and LLM safety and cultural alignment).

Publications

(* indicates equal contribution)

Teaching LLMs to Remember: Parametric Memory via Personalized Adapter Weights.

Binchi Zhang, Tianhao Wang, Jing Yang, Henry Kautz, Jundong Li.
In submission.

Exploiting Symmetry in Low-Rank Decomposition for LoRA Fusion.

Zaiyi Zheng*, **Binchi Zhang***, Liang Wu, Liangjie Hong, Jundong Li.
In submission.

When Safety Becomes An Outlier: Understanding the Retention of LLM Safety Behaviors.

Binchi Zhang*, Hadi Abdullah*, Yiwei Cai, Jundong Li.
Annual Conference on Neural Information Processing Systems (NeurIPS), 2026.

Enhancing Representation Learning of Text-Attributed Graphs in Federated Graph Learning.

Xingbo Fu, Zihan Chen, **Binchi Zhang**, Chen Chen, Jundong Li.

Learning on Graphs Conference (LoG), 2026.

LOKA: Conflict-Aware LLM Knowledge Update with Adaptive Knowledge Memory.

Binchi Zhang, Zhengzhang Chen, Zaiyi Zheng, Jundong Li, Haifeng Chen.

Annual Meeting of the Association for Computational Linguistics (ACL), 2026.

Mind the Gap in Cultural Alignment: Task-Aware Culture Management for Large Language Models.

Binchi Zhang, Xujiang Zhao, Jundong Li, Haifeng Chen, Zhengzhang Chen.

Annual Meeting of the Association for Computational Linguistics (ACL), 2026.

A Survey of Weight Space Learning: Understanding, Representation, and Generation.

Xiaolong Han*, Zehong Wang*, Bo Zhao, **Binchi Zhang**, Jundong Li, Damian Borth, Rose Yu, Haggai Maron, Yanfang Ye, Lu Yin, Ferrante Neri

arXiv preprint 2026.

Safety in Graph Machine Learning: Threats and Safeguards.

Song Wang, Yushun Dong, **Binchi Zhang**, Zihan Chen, Xingbo Fu, Yinhan He, Cong Shen, Chuxu Zhang, Nitesh V. Chawla, Jundong Li.

IEEE Transactions on Knowledge and Data Engineering (TKDE), 2026.

Certified Defense on the Fairness of Graph Neural Networks.

Yushun Dong, **Binchi Zhang**, Hanghang Tong, Jundong Li.

ACM SIGKDD Conference on Knowledge Discovery & Data Mining (KDD), 2026.

GraphTOP: Graph Topology-Oriented Prompting for Graph Neural Networks.

Xingbo Fu, Zhenyu Lei, Zihan Chen, **Binchi Zhang**, Chuxu Zhang, Jundong Li.

Annual Conference on Neural Information Processing Systems (NeurIPS), 2025.

Beyond the Permutation Symmetry of Transformers: The Role of Rotation for Model Fusion.

Binchi Zhang*, Zaiyi Zheng*, Zhengzhang Chen, Jundong Li.

International Conference on Machine Learning (ICML), 2025 (Spotlight).

Virtual Nodes Can Help: Tackling Distribution Shifts in Federated Graph Learning.

Xingbo Fu, Zihan Chen, Yinhan He, Song Wang, **Binchi Zhang**, Chen Chen, Jundong Li.

AAAI Conference on Artificial Intelligence (AAAI), 2025.

Understanding and Modeling Job Marketplace with Pretrained Language Models.

Yaochen Zhu, Liang Wu, **Binchi Zhang**, Song Wang, Qi Guo, Liangjie Hong, Luke Simon, Jundong Li.

ACM International Conference on Information and Knowledge Management (CIKM), Applied Research Paper, 2024

Federated Graph Learning with Graphless Clients.

Xingbo Fu, Song Wang, Yushun Dong, **Binchi Zhang**, Chen Chen, Jundong Li.

Transactions on Machine Learning Research (TMLR), 2024.

IDEA: A Flexible Framework of Certified Unlearning for Graph Neural Networks.

Yushun Dong, **Binchi Zhang**, Zhenyu Lei, Na Zou, Jundong Li.

ACM SIGKDD Conference on Knowledge Discovery & Data Mining (KDD), 2024.

Federated Graph Learning with Structure Proxy Alignment.

Xingbo Fu, Zihan Chen, **Binchi Zhang**, Chen Chen, Jundong Li.

ACM SIGKDD Conference on Knowledge Discovery & Data Mining (KDD), 2024.

Verification of Machine Unlearning is Fragile.

Binchi Zhang, Zihan Chen, Cong Shen, Jundong Li.

International Conference on Machine Learning (ICML), 2024.

Towards Certified Unlearning for Deep Neural Networks.

Binchi Zhang, Yushun Dong, Tianhao Wang, Jundong Li.

International Conference on Machine Learning (ICML), 2024.

Adversarial Attacks on Fairness of Graph Neural Networks.

Binchi Zhang, Yushun Dong, Chen Chen, Yada Zhu, Minnan Luo, Jundong Li.

International Conference on Learning Representations (ICLR), 2024.

RELIANT: Fair Knowledge Distillation for Graph Neural Networks.

Yushun Dong, **Binchi Zhang**, Yiling Yuan, Na Zou, Qi Wang, Jundong Li.

SIAM International Conference on Data Mining (SDM), 2023.

AHEAD: A Triple Attention Based Heterogeneous Graph Anomaly Detection Approach.

Shujie Yang, **Binchi Zhang**, Shangbin Feng, Zhaoxuan Tan, Qinghua Zheng, Ziqi Liu, Minnan Luo.
Chinese Intelligent Automation Conference (CIAC), 2023 (Best Application Paper Finalist).

Federated Graph Machine Learning: A Survey of Concepts, Techniques, and Applications.

Xingbo Fu, **Binchi Zhang**, Yushun Dong, Chen Chen, Jundong Li.
SIGKDD Explorations Newsletter, 2022.
A short version appears in *FedGraph @ CIKM 2022 (Spotlight)*.

Twibot-22: Towards Graph-Based Twitter Bot Detection.

Shangbin Feng*, Zhaoxuan Tan*, Herun Wan*, Ningnan Wang*, Zilong Chen*, **Binchi Zhang***, Qinghua Zheng, Wenqian Zhang, Zhenyu Lei, Shujie Yang, Xinshun Feng, Qingyue Zhang, Hongrui Wang, Yuhao Liu, Yuyang Bai, Heng Wang, Zijian Cai, Yanbo Wang, Lijing Zheng, Zihan Ma, Jundong Li, Minnan Luo.
Advances in Neural Information Processing Systems (NeurIPS), Datasets and Benchmarks Track, 2022.

Tackling the Local Bias in Federated Graph Learning.

Binchi Zhang, Minnan Luo, Shangbin Feng, Ziqi Liu, Jun Zhou, Qinghua Zheng.
arXiv preprint 2021.

Tutorials

Federated Graph Learning: Recent Advances and Future Directions.

Xingbo Fu, Zihan Chen, **Binchi Zhang**, Chen Chen, Jundong Li.
SIAM International Conference on Data Mining (SDM), 2025.

Industrial Experience

Research Intern, LinkedIn.

May 2026 – Aug 2026

LLM Reasoning and Generative Recommendation: Investigate the faithfulness of Chain-of-Thought and implicit reasoning in LLM-based generative recommenders.

Counterfactual Evaluation: Design a counterfactual evaluation framework and sensitivity metrics to quantify reasoning faithfulness and consistency beyond ranking accuracy and surface plausibility.

Reasoning Model Training: Build reproducible SFT and RL pipelines for LLM-based generative recommenders, benchmarking multiple state-of-the-art recommenders and systematically studying reasoning format, curriculum, quality, self-distillation, and reward.

Mentor: Dr. Eric Qasemi, Manager: Dr. Bo Long and Dr. Shuang Song

Research Intern, Visa Research.

May 2025 – Aug 2025

LLM Safety and Alignment: Diagnose a structural failure mode in safety alignment, modeling refusal-based safeguards as low-entropy outlier behaviors vulnerable to continual fine-tuning.

SFT and DPO Training: Develop natural and semantically grounded targets to improve safety training pipelines across SFT and DPO backbones, including trade-offs on harmlessness, helpfulness, and naturalness.

Evaluation and Experimental Design: Introduce a perplexity-based naturalness metric to quantify safety retention and distributional shift under downstream adaptation.

Mentor: Dr. Hadi Abdullah, Manager: Dr. Yiwei Cai

Research Intern, NEC Laboratories America.

Oct 2024 – Mar 2025

Continual Learning with Modular Memory: Design an external LLM knowledge memory to flexibly add new knowledge and remove unwanted knowledge without catastrophic forgetting.

Knowledge Conflict Resolution: Analyze gradient conflicts between learning and unlearning objectives and introduce a knowledge allocation resolution.

Cultural Alignment and Low-Resource Data Synthesis: LLM cultural alignment by aligning both task structures and cultural norms. Design an automatic pipeline that synthesizes task-specific, culture-grounded training data from web sources.

Benchmarking and Evaluation: Construct and evaluate knowledge update benchmarks across LLM updating, unlearning, and editing methods. Cultural evaluation across 10 national cultures and culture-sensitive tasks.

Mentor: Dr. Zhengzhang Chen, Manager: Dr. Haifeng Chen

Academic Experience

Research Assistant, Department of ECE, UVA

Sep 2022 – Present

Advisor: Dr. Jundong Li

Teaching Assistant, Department of ECE, UVA

Sep 2023 – Dec 2023

ECE 6501: Convex Optimization

Data Justice Academy Peer Mentor, UVA

Jun 2023 – Aug 2023

Honors

Thinking Machines Tinker Research Grant, \$5,000 in API credits 2026

OpenAI Cybersecurity Grant Program, \$2,500 in API credits 2025

SIAM Travel Award for SDM 2025

Wilson Bicentennial Grad Fellowship, Department of ECE, UVA 2024

NSF Graduate Student Travel Award 2023

Meritorious Winner, Mathematical Contest in Modeling 2021

Service

Conference Reviewer: NeurIPS, ICLR, ICML, AAAI, AISTATS, WWW, CIKM.

Journal Reviewer: TPAMI, TIFS, TKDD, TNNLS.