GUANGTAO ZHENG

Charlottesville, Virginia, 22903, United States of America

zhguangt@gmail.com ◆ (+1) 518 961 2159 ◆ Homepage ◆ Google Scholar ◆ LinkedIn

SUMMARY

My research is on Artificial Intelligence and Machine Learning. My main research theme is on AI alignment and safety from the perspective of detecting and mitigating spurious correlations, with topics covering: shortcut learning in image and text classification, superficial learning in multimodal models, reward hacking in reinforcement learning, benchmarking superficial learning in AI models, and robust reasoning and planning in agentic systems.

EDUCATION

University of Virginia **Charlottesville, United States**
Doctor of Philosophy in Computer Science (4.0 / 4.0 GPA) Aug 2019 – Jun 2025
- Advisor: Aidong Zhang

University of Science and Technology of China **Hefei, China**
Master of Engineering in Electrical Engineering Sep 2015 – Jun 2018
- Advisor: Chen Gong

Sun Yat-Sen University **Guangzhou, China**
Bachelor of Science in Electrical Engineering Sep 2011 – Jun 2015
- Advisor: Ming Jiang

WORK EXPERIENCE

Center for Advanced AI @ Accenture **Mountain View, California**
Advanced AI Research Scientist Manager Jun 2025 – Present
- Design methods to improve task understanding, reasoning, and planning capabilities of LLMs
- Develop for automatic agent generation to enhance user experience

RESEARCH EXPERIENCE

Mitigating Learning Superficial Patterns in Machine Learning
University of Virginia Jan 2021 – Jun 2025
- Collaborated on the development of algorithms to mitigate reward hacking in reinforcement learning and reduce spurious biases in multimodal models
- Developed algorithms to detect and mitigate the tendency of learning superficial patterns in machine learning models
- Improved the robustness of machine learning models against distribution shifts and improved the generalization to out-of-distribution data and data with subpopulation shifts

Building Foundation Models for Genomics
University of Virginia Jul 2024 – Jul 2025
- Develop tokenization methods for genomic interval data and design model architectures for efficient training
- Design training algorithms to build mappings between genomic data and natural languages in LLMs

Improving Multimodal Understanding of Vision-Language Models
University of Virginia Oct 2024 – May 2025
- Created a benchmark dataset to systematically evaluate multimodal understanding of large vision-language models
- Developed prompt selection methods to improve alignment between vision and language modalities

RECENT PUBLICATIONS

[NeurIPS'25] Wenqian Ye, **Guangtao Zheng**, and Aidong Zhang, Rectifying Shortcut Behaviors in Preference-based Reward Learning, *Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2025

[ICML'25] **Guangtao Zheng**, Wenqian Ye, and Aidong Zhang, NeuronTune: Towards Self-Guided Spurious Bias Mitigation, *The 42nd International Conference on Machine Learning (ICML)*, 2025

[IJCAI'25] **Guangtao Zheng**, Wenqian Ye, and Aidong Zhang, ShortcutProbe: Probing Prediction Shortcuts for Learning Robust Models, *The 34th International Joint Conference on Artificial Intelligence (IJCAI)*, 2025

[KDD'25] Wenqian Ye, **Guangtao Zheng**, and Aidong Zhang, Improving Group Robustness on Spurious Correlation via Evidential Alignment, *The 31th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2025 [**Best Paper Award**]

[ECCV'24] **Guangtao Zheng**, Wenqian Ye, and Aidong Zhang, Benchmarking Spurious Bias in Few-Shot Image Classifiers, *The 18th European Conference on Computer Vision (ECCV)*, 2024

[KDD'24] **Guangtao Zheng**, Wenqian Ye, and Aidong Zhang, Spuriousness-Aware Meta-Learning for Learning Robust Classifiers, *The 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2024

[IJCAI'24] **Guangtao Zheng**, Wenqian Ye, and Aidong Zhang, Learning Robust Classifiers with Self-Guided Spurious Correlation Mitigation, *The 33rd International Joint Conference on Artificial Intelligence (IJCAI)*, 2024

[NARGAB'24] **Guangtao Zheng**, Julia Rymuza, Erfaneh Gharavi, Nathan J LeRoy, Aidong Zhang, and Nathan C Sheffield, Methods for Evaluating Unsupervised Vector Representations of Genomic Regions, *Nucleic Acids Research Genomics and Bioinformatics*, 2024

[AAAI'24] **Guangtao Zheng**, Mengdi Huai, and Aidong Zhang, AdvST: Revisiting Data Augmentations for Single Domain Generalization, *The 38th Annual AAAI Conference on Artificial Intelligence (AAAI)*, 2024

[ICMLW'24] Wenqian Ye, **Guangtao Zheng**, Xu Cao, Yunsheng Ma, and Aidong Zhang, Spurious Correlations in Machine Learning: A Survey, *ICML Workshop on Data-Centric Machine Learning Research*, 2024

[WRBFM'24] Wenqian Ye, **Guangtao Zheng**, Yunsheng Ma, Xu Cao, Bolin Lai, James M. Rehg, and Aidong Zhang, MM-SpuBench: Towards Better Understanding of Spurious Biases in Multimodal LLMs, *NeurIPS Workshop on Responsibly Building the Next Generation of Multimodal Foundational Models*, 2024

[NARGAB'24] Nathan J LeRoy, Jason P Smith, **Guangtao Zheng**, Julia Rymuza, Erfaneh Gharavi, Donald E Brown, Aidong Zhang, and Nathan C Sheffield, Fast Clustering and Cell-Type Annotation of scATAC Data Using Pre-trained Embeddings, *Nucleic Acids Research Genomics and Bioinformatics*, 2024

[BioEng'24] Erfaneh Gharavi, Nathan J LeRoy, **Guangtao Zheng**, Aidong Zhang, Donald E Brown, and Nathan C Sheffield, Joint Representation Learning for Retrieval and Annotation of Genomic Interval Sets, *Bioengineering*, 2024

[NAR'24] Julia Rymuza, Yuchen Sun, **Guangtao Zheng**, Nathan J LeRoy, Maria Murach, Neil Phan, Aidong Zhang, and Nathan C Sheffield, Methods for Constructing and Evaluating Consensus Genomic Interval Sets, *Nucleic Acids Research*, 2024

[SDM'23] **Guangtao Zheng**, Qiuling Suo, Mengdi Huai, and Aidong Zhang, Learning to Learn Task Transformations for Improved Few-Shot Classification, *SIAM International Conference on Data Mining (SDM)*, 2023

[ICDM'22] **Guangtao Zheng**, and Aidong Zhang, Knowledge-Guided Semantics Adjustment for Improved Few-Shot Classification, *IEEE International Conference on Data Mining (ICDM)*, 2022

[ICDMW'21] **Guangtao Zheng**, and Aidong Zhang, Few-Shot Class-Incremental Learning with Meta-Learned Class Structures, *IEEE International Conference on Data Mining (ICDM) Workshop*, 2021

[Bioinfo'21] Erfaneh Gharavi, Aaron Gu, **Guangtao Zheng**, Jason P Smith, Hyun Jae Cho, Aidong Zhang, Donald E Brown, and Nathan C Sheffield, Embeddings of Genomic Region Sets Capture Rich Biological Associations in Lower Dimensions, *Bioinformatics*, 2021

[ACL'20] Hanjie Chen, **Guangtao Zheng**, and Yangfeng Ji, Generating Hierarchical Explanations on Text Classification via Feature Interaction Detection, *In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2020

INVITED TALKS

- Microsoft Applied Research Talk Series, Mitigating Spurious Bias: Building AI Models with Generalizable Knowledge, March 10, 2025
- Standford @ Aghaeepour Laboratory, job talk on my research, including mitigating superficial learning and machine learning applications on Genomics, January 30, 2025
- Oracle Labs East, From Surface Learners to Deep Thinkers: Building AI Models with Generalizable Knowledge, December 18, 2024

SKILLS AND LANGUAGES

- **Languages**: English (Proficient), Chinese (Native)
- **Programming**: Python, MATLAB, C/C++, LaTeX, HTML, CSS, Typst
- **Packages**: PyTorch, Tensorflow, scikit-learn, Gensim, Numpy, Jupyter Notebook, matplotlib, SciPy, pandas
- **Miscellaneous**: GitHub, Hugging Face, AWS, Linux command line

HONORS AND AWARDS

KDD 2025 Best Paper Award (Research Track) — **Toronto, Canada**
Issued by *ACM KDD* — Aug 2025

AAAI 2024 Scholarship and Volunteer — **Vancouver, Canada**
Issued by *AAAI Conference on Artificial Intelligence* — Feb 2024

SDM 2023 Travel Award — **Minneapolis, United States**
Issued by *SIAM International Conference on Data Mining* — Apr 2023

ICDM 2022 Travel Award — **Orlando, United States**
Issued by *IEEE International Conference on Data Mining* — Dec 2022

Computer Science Fellowship — **Charlottesville, United States**
Issued by *University of Virginia* — Aug 2019

OTHERS

- **Reviewer**: IEEE BigData (2020), ACM BCB (2020), ICDM (2020, 2023), AAAI (2021, 2022), KDD (2024,2025), COLM (2024,2025), NeurIPS (2024,2025), ICLR (2025), AISTATS (2025), ICML (2025)
- **Teaching assistant**: Foundations of Data Analysis (CS4964, Spring 2022, University of Virginia), Cloud Computing (CS4740, Spring 2021, University of Virginia), Operating System (CS4414, Fall 2020, University of Virginia)