

Yingqian Cui

CONTACT INFORMATION	428 S Shaw Ln Rm 3308 East Lansing, MI 48824 Office: Engineering Building 3308 E-mail: cuiyingq@msu.edu	Homepage: http://cse.msu.edu/yingqiancui GitHub: https://github.com/Yingqiancui Phone: (+1) 5174889783
RESEARCH INTERESTS	- Reasoning of Large Language Models - Trusworthy AI: Safety, Robustness and Fairness - Theory in Deep Learning	
EDUCATION	Michigan State University, USA - Ph.D., Computer Science and Engineering - Advisor: Prof. Jiliang Tang & Prof. Yue Xing	08/2022 – Present
	Hong Kong Baptist University, HKSAR Bachelor of Science (Honors.) in Mathematics and Statistics	09/2018 – 06/2022
PROFESSIONAL EXPERIENCE	Data Science and Engineering Lab, Michigan State University Research Assistant, Department of Computer Science and Engineering - Topic: Trustworthy AI; Deep Learning Theory; LLM Reasoning - Supervisor: Dr. Jiliang Tang & Dr. Yue Xing	08/2022 – Present
	Microsoft Applied Scientist Intern	05/2026 – 08/2026
	Amazon.com Inc, USA Applied Scientist Intern - Topic 1: Research on advancing fine-grained Chain-of-Thought reasoning in LLMs to better balance efficiency and accuracy - Topic 2: Research on improving test-time scaling search methods for LLM reasoning - Topic 3: Research on the limitations of latent reasoning methods	10/2024 – 01/2026
	The Hong Kong Polytechnic University (PolyU), HKSAR Research Assistant - Topic: Copyright protection of image data against the plagiarism of AI Generative Models - Mentor: Dr. Wenqi Fan	05/2023 – 08/2023
PUBLICATION	Google Scholar: https://scholar.google.com/citations?user=3p67r08AAAAJ&hl=zh-CN&oi=ao Published Papers * indicates equal contribution “DiffusionShield: A Watermark for Data Copyright Protection against Generative Diffusion Models” Yingqian Cui*, Jie Ren*, Han Xu, Pengfei He, Hui Liu, Lichao Sun, Yue Xing, Jiliang Tang ACM SIG Conference on Knowledge Discovery and Data Mining (SIGKDD) Explorations, 2024	

- “FT-Shield: A Watermark Against Unauthorized Fine-tuning in Text-to-Image Diffusion Models”
Yingqian Cui, Jie Ren, Yuping Lin, Han Xu, Pengfei He, Yue Xing, Wenqi Fan, Hui Liu, Jiliang Tang
ACM SIG Conference on Knowledge Discovery and Data Mining (SIGKDD) Explorations, 2024
- “A Theoretical Understanding of Chain-of-Thought: Coherent Reasoning and Error-Aware Demonstration”
Yingqian Cui, Pengfei He, Xianfeng Tang, Qi He, Chen Luo, Jiliang Tang, Yue Xing
Artificial Intelligence and Statistics (AISTATS), 2025
- “Superiority of multi-head attention in in-context linear regression”
Yingqian Cui, Jie Ren, Pengfei He, Hui Liu, Jiliang Tang, Yue Xing
Artificial Intelligence and Statistics (AISTATS), 2025
- “Towards the Effect of Examples on In-Context Learning: A Theoretical Case Study”
Pengfei He, **Yingqian Cui**, Han Xu, Hui Liu, Makoto Yamada, Jiliang Tang, Yue Xing
Stat, 2024
- “Stepwise perplexity-guided refinement for efficient chain-of-thought reasoning in large language models”
Yingqian Cui, Pengfei He, Jingying Zeng, Hui Liu, Xianfeng Tang, Zhenwei Dai, Yan Han, Chen Luo, Jing Huang, Zhen Li, Suhang Wang, Yue Xing, Jiliang Tang, Qi He
Findings of the Association for Computational Linguistics: ACL 2025
- “Sharpness-Aware Data Poisoning Attack”
Pengfei He, Han Xu, Jie Ren, **Yingqian Cui**, Hui Liu, Charu C. Aggarwal, Jiliang Tang
Spotlight paper (5%) of International Conference on Learning Representations (ICLR), 2024
- “Stealthy Backdoor Attack via Confidence-driven Sampling”
Pengfei He, Han Xu, Jie Ren, **Yingqian Cui**, Shenglai Zeng, Yue Xing, Jiliang Tang, Makoto Yamada, Mohammad Sabokrou
Transactions on Machine Learning Research (TMLR), 2024
- “A robust semantics-based watermark for large language model against paraphrasing”
Jie Ren, Han Xu, Yiding Liu, **Yingqian Cui**, Shuaiqiang Wang, Dawei Yin, Jiliang Tang
Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2024
- “On the Generalization of Training-based ChatGPT Detection Methods”
Han Xu, Jie Ren, Pengfei He, Shenglai Zeng, **Yingqian Cui**, Amy Liu, Hui Liu, Jiliang Tang
Findings of the Association for Computational Linguistics: EMNLP 2024
- “Six-CD: Benchmarking Concept Removals for Benign Text-to-image Diffusion Models”
Jie Ren, Kangrui Chen, **Yingqian Cui**, Shenglai Zeng, Hui Liu, Yue Xing, Jiliang Tang, Lingjuan Lyu
Conference on Computer Vision and Pattern Recognition (CVPR), 2025
- “Retrieval Heads are Dynamic”
Yuping Lin, Zitao Li, Yue Xing, Pengfei He, **Yingqian Cui**, Yaliang Li, Bolin Ding, Jingren Zhou, Jiliang Tang
Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (ACL), 2026
- “Adaptive Test-Time Reasoning via Reward-Guided Dual-Phase Search”
Yingqian Cui, Zhenwei Dai, Pengfei He, Bing He, Hui Liu, Xianfeng Tang, Jingying Zeng, Suhang Wang, Yue Xing, Jiliang Tang, Benoit Dumoulin
Findings of the Association for Computational Linguistics: ACL 2026

Submitted Papers

- “How Do Latent Reasoning Methods Perform Under Weak and Strong Supervision? (Preprint)”
Yingqian Cui, Zhenwei Dai, Bing He, Zhan Shi, Hui Liu, Rui Sun, Zhiji Liu, Yue Xing, Jiliang Tang, Benoit Dumoulin
- “Adaptive Multi-Step Lookahead Decoding for Diffusion Language Models (Preprint)”
Yingqian Cui, Wei Deng, Lantao Mei, Hang Li, Charu C Aggarwal, Hui Liu, Yue Xing
- “Copyright Protection in Generative AI: A Technical Perspective’ (Preprint)”
Jie Ren, Han Xu, Pengfei He, **Yingqian Cui**, Shenglai Zeng, Jiankun Zhang, Hongzhi Wen, Jiayuan Ding, Hui Liu, Yi Chang, Jiliang Tang
- “SoK: Machine Unlearning for Large Language Models” (Preprint)
Jie Ren, Yue Xing, **Yingqian Cui**, Charu C Aggarwal, Hui Liu
- Crafting Reversible SFT Behaviors in Large Language Models (Preprint)
Yuping Lin, Pengfei He, Yue Xing, **Yingqian Cui**, Jiayuan Ding, Subhabrata Mukherjee, Hui Liu, Zhen Xiang
- “** Important** You should give me full credits!”: Exploring Prompt Injection Attacks on LLM-Based Automatic Grading Systems (Preprint)
Hang Li, Fedor Filippov, Yuping Lin, Pengfei He, Kaiqi Yang, Yucheng Chu, **Yingqian Cui**, Hui Liu, Jiliang Tang

Workshop Papers

- “EnTruth: Enhancing the Traceability of Unauthorized Dataset Usage in Text-to-image Diffusion Models with Minimal and Robust Alterations”
Jie Ren, **Yingqian Cui**, Chen Chen, Vikash Sehwal, Yue Xing, Jiliang Tang, Lingjuan Lyu
NeurIPS 2025 Workshop GenProCC

HONORS AND AWARDS

- SDM’24 Doctoral Forum Travel Award 2024
- AISTATS 2025 Travel Award 2025
- Scholastic Award for BSc (Hons.) in Mathematics and Statistics (top 2%), HKBU 2022
- President’s Honour Roll, HKBU 2018 - 2022
- Mr Li Men Jan Prize in Mathematics, HKSAR 2022
- C.V. Starr Scholarship Fund, HKSAR 2021
- Former Civil Engineering Department Head Dr. Yan Sze Kwan Memorial Scholarship, HKBU 2021
- Scholarship for Outstanding Mainland Students, Lo Ka Chung Charitable Foundation Ltd, HKBU 2020
- Mr. & Mrs. Lau Chor Tak Scholarship Scheme for Outstanding Mainland Students, HKBU 2019
- Sir Tseng Chi Lu Scholarship, HKBU 2019

TEACHING EXPERIENCE

- Teaching Assistant for CSE 482 Big data analytics (Undergraduate Level), MSU 2023
– Duties included online discussions, office hours and grading.
- Teaching Assistant for CSE 480 Database System (Undergraduate Level), MSU 2024
– Duties included online discussions and office hours.

SERVICES

Conference Reviewer & Journal Reviewer

- Transactions on Knowledge and Data Engineering (TKDE) 2023-2025
- ACL Rolling Review (ARR) 2024-2026
- Transactions on Knowledge Discovery from Data (TKDD) 2023-2025
- ACM International Conference on Information and Knowledge Management (CIKM) 2024-2025
- IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) 2024
- Artificial Intelligence and Statistics (AISTATS) 2024-2025
- The IEEE / CVF Computer Vision and Pattern Recognition Conference (CVPR) 2024
- Conference on Neural Information Processing Systems (NeurIPS) 2025
- International Conference on Learning Representations (ICLR) 2025

Conference External Reviewer

- SIGKDD Conference on Knowledge Discovery and Data Mining (KDD) 2023
- Conference on Neural Information Processing Systems (NeurIPS) 2023
- International Conference on Machine Learning (ICML) 2023 - 2024

Workshop Reviewer

- ICML 2025 Methods and Opportunities at Small Scale (MOSS) Workshop 2025
- ICML 2024 Workshop on Theoretical Foundations of Foundation Models (TF2M) 2024
- NeurIPS 2023 Workshop on Diffusion Models 2023

Student Representative & Ambassador

- Student Representative of Science Faculty Board, HKBU 2021 - 2022
- Student Representative of BSC MATH & STAT program, HKBU 2019 - 2022
- Non-Local Mentor, Undergraduate Hall Residence, HKBU 2019 - 2020