

# Yue Zhao

---

## CONTACT INFORMATION

✉ [yue.z@usc.edu](mailto:yue.z@usc.edu)  
🐙 [github.com/yzhao062](https://github.com/yzhao062)  
in [linkedin.com/in/yzhao062](https://www.linkedin.com/in/yzhao062)  
🏠 <https://viterbi-web.usc.edu/~yzhao010/>  
👤 USC Faculty Directory  
🔍 Google Scholar  
★ 22,000+ GitHub Stars

213-821-2369  
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.

- ❑ LLM Safety & Reliability
- ❑ Hallucination Mitigation
- ❑ Jailbreak Detection
- ❑ Multi-agent Reliability
- ❑ Privacy & Model Extraction Risks
- ❑ Robust Reasoning & Routing
- ❑ Failure Mode Analysis

#### 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.

- ❑ Robustness
- ❑ OOD Generalization
- ❑ Anomaly Detection
- ❑ Model Selection
- ❑ Benchmarks & Evaluation
- ❑ Multimodal & Graph Settings
- ❑ Data Efficiency

### 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.

- ❑ System-Level Evaluation
- ❑ Benchmarking & Stress Testing
- ❑ Scalable Evaluation Pipelines
- ❑ Reproducibility & Measurement
- ❑ Agentic Workflows
- ❑ 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.

- ❑ AI for Science
- ❑ Scientific Foundation Models
- ❑ Climate & Weather Modeling
- ❑ Healthcare & Biomedicine
- ❑ Computational Social Systems
- ❑ Decision Modeling

FULL-TIME	<b>University of Southern California</b>	
PROFESSIONAL	<i>Thomas Lord Department of Computer Science</i>	
EXPERIENCE	Assistant Professor (Tenure-Track)	Aug. 2023 - Present
	<ul style="list-style-type: none"> <li>• <u>F</u>oundations <u>O</u>f <u>R</u>obust <u>T</u>rustworthy <u>I</u>ntelligent <u>S</u>ystems (<b>FORTIS</b>) Lab: <a href="#">Link</a></li> <li>• USC Machine Learning Center (MaSCle): <a href="#">Link</a></li> </ul>	

### Stealth Startup

<i>AI Research &amp; Systems</i>	
External Technical Advisor (Part-Time, Contract-based)	Mar. 2024 – Present

### PwC Canada

<i>Consulting &amp; Deals</i>	
Senior Consultant (Data Scientist)	Aug. 2017 - Jun. 2019
Consultant (Data Scientist)	Feb. 2017 - Jul. 2017

EDUCATION	<b>Carnegie Mellon University</b>	Pittsburgh, PA
	<i>Ph.D. in Information Systems and Management</i>	Sep. 2019 - May. 2023
	<ul style="list-style-type: none"> <li>• <b>Affiliation:</b> CMU automated learning systems group (Catalyst) and Data Analytics Techniques Algorithms (DATA) Lab</li> <li>• <b>Advisors and Mentors:</b> 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.</li> </ul>	

### University of Toronto

<i>Master of Science in Computer Science</i>	Toronto, ON
	Sep. 2015 - Dec. 2016

### University of Cincinnati

<i>Bachelor of Science in Computer Engineering</i>	Cincinnati, OH
<b>Minor:</b> <i>Computer Science and Mathematics</i>	Sep. 2010 - May. 2015

### AWARDS, GRANTS, AND FUNDING

#### As Principal Investigator (August 2023 onwards)

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.*

Note: \*first authors and <sup>†</sup>corresponding authors if more than one.

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**  
**SSRN 5917963**
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. 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  
**Under submission**  
**arXiv preprint arXiv:2512.00752**
87. 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**
86. 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**
85. Haoyan Xu, Ruizhi Qian, Jiate Li, Yushun Dong, Minghao Lin, Hanson Yan, Zhengtao Yao, Qinghua Liu, Junhao Dong, Ruopeng Huang, Yue Zhao<sup>†</sup>, Mengyuan Li<sup>†</sup>  
 A Systematic Study of Model Extraction Attacks on Graph Foundation Models  
**Under submission**  
**arXiv preprint arXiv:2511.11912**
84. 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**
83. 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**
82. 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**
81. Yuehan Qin, Li Li, Defu Cao, Tiankai Yang, Yue Zhao  
 M3OOD: Automatic Selection of Multimodal OOD Detectors  
**Under submission**  
**arXiv preprint arXiv:2508.11936**

80. 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**
79. 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  
**Under submission**  
**arXiv preprint arXiv:2505.23713**
78. 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**
77. 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**
76. 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  
**Under submission**  
**arXiv preprint arXiv:2504.19373**
75. Yuehan Qin, Shawn Li, Yi Nian, Xinyan Velocity Yu, Yue Zhao<sup>†</sup>, Xuezhe Ma<sup>†</sup>  
Don't Let It Hallucinate: Premise Verification via Retrieval-Augmented Logical Reasoning  
**Under submission**  
**arXiv preprint arXiv:2504.06438**
74. 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**
73. Chengxuan Qian, Shuo Xing, Shawn Li, Yue Zhao, Zhengzhong Tu  
DecAlign: Hierarchical Cross-Modal Alignment for Decoupled Multimodal Representation Learning  
**Under submission**  
**arXiv preprint arXiv:2503.11892**
72. Xiong Xiao Xu, Haoran Wang, Yueqing Liang, Philip S. Yu, Yue Zhao, Kai Shu  
Can Multimodal LLMs Perform Time Series Anomaly Detection?  
**Under submission**  
**arXiv preprint arXiv:2502.17812**
71. 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**
70. 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  
**Under submission**  
**arXiv preprint arXiv:2502.14296**  
<https://trustgen.github.io/>
69. 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**

68. Lincan Li, Jiaqi Li, Catherine Chen<sup>†</sup>, Fred Gui<sup>†</sup>, other collaborators, [Yue Zhao<sup>†</sup>](#), Yushun Dong<sup>†</sup>  
Political-LLM: Large Language Models in Political Science  
**Under submission**  
**arXiv preprint arXiv:2412.06864**
67. Chenxiao Yu, Jinyi Ye, Yuangang Li, Zhaotian Weng, Zheng Li, Emilio Ferrara, Xiyang Hu<sup>†</sup>, [Yue Zhao<sup>†</sup>](#)  
A Large-Scale Simulation on Large Language Models for Decision-Making in Political Science  
**Under submission**  
**arXiv preprint arXiv:2412.15291**
66. 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**
65. 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

64. Haoyan Xu, Kay Liu, Zhengtao Yao, Philip S. Yu, Kaize Ding<sup>†</sup>, [Yue Zhao<sup>†</sup>](#)  
LEGO-Learn: Label-Efficient Graph Open-Set Learning  
*Transactions on Machine Learning Research (TMLR)*, 2025
63. 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
62. 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)
61. [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)
60. 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)
59. [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)
58. [Yue Zhao](#), Zain Nasrullah, Zheng Li  
PyOD: A Python Toolbox for Scalable Outlier Detection  
*Journal of Machine Learning Research (JMLR)*, 2019.

#### Conference & Workshop Papers

57. Bo Ni, Yu Wang, Leyao Wang, Branislav Kveton, Franck Dernoncourt, Yu Xia, Hongjie Chen, Reuben Luera, Samyadeep Basu, Subhojyoti 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, Namyoung 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
56. 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.
55. 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
54. Tiankai 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.
53. Ruosi Shao, Md Shamim Seraj, Kangyi Zhao, Yingtao Luo, Lincan Li, Bolin Shen, Averil 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.
52. Li Li, Peilin Cai, Ryan A. Rossi, Franck Dernoncourt, Branislav Kveton, Junda Wu, Tong Yu, Lixin Song, Tiankai Yang, Yuehan Qin, Nesreen K. Ahmed, Samyadeep Basu, Subhojyoti 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
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, Tiankai 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, Tiankai 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  
TyphoFormer: 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. Tiankai 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, Subhojyoti 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, Tiankai 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, Zhuo 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, Tianfan Fu, [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<sup>†</sup>, Qingsong Wen, Xiyang Hu<sup>†</sup>, [Yue Zhao](#)<sup>†</sup>  
ADGym: Design Choices for Deep Anomaly Detection.  
*Advances in Neural Information Processing Systems (NeurIPS)*, 2023  
(<sup>†</sup>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\*, Yingdong 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 Zunkhawaka, 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. Hryniewicki, 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. Hryniewicki  
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. Hryniewicki  
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. Hryniewicki, 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	<b>NortonLifeLock Research Group</b>	
	Machine Learning Research Intern	2022
	<b>Microsoft Research</b>	
	Machine Learning Research Intern	2022
	<b>Stanford University, Computer Science Department</b>	
	Visiting Student Researcher (Prof. Jure Leskovec)	2021
	<b>IQVIA, Analytics Center of Excellence</b>	
	Machine Learning Research Intern	2020
	<b>Siemens PLM Software USA</b>	
	Software Engineer (Intern & Contract)	Mar. 2012 - Dec. 2014

TEACHING EXPERIENCE	<b>University of Southern California</b>	Los Angeles, CA
	<b>Instructor</b>	Fall 2026 (scheduled)
	<i>CSCI 566 Deep Learning and Its Applications</i>	
	<b>Instructor</b>	Spring 2026
	<i>CSCI 699 Adversarial and Trustworthy Foundation Models</i>	
	<b>Instructor</b>	Spring 2025
	<i>CSCI 566 Deep Learning and Its Applications</i>	
	<b>Instructor</b>	Spring 2024
	<i>CSCI 566 Deep Learning and Its Applications</i>	
	<b>Carnegie Mellon University</b>	Pittsburgh, PA
	<b>Teaching Assistant</b>	Fall 2022
	<i>Managing Digital Business</i> (Prof. David Riel)	
	<b>Teaching Assistant &amp; co-Instructor</b> (lectures on AutoML and MLSys)	Spring 2022 – Fall 2020
	<i>Intro to Artificial Intelligence</i> (Prof. David Steier)	
	<b>Teaching Assistant</b>	Spring 2022
	<i>Digital Transformation</i> (Prof. David Riel)	
	<b>Teaching Assistant</b> (helping on course topics)	Fall 2021
	<i>Statistics for IT Managers</i> (Prof. Daniel Nagin)	
	<b>University of Toronto</b>	Toronto, ON
	<b>Teaching Assistant &amp; Lab Session Instructor</b>	Fall 2015
	<i>Embedded Systems</i> (Prof. Philip Anderson)	
	<b>University of Cincinnati</b>	Cincinnati, OH
	<b>Teaching Assistant &amp; Lab Session Instructor</b>	Fall 2014
	<i>Intro to Programming</i> (Prof. George Purdy)	

#### 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 Novelty (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 (EJMBE)
- The Journal of Open Source Software (JOSS)

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