

Baixiang Huang

baixianghuang.github.io [Google Scholar](#) [LinkedIn](#) baixiang.huang@emory.edu

EDUCATION

Emory University	Jan 2023 – Present (Expected Graduation: Dec 2026)
Ph.D. in Computer Science	Advisor: Kai Shu, GPA: 4.0
Illinois Institute of Technology	Jan 2023 – Dec 2024
Ph.D. in Computer Science (Transfer Out)	Advisor: Kai Shu
National University of Singapore	Aug 2020 – Feb 2022
Master's Degree in Computer Science	Advisor: Bryan Hooi

PROFESSIONAL EXPERIENCE

Amazon Search	May 2025 – Aug 2025
Applied Scientist Intern	Mentor: Limeng Cui
Proposed a model editing framework for LLM personalization.	
[PDF] [GitHub] [Project Website]	

PUBLICATIONS AND PREPRINTS

- [1] Model Editing as a Double-Edged Sword: Steering Agent Ethical Behavior Toward Beneficence or Harm
Baixiang Huang, Zhen Tan, Haoran Wang, Zijie Liu, Dawei Li, Ali Payani, Huan Liu, Tianlong Chen, Kai Shu.
Proceedings of the AAAI Conference on Artificial Intelligence (AAAI 2026 Oral)
[PDF] [GitHub] [Project Website]
- [2] Can Editing LLMs Inject Harm?
Canyu Chen*, **Baixiang Huang***, Zekun Li, Zhaorun Chen, Shiyang Lai, Xiong Xiao Xu, Jia-Chen Gu, Jindong Gu, Huaxiu Yao, Chaowei Xiao, Xifeng Yan, William Wang, Philip Torr, Dawn Song, Kai Shu. (* denotes equal contribution)
Proceedings of the AAAI Conference on Artificial Intelligence (AAAI 2026)
[PDF] [GitHub] [Project Website]
- [3] Can Knowledge Editing Really Correct Hallucinations?
Baixiang Huang, Canyu Chen, Xiong Xiao Xu, Ali Payani, Kai Shu.
The Thirteenth International Conference on Learning Representations (ICLR 2025)
[PDF] [GitHub] [Project Website]
- [4] SST: Multi-Scale Hybrid Mamba-Transformer Experts for Long-Short Range Time Series Forecasting
Xiong Xiao Xu, Canyu Chen, Yueqing Liang, **Baixiang Huang**, Guangji Bai, Liang Zhao, Kai Shu.
Proceedings of the 34th ACM International Conference on Information and Knowledge Management (CIKM 2025) [PDF] [GitHub]
- [5] Can Large Language Models Identify Authorship?
Baixiang Huang, Canyu Chen, Kai Shu.

Empirical Methods in Natural Language Processing (EMNLP Findings 2024)

[PDF] [GitHub] [Project Website]

- [6] Authorship Attribution in the Era of LLMs: Problems, Methodologies, and Challenges

Baixiang Huang, Canyu Chen, Kai Shu.

ACM SIGKDD Explorations 2024 [PDF] [GitHub] [Project Website]

- [7] TAP: A Comprehensive Data Repository for Traffic Accident Prediction in Road Networks

Baixiang Huang, Bryan Hooi, Kai Shu.

Proceedings of the 31st ACM International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL 2023) [PDF] [GitHub]

Preprints

- [1] Rethinking the Value of Multi-Agent Workflow: A Strong Single Agent Baseline

Jiawei Xu, Arief Koesdwiady, Sisong Bei, Yan Han, **Baixiang Huang**, Dakuo Wang, Yutong Chen, Zheshen Wang, Peihao Wang, Pan Li, Ying Ding.

arXiv preprint 2026. [PDF]

- [2] Rethinking Memory Mechanisms of Foundation Agents in the Second Half

Wei-Chieh Huang, Weizhi Zhang, Yueqing Liang, Yuanchen Bei, Yankai Chen, Tao Feng, Xinyu Pan, Zhen Tan, Yu Wang, Tianxin Wei, Shanglin Wu, Ruiyao Xu, Liangwei Yang, Rui Yang, Woosong Yang, Chin-Yuan Yeh, Hanrong Zhang, Haozhen Zhang, Siqi Zhu, Henry Peng Zou, Wanjia Zhao, Song Wang, Wujiang Xu, Zixuan Ke, Zheng Hui, Dawei Li, Yaozu Wu, Langzhou He, Chen Wang, Xiong Xiao Xu, **Baixiang Huang**, Juntao Tan, Shelby Heinecke, Huan Wang, Caiming Xiong, Ahmed A Metwally, Jun Yan, Chen-Yu Lee, Hanqing Zeng, Yinglong Xia, Xiaokai Wei, Ali Payani, Haitong Ma, Wenya Wang, Chengguang Wang, Yu Zhang, Xin Wang, Yongfeng Zhang, Jiaxuan You, Hanghang Tong, Xiao Luo, Yizhou Sun, Wei Wang, Julian McAuley, James Zou, Jiawei Han, Philip S Yu, Kai Shu.

arXiv preprint 2026. [PDF]

- [3] Towards Effective Model Editing for LLM Personalization

Baixiang Huang, Limeng Cui, Jiapeng Liu, Haoran Wang, Jiawei Xu, Zhuoyue Tan, Yutong Chen, Chen Luo, Yi Liu, Kai Shu.

arXiv preprint 2025. [PDF] [GitHub] [Project Website]

- [4] Who's Your Judge? On the Detectability of LLM-Generated Judgments

Dawei Li, Zhen Tan, Chengshuai Zhao, Bohan Jiang, **Baixiang Huang**, Pingchuan Ma, Abdullah Alnaibari, Kai Shu, Huan Liu.

arXiv preprint 2025. [PDF] [GitHub] [Project Website]

- [5] Privacy-Aware Decoding: Mitigating Privacy Leakage of Large Language Models in Retrieval-Augmented Generation

Haoran Wang, Xiong Xiao Xu, **Baixiang Huang**, Kai Shu.

arXiv preprint 2025. [PDF] [GitHub]

TEACHING

- Undergraduate TA for CS253: Data Structures and Algorithms, Fall 2025
- Undergraduate TA for CS470: Data Mining, Spring 2025
- Graduate TA for CS536: Science of Programming, Fall 2024

ACADEMIC SERVICE

- Reviewer: ICLR {2024, 2025}, NeurIPS {2026}, KDD {2024, 2026}, SIGIR {2023, 2024}, CIKM {2023}, ICDM {2023}, TheWebConf {2023}
- Program Committee: AAAI {2023, 2024, 2025}, IEEE CogMI {2025}

PROJECT EXPERIENCE

- Towards Effective and Robust Inference-time Knowledge Editing, Cisco, 2025-2026
- Towards Fairness in the Real World under Generalization, Privacy and Robustness Challenges, NSF, 2024-2029
- An Integrated Open-Source Ecosystem for Anomaly Detection, NSF, 2024-2026
- Targeting Challenges in Computational Disinformation Research to Enhance Attribution, Detection, and Explanation, NSF, 2023-2026.
- Countering Misinformation in the Era of Large Language Models, DHS, 2023-2025
- Microsoft Accelerate Foundation Models Research Award, Microsoft, 2023-2025
- Towards Factual Large Language Models, Cisco, 2023-2024

SKILLS

- Frameworks: PyTorch, TensorFlow, MyBatis, Spring Boot, Vue
- Programming Languages: Java, Python, JavaScript, C++, HTML/CSS