

# GUANGTAO ZHENG

Charlottesville, Virginia, 22903, United States of America

gz5hp@virginia.edu ♦ (+1) 518 961 2159 ♦ gtzheng.github.io ♦ linkedin.com/in/guangtao-zheng ♦ github.com/gtzheng

## SUMMARY

---

I specialize in developing robust models resilient to spurious patterns, with expertise in Computer Vision, NLP, and Bioinformatics. My work focuses on optimizing LLMs and Generative AI, including prompt engineering and model customization.

## EDUCATION

---

### University of Virginia

*Doctor of Philosophy in Computer Science (4.0 / 4.0 GPA),*

- Advisor: Aidong Zhang

Charlottesville, United States

Aug 2019 – Present

### University of Science and Technology of China

*Master of Engineering in Electrical Engineering*

- Advisor: Chen Gong

Hefei, China

Sep 2015 – Jun 2018

### Sun Yat-Sen University

*Bachelor of Science in Electrical Engineering*

- Advisor: Ming Jiang

Guangzhou, China

Sep 2011 – Jun 2015

## WORK EXPERIENCE

---

### University of Virginia

*Research Assistant*

- Lead research on enhancing robustness in machine learning models, with a particular focus on visual and language domains
- Publish papers in leading AI and data mining conferences; develop and maintain open-source solutions
- Collaborate with bioengineering researchers to develop innovative machine learning models utilizing genomic interval data

Charlottesville, Virginia

Aug 2019 – Present

## SELECTED PROJECTS

---

### Large Language Model Optimization for Genomics

*University of Virginia*

Jul 2024 – Present

- Developed and fine-tuned state-of-the-art LLMs for answering questions in genomics
- Implemented advanced techniques like Retrieval-Augmented Generation to enhance contextual understanding of the models
- Employed data and model parallelism techniques for efficient training of large-scale models
- Leveraged PyTorch Distributed Data Parallel (DDP) on a multi-node GPU cluster for optimal resource utilization

### Benchmarking Spurious Biases in Multimodal LLMs

*University of Virginia*

Apr 2024 – Present

- Identify and formulate spurious biases in multimodal LLMs
- Design prompts to generate vision-question answers for benchmarking multimodal LLMs

### Mitigating Spurious Biases in Deep Image Classifiers

*University of Virginia*

Jun 2023 – Present

- Created an automatic spurious bias detection method using vision-language models
- Enhanced model robustness through meta-learning and balanced training; published findings at KDD and IJCAI

## RECENT PUBLICATIONS

---

- [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

## 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
- **Deep Neural Networks:** ResNet, ViT, LLMs (e.g. BERT/Llama/GPTs), VLMs (e.g., CLIP, BLIP)
- **Miscellaneous:** GitHub, Hugging Face, AWS, Linux command line

## HONORS AND AWARDS

---

<b>AAAI 2024 Scholarship and Volunteer</b> Issued by <i>AAAI Conference on Artificial Intelligence</i>	<b>Vancouver, Canada</b> Feb 2024
<b>SDM 2023 Travel Award</b> Issued by <i>SIAM International Conference on Data Mining</i>	<b>Minneapolis, United States</b> Apr 2023
<b>ICDM 2022 Travel Award</b> Issued by <i>IEEE International Conference on Data Mining</i>	<b>Orlando, United States</b> Dec 2022
<b>Computer Science Fellowship</b> Issued by <i>University of Virginia</i>	<b>Charlottesville, United States</b> Aug 2019

## OTHERS

---

- **Reviewer:** IEEE BigData (2020), ACM BCB (2020), ICDM (2020, 2023), AAAI (2021, 2022), KDD (2024,2025), COML (2024), NeurIPS (2024), ICLR (2025), AISTATS (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)