

Yue Zhao

CONTACT	✉ yue.z@usc.edu	213-821-2369
INFORMATION	GitHub LinkedIn Viterbi Web USC Faculty Directory Google Scholar  22,000+ GitHub Stars	CS Department, GCS Hall Los Angeles, CA United States, 90089 Department of Computer Science University of Southern California Top ~700 Worldwide
RESEARCH SUMMARY	My research builds reliable, safe, and scalable AI systems . I organize my work into two tiers: (1) advancing the scientific foundations of safety, robustness, and reliability in modern AI systems, and (2) developing system-level evaluation frameworks and applications that translate these foundations into real-world impact.	

Tier 1: Foundations of Reliable & Safe AI

I study why and how modern AI systems fail under distribution shift, uncertainty, and strategic pressure, and develop methods to make their behavior more predictable and reliable. This tier integrates two complementary research threads:

1. LLM & Agent Safety

Understanding and mitigating failure modes in large language models and agentic systems, including hallucinations, jailbreaks, privacy leakage, model extraction, and multi-agent instability.

- | | |
|---|---|
| <input type="checkbox"/> LLM Safety & Reliability | <input type="checkbox"/> Privacy & Model Extraction Risks |
| <input type="checkbox"/> Hallucination Mitigation | <input type="checkbox"/> Robust Reasoning & Routing |
| <input type="checkbox"/> Jailbreak Detection | <input type="checkbox"/> Failure Mode Analysis |
| <input type="checkbox"/> Multi-agent Reliability | |

2. Robustness & Failure Detection (Methodological Foundations)

Developing algorithms and benchmarks to identify abnormal or unreliable behavior, grounded in robustness, out-of-distribution generalization, and anomaly detection.

- | | |
|---|--|
| <input type="checkbox"/> Robustness | <input type="checkbox"/> Benchmarks & Evaluation |
| <input type="checkbox"/> OOD Generalization | <input type="checkbox"/> Multimodal & Graph Settings |
| <input type="checkbox"/> Anomaly Detection | <input type="checkbox"/> Data Efficiency |
| <input type="checkbox"/> Model Selection | |

Tier 2: System-Level Evaluation & Scientific/Societal Impact

I adopt a system-oriented perspective to evaluate, stress-test, and deploy reliable AI in realistic settings, and apply these methods to domains where failures carry high cost.

3. Evaluation & Benchmarking (System Perspective)

Designing scalable evaluation frameworks, benchmarks, and workflows that probe model and agent behavior under realistic and adversarial conditions. Open-source tools include **PyOD** (**35M+** downloads) and related libraries with **22K+** GitHub stars.

- | | |
|--|--|
| <input type="checkbox"/> System-Level Evaluation | <input type="checkbox"/> Reproducibility & Measurement |
| <input type="checkbox"/> Benchmarking & Stress Testing | <input type="checkbox"/> Agentic Workflows |
| <input type="checkbox"/> Scalable Evaluation Pipelines | <input type="checkbox"/> Deployment-Aware Analysis |

4. AI for Science & Society (Applications)

Applying reliable foundation models to high-impact scientific and societal domains, including climate and weather forecasting, healthcare and biomedicine, and political or social decision-making.

- | | |
|---|---|
| <input type="checkbox"/> AI for Science | <input type="checkbox"/> Healthcare & Biomedicine |
| <input type="checkbox"/> Scientific Foundation Models | <input type="checkbox"/> Computational Social Systems |
| <input type="checkbox"/> Climate & Weather Modeling | <input type="checkbox"/> Decision Modeling |

FULL-TIME	University of Southern California	
PROFESSIONAL	<i>Thomas Lord Department of Computer Science</i>	
EXPERIENCE	Assistant Professor (Tenure-Track)	Aug. 2023 - Present
	<ul style="list-style-type: none"> • Foundations Of Robust Trustworthy Intelligent Systems (FORTIS) Lab: Link • USC Machine Learning Center (MaSCle): Link 	
	Stealth Startup	
	<i>AI Research & Systems</i>	
	External Technical Advisor (Part-Time, Contract-based)	Mar. 2024 – Present
	PwC Canada	
	<i>Consulting & Deals</i>	
	Senior Consultant (Data Scientist)	Aug. 2017 - Jun. 2019
	Consultant (Data Scientist)	Feb. 2017 - Jul. 2017
EDUCATION	Carnegie Mellon University	Pittsburgh, PA
	<i>Ph.D. in Information Systems and Management</i>	Sep. 2019 - May. 2023
	<ul style="list-style-type: none"> • Affiliation: CMU automated learning systems group (Catalyst) and Data Analytics Techniques Algorithms (DATA) Lab • Advisors and Mentors: CMU: Prof. Leman Akoglu, Prof. Zhihao Jia, and Prof. George Chen. I collaborate with Prof. Jure Leskovec at Stanford, and Prof. Philip S. Yu at UIC. 	
	University of Toronto	Toronto, ON
	<i>Master of Science in Computer Science</i>	Sep. 2015 - Dec. 2016
	University of Cincinnati	Cincinnati, OH
	<i>Bachelor of Science in Computer Engineering</i>	Sep. 2010 - May. 2015
	Minor: Computer Science and Mathematics	

AWARDS, GRANTS, AND FUNDING			
As Principal Investigator (August 2023 onwards)			
USC CCSL Seed Funding	<i>Small Grant</i>	Jan. 2026	
Second Prize CCC Award @ IEEE ICDM, BlueSky Track	<i>Recognition</i>	Nov. 2025	
Best Short Paper Award @ ACM SIGSPATIAL	<i>Recognition</i>	Nov. 2025	
NSF POSE I	<i>Funding</i>	Aug. 2025	
Capital One Research Awards	<i>Grant</i>	Oct. 2024	
Amazon Research Awards	<i>Gift</i>	Aug. 2024	
Best Paper Award @ KDD Resource-Efficient Learning Workshop	<i>Recognition</i>	Aug. 2024	
NSF ATD	<i>Funding</i>	Aug. 2024	
NSF POSE II	<i>Funding</i>	Jun. 2024	
Google Cloud Research Innovators	<i>Recognition</i>	Mar. 2024	
AAAI New Faculty Highlights	<i>Recognition</i>	Feb. 2024	

Note: Monetary values represent my portion of the funding. Total project budgets may be larger.

Prior to Principal Investigator Role (Before August 2023)

Meta 2022 AI4AI Research Award (student co-PI)	<i>Recognition</i>	Oct. 2022
The Norton Labs Graduate Fellowship	<i>Fellowship</i>	Mar. 2022
CMU Presidential Fellowship	<i>Fellowship</i>	2019
Mitacs-Accelerate Research and Development Funding	<i>Funding</i>	2016-2017
University Global Award and Scholarship	<i>Scholarship</i>	2010-2015
Mantei/Mae Award & Scholar	<i>Award</i>	2012-2015
Engineer of the Month	<i>Recognition</i>	Jun. 2014

Note: Monetary values are omitted for awards and recognitions received prior to PI role.

Preprints & Under Submission

Note: *first authors and [†]corresponding authors if more than one.

96. Zhisheng Qi, Utkarsh Sahu, Li Ma, Haoyu Han, Ryan Rossi, Franck Dernoncourt, Mahantesh Halappanavar, Nesreen Ahmed, Yushun Dong, Yue Zhao, Yu Zhang, Yu Wang
 Benchmarking Knowledge-Extraction Attack and Defense on Retrieval-Augmented Generation
Under submission
arXiv preprint arXiv:2602.09319
95. Rujie Ye, Jiayi Zhang, Zhuoxin Liu, Zihao Zhu, Siyuan Yang, Li Li, Tianfu Fu, Franck Dernoncourt, Yue Zhao, Jiacheng Zhu, Ryan Rossi, Wenhao Chai, Zhengzhong Tu
 Agent Banana: High-Fidelity Image Editing with Agentic Thinking and Tooling
Under submission
arXiv preprint arXiv:2602.09084
94. Jiate Li, Defu Cao, Li Li, Wei Yang, Yuehan Qin, Chenxiao Yu, Tiannuo Yang, Ryan A. Rossi, Yan Liu, Xiyang Hu, Yue Zhao
 “Someone Hid It”: Query-Agnostic Black-Box Attacks on LLM-Based Retrieval
Under submission
arXiv preprint arXiv:2602.00364
93. Shawn Li, Chenxiao Yu, Zhiyu Ni, Hao Li, Charith Peris, Chaowei Xiao, Yue Zhao
 Defenses Against Prompt Attacks Learn Surface Heuristics
Under submission
arXiv preprint arXiv:2601.07185
92. Xiaolin Zhou, Zheng Luo, Yicheng Gao, Qixuan Chen, Xiyang Hu, Yue Zhao, Ruishan Liu
 Fairness or Fluency? An Investigation into Language Bias of Pairwise LLM-as-a-Judge
Under submission
arXiv preprint arXiv:2601.13649
91. Chenxiao Yu, Bowen Yi, Farzan Karimi-Malekabadi, Suhaib Abdurahman, Jinyi Ye, Shrikanth Narayanan, Yue Zhao, Morteza Dehghani
 Tracing Moral Foundations in Large Language Models
Under submission
arXiv preprint arXiv:2601.05437
90. Yixuan Du, Chenxiao Yu, Haoyan Xu, Ziyi Wang, Yue Zhao, Xiyang Hu
 Multimodal Generative Engine Optimization: Rank Manipulation for Vision-Language Model Rankers
Under submission
arXiv preprint arXiv:2601.12263
89. Jinbo Liu, Defu Cao, Yifei Wei, Tianyao Su, Yuan Liang, Yushun Dong, Yan Liu, Yue Zhao, Xiyang Hu
 Topology Matters: Measuring Memory Leakage in Multi-Agent LLMs
Under submission
arXiv preprint arXiv:2512.04668
88. Kay Liu, Yuwei Han, Haoyan Xu, Henry Peng Zou, Yue Zhao, Philip S. Yu
 TAGFN: A Text-Attributed Graph Dataset for Fake News Detection in the Age of LLMs
Under submission
arXiv preprint arXiv:2511.21624
87. Haoyan Xu, Ruizhi Qian, Zhengtao Yao, Ziyi Liu, Li Li, Yuqi Li, Yanshu Li, Wenqing Zheng, Daniele Rosa, Daniel Barcklow, Senthil Kumar, Jieyu Zhao, Yue Zhao
 LLM-Powered Text-Attributed Graph Anomaly Detection via Retrieval-Augmented Reasoning
Under submission
arXiv preprint arXiv:2511.17584
86. Haoyan Xu, Ruizhi Qian, Jiate Li, Yushun Dong, Minghao Lin, Hanson Yan, Zhengtao Yao, Qinghua Liu, Junhao Dong, Ruopeng Huang, Yue Zhao[†], Mengyuan Li[†]
 A Systematic Study of Model Extraction Attacks on Graph Foundation Models
Under submission
arXiv preprint arXiv:2511.11912

85. Yuexing Hao, Yue Huang, Haoran Zhang, Chenyang Zhao, Zhenwen Liang, Paul Pu Liang, Yue Zhao, Lichao Sun, Saleh Kalantari, Xiangliang Zhang, Marzyeh Ghassemi
The Role of Computing Resources in Publishing Foundation Model Research
Under submission
arXiv preprint arXiv:2510.13621
84. Wang Wei, Tiankai Yang, Hongjie Chen, Yue Zhao, Franck Dernoncourt, Ryan A. Rossi, Hoda Eldardiry
Learning to Route LLMs from Bandit Feedback: One Policy, Many Trade-offs
Under submission
arXiv preprint arXiv:2510.07429
83. Langzhou He, Junyou Zhu, Fangxin Wang, Junhua Liu, Haoyan Xu, Yue Zhao, Philip S. Yu, Qitian Wu
Can Molecular Foundation Models Know What They Don't Know? A Simple Remedy with Preference Optimization
Under submission
arXiv preprint arXiv:2509.25509
82. Yuehan Qin, Li Li, Defu Cao, Tiankai Yang, Yue Zhao
M3OOD: Automatic Selection of Multimodal OOD Detectors
Under submission
arXiv preprint arXiv:2508.11936
81. Bolin Shen, Eren Erman Ozguven, Yue Zhao, Guang Wang, Yiqun Xie, Yushun Dong
Learning from the Storm: A Multivariate Machine Learning Approach to Predicting Hurricane-Induced Economic Losses
Under submission
arXiv preprint arXiv:2506.17964
80. Haoyan Xu, Zhengtao Yao, Xuzhi Zhang, Ziyi Wang, Langzhou He, Yushun Dong, Philip S. Yu, Mengyuan Li, Yue Zhao
GLIP-OOD: Zero-Shot Graph OOD Detection with Foundation Model
Under submission
arXiv preprint arXiv:2504.21186
79. Haoyan Xu, Zhengtao Yao, Ziyi Wang, Zhan Cheng, Xiyang Hu, Mengyuan Li, Yue Zhao
Graph Synthetic Out-of-Distribution Exposure with Large Language Models
Under submission
arXiv preprint arXiv:2504.21198
78. Yiming Tang, Yi Fan, Chenxiao Yu, Tiankai Yang, Yue Zhao, Xiang Hu
StealthRank: LLM Ranking Manipulation via Stealthy Prompt Optimization
Under submission
arXiv preprint arXiv:2504.05804
77. Kaixiang Zhao, Lincan Li, Kaize Ding, Neil Zhenqiang Gong, Yue Zhao, Yushun Dong
A Survey of Model Extraction Attacks and Defenses in Distributed Computing Environments
Under submission
arXiv preprint arXiv:2502.16065
76. Shixuan Li, Wei Yang, Peiyu Zhang, Xiongye Xiao, Defu Cao, Yuehan Qin, Xiaole Zhang, Yue Zhao, Paul Bogdan
ClimateLLM: Efficient Weather Forecasting via Frequency-Aware Large Language Models
Under submission
arXiv preprint arXiv:2502.11059
75. Lincan Li, Jiaqi Li, Catherine Chen[†], Fred Gui[†], other collaborators, Yue Zhao[†], Yushun Dong[†]
Political-LLM: Large Language Models in Political Science
Under submission
arXiv preprint arXiv:2412.06864
74. Chenxiao Yu, Jinyi Ye, Yuangang Li, Zhaotian Weng, Zheng Li, Emilio Ferrara, Xiyang Hu[†], Yue Zhao[†]
A Large-Scale Simulation on Large Language Models for Decision-Making in Political Science
Under submission
arXiv preprint arXiv:2412.15291
73. Junda Wu, Hanjia Lyu, Yu Xia, Zhehao Zhang, Joe Barrow, Ishita Kumar, Mehnoosh Mirtahebi, Hongjie Chen, Ryan A. Rossi, Franck Dernoncourt, Tong Yu, Ruiyi Zhang, Jiuxiang Gu, Nesreen K. Ahmed, Yu Wang, Xiang Chen, Hanieh Deilamsalehy, Namyong Park, Sungchul Kim, Huanrui Yang,

Subrata Mitra, Zhengmian Hu, Nedim Lipka, Yue Zhao, Jiebo Luo, Julian McAuley
Personalized Multimodal Large Language Models: A Survey
Under submission
arXiv preprint arXiv:2412.02142

72. Han Bao, Yue Huang, Yanbo Wang, Jiayi Ye, Xiangqi Wang, Xiuying Chen, Yue Zhao, Tianyi Zhou, Mohamed Elhoseiny, Xiangliang Zhang
AutoDavis: Automatic and Dynamic Evaluation Protocol of Large Vision-Language Models on Visual Question-Answering?
ICML 2025 DataWorld Workshop
arXiv preprint arXiv:2410.21259

Peer-reviewed Journal Papers

71. Yuehan Qin, Shawn Li, Yi Nian, Xinyan Velocity Yu, Yue Zhao[†], Xuezhe Ma[†]
Don't Let It Hallucinate: Premise Verification via Retrieval-Augmented Logical Reasoning
Transactions on Machine Learning Research (TMLR), 2026
70. Haoyan Xu, Kay Liu, Zhengtao Yao, Philip S. Yu, Kaize Ding[†], Yue Zhao[†]
LEGO-Learn: Label-Efficient Graph Open-Set Learning
Transactions on Machine Learning Research (TMLR), 2025
69. Hao Dong, Gaetan Frusque, Yue Zhao, Eleni Chatzi, Olga Fink
NNG-Mix: Improving Semi-supervised Anomaly Detection with Pseudo-anomaly Generation
IEEE Transactions on Neural Networks and Learning Systems (TNNLS), 2024
68. Ling Yang*, Zhilong Zhang*, Yang Song, Shenda Hong, Runsheng Xu, Yue Zhao, Wentao Zhang, Bin Cui, Ming-Hsuan Yang
Diffusion Models: A Comprehensive Survey of Methods and Applications
ACM Computing Surveys (CSUR), 2023
(*equal contribution)
67. Yue Zhao*^{*}, Martin Q. Ma*, Xiaorong Zhang, Leman Akoglu
The Need for Unsupervised Outlier Model Selection: A Review and Evaluation of Internal Evaluation Strategies
ACM SIGKDD Explorations Newsletter (SIGKDD Explor.), 2023
(*equal contribution)
66. Kexin Huang*, Tianfan Fu*, Wenhao Gao*, Yue Zhao, Yusuf Roohani, Jure Leskovec, Connor W. Coley, Cao Xiao, Jimeng Sun, Marinka Zitnik
Artificial Intelligence Foundation for Therapeutic Science
Nature Chemical Biology (NCHEMB), 2022
(*equal contribution)
65. Yue Zhao*^{*}, Zheng Li*, Xiyang Hu, Nicola Botta, Cezar Ionescu, George H. Chen
ECOD: Unsupervised Outlier Detection Using Empirical Cumulative Distribution Functions
IEEE Transactions on Knowledge and Data Engineering (TKDE), 2022.
(*equal contribution)
64. Yue Zhao, Zain Nasrullah, Zheng Li
PyOD: A Python Toolbox for Scalable Outlier Detection
Journal of Machine Learning Research (JMLR), 2019.

Conference & Workshop Papers

63. Shawn Li, Ryan Rossi, Sungchul Kim, Sunav Choudhary, Franck Dernoncourt, Puneet Mathur, Zhengzhong Tu, Yue Zhao
Charts Are Not Images: On the Challenges of Scientific Chart Editing
International Conference on Learning Representations (ICLR), 2026
62. Weidi Luo, Qiming Zhang, Tianyu Lu, Xiaogeng Liu, Yue Zhao, Zhen Xiang, Chaowei Xiao
Doxing via the Lens: Revealing Privacy Leakage in Image Geolocation for Agentic Multi-Modal Large Reasoning Model
International Conference on Learning Representations (ICLR), 2026

61. Yue Huang, Chujie Gao, Siyuan Wu, Haoran Wang, Xiangqi Wang, Yujun Zhou, Yanbo Wang, Jiayi Ye, Jiawen Shi, Qihui Zhang, Yuan Li, Han Bao, Zhaoyi Liu, Tianrui Guan, Dongping Chen, Ruoxi Chen, other authors, Yue Zhao, other authors, Xiangliang Zhang
On the Trustworthiness of Generative Foundation Models: Guideline, Assessment, and Perspective
International Conference on Learning Representations (ICLR), 2026
<https://trustgen.github.io/>
60. Chengxuan Qian, Shuo Xing, Shawn Li, Yue Zhao, Zhengzhong Tu
DecAlign: Hierarchical Cross-Modal Alignment for Decoupled Multimodal Representation Learning
International Conference on Learning Representations (ICLR), 2026
59. Xiongxiao Xu, Haoran Wang, Yueqing Liang, Philip S. Yu, Yue Zhao, Kai Shu
Can Multimodal LLMs Perform Time Series Anomaly Detection?
The Web Conference (WWW), 2026
58. Bo Ni, Yu Wang, Leyao Wang, Branislav Kveton, Franck Dernoncourt, Yu Xia, Hongjie Chen, Reuben Luera, Samyadeep Basu, Subhojoyoti Mukherjee, Puneet Mathur, Nesreen K. Ahmed, Junda Wu, Li Li, Huixin Zhang, Ruiyi Zhang, Tong Yu, Sungchul Kim, Jiuxiang Gu, Zhengzhong Tu, Alexa Siu, Zichao Wang, Seunghyun Yoon, Nedim Lipka, Namyong Park, Zihao Lin, Trung Bui, Yue Zhao, Tyler Derr, Ryan A. Rossi
A Survey on LLM-based Conversational User Simulation
Conference of the European Chapter of the Association for Computational Linguistics (EACL), 2026
57. Ojas Nimase, Yue Zhao, Yushun Dong
Navigating Between Explainability and Extractability in Machine Learning as a Service
IEEE International Conference on Data Mining (ICDM) BlueSky Track, Second Prize CCC Award, 2025.
56. Yuangang Li, Yiqing Shen, Yi Nian, Jiechao Gao, Ziyi Wang, Chenxiao Yu, Shawn Li, Jie Wang, Xiyang Hu, Yue Zhao
Mitigating Hallucinations in Large Language Models via Causal Reasoning
Thirty-Seventh AAAI Conference on Artificial Intelligence (AAAI), 2026
55. Tiansai Yang, Junjun Liu, Wingchun Siu, Jiahang Wang, Zhuangzhuang Qian, Chanjuan Song, Cheng Cheng, Xiyang Hu, Yue Zhao
AD-AGENT: A Multi-agent Framework for End-to-end Anomaly Detection
Findings of the Association for Computational Linguistics: IJCNLP-AACL, 2025.
54. Ruosi Shao, Md Shamim Seraj, Kangyi Zhao, Yingtao Luo, Lincan Li, Bolin Shen, Averi Bates, Yue Zhao, Chongle Pan, Lisa Hightow-Weidman, Shayok Chakraborty, Yushun Dong
LLM-Empowered Patient-Provider Communication: A Data-Centric Survey From a Clinical Perspective
Findings of the Association for Computational Linguistics: IJCNLP-AACL, 2025.
53. Li Li, Peilin Cai, Ryan A. Rossi, Franck Dernoncourt, Branislav Kveton, Junda Wu, Tong Yu, Lixin Song, Tiansai Yang, Yuehan Qin, Nesreen K. Ahmed, Samyadeep Basu, Subhojoyoti Mukherjee, Ruiyi Zhang, Yuxiao Zhou, Zichao Wang, Yue Huang, Yu Wang, Xiangliang Zhang, Philip S. Yu, Xiyang Hu, Yue Zhao
A Personalized Conversational Benchmark: Towards Simulating Personalized Conversations
NeurIPS Workshop on Multi-Turn Interactions in Large Language Models (MTI-LLM), Spotlight, 2025.
arXiv preprint arXiv:2505.14106
52. Zixiang Xu, Yanbo Wang, Yue Huang, Jiayi Ye, Haomin Zhuang, Zirui Song, Lang Gao, Chenxi Wang, Zhaorun Chen, Yujun Zhou, Sixian Li, Wang Pan, Yue Zhao, Jieyu Zhao, Xiangliang Zhang, Xiuying Chen
SocialMaze: A Benchmark for Evaluating Social Reasoning in Large Language Models
NeurIPS Workshop on Socially Responsible and Trustworthy Foundation Models (ResponsibleFM)
arXiv preprint arXiv:2505.23713
51. Yanbo Wang, Zixiang Xu, Yue Huang, Xiangqi Wang, Zirui Song, Lang Gao, Chenxi Wang, Xiangru Tang, Yue Zhao, Arman Cohan, Xiangliang Zhang, Xiuying Chen
DyFlow: Dynamic Workflow Framework for Agentic Reasoning
Advances in Neural Information Processing Systems (NeurIPS), 2025
50. Shawn Li, Jiashu Qu, Yuxiao Zhou, Yuehan Qin, Tiansai Yang, Yue Zhao
Treble Counterfactual VLMs: A Causal Approach to Hallucination
Findings of the Association for Computational Linguistics: EMNLP, 2025.

49. Yuangang Li, Jiaqi Li, Zhuo Xiao, Tianskai Yang, Yi Nian, Xiyang Hu, Yue Zhao
NLP-ADBench: NLP Anomaly Detection Benchmark
Findings of the Association for Computational Linguistics: EMNLP, 2025.
48. Lincan Li, Eren Erman Ozguven, Yue Zhao, Guang Wang, Yiqun Xie, Yushun Dong
TyphoonFormer: Language-Augmented Transformer for Accurate Typhoon Track Forecasting
ACM International Conference on Advances in Geographic Information Systems (SIGSPATIAL),  Best Short Paper Award, 2025.
47. Yi Nian*, Shenzhe Zhu*, Yuehan Qin, Shawn Li, Ziyi Wang, Chaowei Xiao, Yue Zhao
JailDAM: Jailbreak Detection with Adaptive Memory for Vision-Language Model
Conference on Language Modeling (COLM), 2025.
46. Shawn Li, Peilin Cai, Yuxiao Zhou, Zhiyu Ni, Renjie Liang, You Qin, Yi Nian, Zhengzhong Tu, Xiyang Hu, Yue Zhao
Secure On-Device Video OOD Detection Without Backpropagation
International Conference on Computer Vision (ICCV), 2025.
45. Zerui Xu, Fang Wu, Yue Zhao
Retrieval-Reasoning Large Language Model-based Synthetic Clinical Trial Generation
ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD Workshop on AI Agent for Information Retrieval), 2025.
ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM BCB, 2025.
44. Haoyan Xu*, Zhengtao Yao*, Yushun Dong, Ziyi Wang, Ryan A. Rossi, Mengyuan Li, Yue Zhao
Few-Shot Graph Out-of-Distribution Detection with LLMs
European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD), 2025.
43. Tianskai Yang*, Yi Nian*, Shawn Li, Ruiyao Xu, Yuangang Li, Jiaqi Li, Xiyang Hu, Ryan Rossi, Kaize Ding, Xia Hu, Yue Zhao
AD-LLM: Benchmarking Large Language Models for Anomaly Detection
Findings of the Association for Computational Linguistics (ACL Findings), 2025.
42. Yu Xia, Subhojoyti Mukherjee, Zhouhang Xie, Junda Wu, Xintong Li, Ryan Aponte, Hanjia Lyu, Joe Barrow, Hongjie Chen, Franck Dernoncourt, Branislav Kveton, Tong Yu, Ruiyi Zhang, Jiuxiang Gu, Nesreen K. Ahmed, Yu Wang, Xiang Chen, Hanieh Deilamsalehy, Sungchul Kim, Zhengmian Hu, Yue Zhao, Nedim Lipka, Seunghyun Yoon, Ting-Hao Kenneth Huang, Zichao Wang, Puneet Mathur, Soumyabrata Pal, Koyel Mukherjee, Zhehao Zhang, Namyong Park, Thien Huu Nguyen, Jiebo Luo, Ryan A. Rossi, Julian McAuley
From Selection to Generation: A Survey of LLM-based Active Learning
Annual Meeting of the Association for Computational Linguistics (ACL), 2025.
41. Kaixiang Zhao, Lincan Li, Kaize Ding, Neil Zhenqiang Gong, Yue Zhao, Yushun Dong
A Survey on Model Extraction Attacks and Defenses for Large Language Models
ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD Lecture-Style Tutorial Track), 2025.
40. Shawn Li, Huixian Gong, Hao Dong, Tianskai Yang, Zhengzhong Tu, Yue Zhao
DPU: Dynamic Prototype Updating for Multimodal Out-of-Distribution Detection
Conference on Computer Vision and Pattern Recognition (CVPR),  Highlight, 2025
39. Hanhui Wang, Yihua Zhang, Ruizheng Bai, Yue Zhao, Sijia Liu, Zhengzhong Tu
Edit Away and My Face Will Not Stay: Personal Biometric Defense against Malicious Generative Editing
Conference on Computer Vision and Pattern Recognition (CVPR), 2025
38. Yanbo Wang, Jiayi Ye, Siyuan Wu, Chujie Gao, Yue Huang, Xiuying Chen, Yue Zhao, Xiangliang Zhang
TRUSTEVAL: A Dynamic Evaluation Toolkit on Trustworthiness of Generative Foundation Models
Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL Demo Track), 2025.
37. Yuehan Qin*, Yichi Zhang*, Yi Nian*, Xueying Ding, Yue Zhao
MetaOOD: Meta-learning for Automatic Out-of-Distribution Detection Model Selection
International Conference on Learning Representations (ICLR), 2025
ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD Workshop on Resource-Efficient Learning for Knowledge Discovery),  Best Paper Award, 2024.

36. Sihan Chen, Zhuangzhuang Qian, Wingchun Siu, Xingcan Hu, Jiaqi Li, Shawn Li, Yuehan Qin, Tiankai Yang, Zhusuo Xiao, Wanghao Ye, Yichi Zhang, Yushun Dong, Yue Zhao
 PyOD 2: A Python Library for Outlier Detection with LLM-powered Model Selection
International World Wide Web Conference (The Web Conference Demo Track), 2025
35. Sizhe Liu, Yizhou Lu, Siyu Chen, Xiyang Hu, Yue Zhao
 DrugAgent: Automating AI-aided Drug Discovery Programming through LLM Multi-Agent Collaboration
AAAI Workshop on Foundation Models for Biological Discoveries (FMs4Bio), 2025.
34. Hao Dong, Yue Zhao, Eleni Chatzi, Olga Fink
 MultiOOD: Scaling Out-of-Distribution Detection for Multiple Modalities
Advances in Neural Information Processing Systems (NeurIPS), Spotlight, 2024
33. Xueying Ding, Yue Zhao, Leman Akoglu
 Fast Unsupervised Deep Outlier Model Selection with Hypernetworks
ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2024
32. Lichao Sun, Yue Huang, Haoran Wang, Siyuan Wu, Qihui Zhang, Chujie Gao, Yixin Huang, Wenhan Lyu, Yixuan Zhang, Xiner Li, Zhengliang Liu, Yixin Liu, Yijue Wang, Zhikun Zhang, 50+ collaborative authors, Yue Zhao
 TrustLLM: Trustworthiness in Large Language Models
International Conference on Machine Learning (ICML), 2024
31. Songtao Liu, Hanjun Dai, Yue Zhao, Peng Liu
 Preference Optimization for Molecule Synthesis with Conditional Residual Energy-based Models
International Conference on Machine Learning (ICML), Oral, 2024
30. Yue Zhao, Leman Akoglu
 Hyperparameter Optimization for Unsupervised Outlier Detection
International Conference on Automated Machine Learning (AutoML), 2024
29. Yue Zhao
 Towards Reproducible, Automated, and Scalable Anomaly Detection
AAAI Conference on Artificial Intelligence (AAAI), New Faculty Highlights, 2024
28. Minqi Jiang*, Chaochuan Hou*, Ao Zheng*, Songqiao Han, Hailiang Huang[†], Qingsong Wen, Xiyang Hu[†], Yue Zhao[†]
 ADGym: Design Choices for Deep Anomaly Detection.
Advances in Neural Information Processing Systems (NeurIPS), 2023
 ([†]Corresponding author)
27. Jaemin Yoo, Yue Zhao, Lingxiao Zhao, Leman Akoglu
 DSV: An Alignment Validation Loss for Self-supervised Outlier Model Selection
European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD), 2023
26. Peng Xu, Lin Zhang, Xuanzhou Liu, Jiaqi Sun, Yue Zhao, Haiqin Yang, Bei Yu
 Do Not Train It: A Linear Neural Architecture Search of Graph Neural Networks
International Conference on Machine Learning (ICML), 2023
25. Yue Zhao, Guoqing Zheng, Subhabrata Mukherjee, Robert McCann, Ahmed Awadallah
 ADMoE: Anomaly Detection with Mixture-of-Experts from Noisy Labels
Thirty-Seventh AAAI Conference on Artificial Intelligence (AAAI), 2023
24. Yue Zhao, George H. Chen, Zhihao Jia
 TOD: GPU-accelerated Outlier Detection via Tensor Operations
International Conference on Very Large Data Bases (VLDB), 2023
23. Songqiao Han*, Xiyang Hu*, Hailiang Huang*, Minqi Jiang*, Yue Zhao*
 ADBench: Anomaly Detection Benchmark
Advances in Neural Information Processing Systems (NeurIPS), 2022
 (*equal contribution & the corresponding author)
22. Yue Zhao*, Kay Liu*, Yingtong Dou*, et al.
 Benchmarking Node Outlier Detection on Graphs
Advances in Neural Information Processing Systems (NeurIPS), 2022
 (*equal contribution)

21. Yue Zhao, Xiaorong Zhang, Leman Akoglu
ELECT: Toward Unsupervised Outlier Model Selection
IEEE International Conference on Data Mining (ICDM), 2022.
20. Zhiming Xu, Xiao Huang, Yue Zhao, Yushun Dong, Jundong Li
Contrastive Attributed Network Anomaly Detection with Data Augmentation
Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), 2022.
19. Yue Zhao, Ryan A. Rossi, Leman Akoglu
Automatic Unsupervised Outlier Model Selection
Advances in Neural Information Processing Systems (NeurIPS), 2021.
18. Kwei-Herng Lai, Daochen Zha, Junjie Xu, Yue Zhao, Guanchu Wang, Xia Hu
Revisiting Time Series Outlier Detection: Definitions and Benchmarks
Advances in Neural Information Processing Systems (NeurIPS), 2021
17. Kexin Huang*, Tianfan Fu*, Wenhao Gao*, Yue Zhao, Yusuf Roohani, Jure Leskovec, Connor W. Coley, Cao Xiao, Jimeng Sun, Marinka Zitnik
Therapeutics Data Commons: Machine Learning Datasets and Tasks for Drug Discovery and Development
Advances in Neural Information Processing Systems (NeurIPS), 2021
(*equal contribution)
16. Yue Zhao*, Xiyang Hu*, Cheng Cheng, Cong Wang, Changlin Wan, Wen Wang, Jianing Yang, Haoping Bai, Zheng Li, Cao Xiao, Yunlong Wang, Zhi Qiao, Jimeng Sun, Leman Akoglu
SUOD: Accelerating Large-scale Unsupervised Heterogeneous Outlier Detection
Conference on Machine and Learning Systems (MLSys), 2021. (*equal contribution)
15. Kwei-Herng Lai*, Daochen Zha*, Guanchu Wang, Junjie Xu, Yue Zhao, Devesh Kumar, Yile Chen, Purav Zumkhawaka, Minyang Wan, Diego Martinez and Xia Ben Hu
TODS: An Automated Time Series Outlier Detection System (Demo paper)
Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI), 2021.
(*equal contribution)
14. Meng-Chieh Lee, Yue Zhao, Aluna Wang, Pierre Jinghong Liang, Leman Akoglu, Vincent S. Tseng, Christos Faloutsos
AutoAudit: Mining Accounting and Time-Evolving Graphs
IEEE International Conference on Big Data (Big Data), 2020
13. Changlin Wan, Dongya Jia, Yue Zhao, Wennan Chang, Sha Cao, Xiao Wang, and Chi Zhang
A Data Denoising Approach to Optimize Functional Clustering of Single Cell RNA-sequencing Data
IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2020
12. Yue Zhao, Xueying Ding, Jianing Yang, Haoping Bai.
SUOD: Toward Scalable Unsupervised Outlier Detection
Workshops at the Thirty-Fourth AAAI Conference on Artificial Intelligence, 2020.
Extended version published in MLSys 2021.
11. Zheng Li, Yue Zhao, Nicola Botta, Cezar Ionescu, Xiyang Hu
COPOD: Copula-Based Outlier Detection
IEEE International Conference on Data Mining (ICDM), 2020.
10. Zheng Li, Yue Zhao, Jialin Fu
SYNC: A Copula based Framework for Generating Synthetic Data from Aggregated Sources
IEEE International Conference on Data Mining Workshops (ICDMW), 2020.
9. Yiqun Mei, Yue Zhao, Wei Liang
DSR: An Accurate Single Image Super Resolution Approach for Various Degradations
IEEE International Conference on Multimedia and Expo (ICME), 2020, London, UK.
8. Yue Zhao, Xuejian Wang*, Cheng Cheng*, Xueying Ding*
Combining Machine Learning Models and Scores using combo Library (Demo paper)
Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI), 2020.
(*equal contribution)
7. Colin Wan, Zheng Li, Alicia Guo, Yue Zhao
SynC: A Unified Framework for Generating Synthetic Population with Gaussian Copula
Workshops at the Thirty-Fourth AAAI Conference on Artificial Intelligence, 2020.
Extended version published in ICDMW 2020.

6. Zain Nasrullah, Yue Zhao
Music Artist Classification with Convolutional Recurrent Neural Networks
IEEE International Joint Conference on Neural Networks (IJCNN), 2019, Hungary.
5. Yue Zhao, Zain Nasrullah, Maciej K. Hryniwicki, Zheng Li
LSCP: Locally Selective Combination in Parallel Outlier Ensembles
SIAM International Conference on Data Mining (SDM), 2019, Calgary, Canada.
4. Yue Zhao, Maciej K. Hryniwicki
DCSO: Dynamic Combination of Detector Scores for Outlier Ensembles
ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD Workshop on Outlier Detection De-constructed), 2018, London, UK.
Extended version published in SDM 2019, renamed to LSCP.
3. Yue Zhao, Maciej K. Hryniwicki
XGBOD: Improving Supervised Outlier Detection with Unsupervised Representation Learning
IEEE International Joint Conference on Neural Networks (IJCNN), 2018, Rio, Brazil.
2. Yue Zhao, Maciej K. Hryniwicki, Francesca Cheng, Boyang Fu, Xiaoyu Zhu
Employee Turnover Prediction with Machine Learning: A Reliable Approach
Intelligent System Conference (Intellisys), 2018, London, UK.
1. Yue Zhao*, Zhongtian Qiu*, Yiqing Yang*, Weiwei Li*, Mingming Fan
An Empirical Study of Touch-based Authentication Methods on Smartwatches
ACM International Symposium on Wearable Computers (ISWC), 2017, Maui, USA. (*equal contribution)

INTERNSHIP EXPERIENCE	NortonLifeLock Research Group Machine Learning Research Intern	2022
	Microsoft Research Machine Learning Research Intern	2022
	Stanford University, Computer Science Department Visiting Student Researcher (Prof. Jure Leskovec)	2021
	IQVIA, Analytics Center of Excellence Machine Learning Research Intern	2020
	Siemens PLM Software USA Software Engineer (Intern & Contract)	Mar. 2012 - Dec. 2014
TEACHING EXPERIENCE	<p>University of Southern California Instructor <i>CSCI 566 Deep Learning and Its Applications</i></p> <p>Instructor <i>CSCI 699 Adversarial and Trustworthy Foundation Models</i></p> <p>Instructor <i>CSCI 566 Deep Learning and Its Applications</i></p> <p>Instructor <i>CSCI 566 Deep Learning and Its Applications</i></p> <p>Carnegie Mellon University Teaching Assistant <i>Managing Digital Business</i> (Prof. David Riel)</p> <p>Teaching Assistant & co-Instructor (lectures on AutoML and MLSys) <i>Intro to Artificial Intelligence</i> (Prof. David Steier)</p> <p>Teaching Assistant <i>Digital Transformation</i> (Prof. David Riel)</p> <p>Teaching Assistant (helping on course topics) <i>Statistics for IT Managers</i> (Prof. Daniel Nagin)</p>	Los Angeles, CA Fall 2026 (scheduled) Spring 2026 Spring 2025 Spring 2024 Pittsburgh, PA Fall 2022 Spring 2022 – Fall 2020 Spring 2022 Fall 2021

University of Toronto
Teaching Assistant & Lab Session Instructor
Embedded Systems (Prof. Philip Anderson)

Toronto, ON
Fall 2015

University of Cincinnati
Teaching Assistant & Lab Session Instructor
Intro to Programming (Prof. George Purdy)

Cincinnati, OH
Fall 2014

Ph.D. STUDENTS

- Haoyan Xu (USC, ECE Ph.D., 2024 Spring-), co-advised by Mengyuan Li,  Capital One Fellowship
- Yuehan Qin (USC, CS Ph.D., 2024 Fall-)
- Tiankai Yang (USC, CS Ph.D., 2024 Fall-)
- Shawn Li (USC, CS Ph.D., 2024 Fall-),  Capital One Fellowship, Amazon ML Fellowship
- Jiate Li (USC, CS Ph.D., 2025 Fall-)

SERVICES

Conference/Workshop Organizing Committee

- Workflow Co-Chair for KDD 2023
- Co-organizer, AI for Financial Fraud Detection & Prevention Workshop @ 6th ACM International Conference on AI in Finance
- Co-organizer, SURGeLLM: Structured Understanding, Retrieval, and Generation in LLMs era Workshop @ ACL 2026

External Reviewer for Funding Proposals

- Dutch Research Council (NWO)

Journal Editor

- Associate Editor, ACM Transactions on AI for Science (TAIS), 2025–present
- Associate Editor, IEEE Transactions on Neural Networks and Learning Systems (TNNLS), 2024–present
- Action Editor, Journal of Data-centric Machine Learning Research (DMLR), 2024–present

Program Committee (PC) or Area Chair (AC) for Conferences and Workshops

- ICLR 2025 (AC), ICLR 2026 (AC)
- AAAI 2021, 2022, 2023, 2025 (Senior PC), 2026 (Senior PC)
- ICML 2024, 2025 (AC), 2026 (AC)
- NeurIPS 2021, 2022, 2023, 2025 (AC)
- AISTATS 2024, 2025 (AC)
- MLSys 2024, 2026
- KDD 2020, 2021, 2022, 2023
- IJCAI 2022, 2023
- AAAI Demonstrations 2021, 2022
- MICCAI 2020, 2021, 2022
- ICDM 2020
- KDD Workshop on Outlier Detection and Description (ODD), 2021
- KDD Workshop on Anomaly and Novelty Detection (ANDEA), 2021, 2022
- IJCAI Workshop on Artificial Intelligence for Anomalies and Novelties (AI4AN), 2020, 2021
- INFORMS Workshop on Data Science 2021

Journal Reviewer

- Journal of Machine Learning Research (JMLR)
- PNAS Nexus
- Machine Learning
- IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)
- IEEE Transactions on Knowledge and Data Engineering (TKDE)
- IEEE Internet of Things Journal (IoT-J)
- IEEE Intelligent Systems
- IEEE Journal on Selected Areas in Communications (J-SAC)
- Data Mining and Knowledge Discovery (DMAI)
- ACM Transactions on Management Information Systems (TMIS)
- Knowledge and Information Systems (KAIS)
- INFORMS Journal on Computing (IJOC)
- Big Data
- Artificial Intelligence Review (AIRE)
- Neurocomputing
- IEEE Transactions on Systems, Man, and Cybernetics: Systems
- IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)
- IEEE Network Magazine
- IEEE Computational Intelligence Magazine (CIM)
- BioData Mining
- European Journal of Management and Business Economics (EJM&BE)
- The Journal of Open Source Software (JOSS)

TALKS AND LECTURES	USC symposium on Frontiers of ML/AI NUS Tea Talk SFU@NeurIPS'24 KAIST	<i>Towards Robust AI: Advances in Outlier and OOD Detection</i> <i>Towards Robust AI: Advances in Outlier and OOD Detection</i> <i>Towards Robust AI: Advances in Outlier and OOD Detection</i> <i>Unsupervised Model Selection: Automation with Meta-learning and LLMs</i>	Mar. 2025 Jan. 2025 Dec. 2024 Nov. 2024
	Kennesaw State University	<i>Unsupervised Model Selection: Automation with Meta-learning and LLMs</i>	Oct. 2024
	LinkedIn Anti-Abuse AI Amazon Security AI New York University University of Washington Microsoft USC Retreat on AI and Engineering Safety Visa Research	<i>Outlier Detection: Automation, Systems, and GenAI</i> <i>Outlier Detection: Automation, Systems, and GenAI</i> <i>Safety Measures for LLMs</i>	Aug. 2024 Aug. 2024 Aug. 2024 Jun. 2024 Jun. 2024 Apr. 2024
	USC Symposium on Frontiers of Generative AI	<i>Towards Reproducible, Automated, and Scalable AD</i> <i>Generative AI for Anomaly Detection</i>	Apr. 2024 Mar. 2024
	AAAI New Faculty Highlights (invited)	<i>Towards Reproducible, Automated, and Scalable AD</i>	Feb. 2024
	U of Nevada, Las Vegas Samsung Seminar KDD SoCal Day CMU Catalyst KAUST Emory University USC UC Davis Stony Brook University University of Chicago UC Merced CMU PDL Meeting CMU Data Science Seminar LoG Seminar Intuit Rice University Microsoft Research Wells Fargo Columbia University Morgan Stanley Microsoft Research Morgan Stanley Rutgers Business School Tesla Catalyst, CMU E&Y Canada University of Nottingham	<i>Automated and Scalable ML Algorithms and Systems</i> <i>Automated and Scalable Anomaly Detection Systems</i> <i>Enable Applications by ML with Noisy Inputs</i> <i>How (Not) to Fail Your Academic Job Search</i> <i>Automated and Scalable ML Algorithms and Systems</i> <i>Automated and Scalable ML Algorithms and Systems</i> <i>Guest Lecture Automated Anomaly Detection</i> <i>Large-scale Graph Anomaly Detection</i> <i>Anomaly Detection for Financial Risk Modeling</i> <i>Large-scale Anomaly Detection with Automation</i> <i>Weakly-supervised Anomaly Detection</i> <i>Anomaly Detection for Financial Risk Modeling</i> <i>Guest Lecture Anomaly Detection</i> <i>Automated Outlier Detection</i> <i>Automated Outlier Detection</i> <i>Large-scale Anomaly Detection Systems</i> <i>Outlier Model Selection</i> <i>Large-scale Anomaly Detection Systems</i> <i>Systems for Data Mining Algorithms</i> <i>ML applications in Data Analytics</i> <i>General Machine Learning Applications</i>	Oct. 2023 Aug. 2023 Aug. 2023 May. 2023 Apr. 2023 Apr. 2023 Mar. 2023 Mar. 2023 Feb. 2023 Feb. 2023 Feb. 2023 Feb. 2023 Jan. 2023 Nov. 2022 Oct. 2022 Aug. 2022 Sep. 2022 Sep. 2022 Aug. 2022 Jul. 2022 Jun. 2022 Jun. 2022 Mar. 2022 Mar. 2022 Feb. 2022 Dec. 2021 Oct. 2021 Jan. 2021