ZIWEI ZHU

https://zziwei.github.io/
zzhu20 at gmu dot edu

George Mason University
Department of Computer Science
Fairfax, VA 22030

EDUCATION

Ph.D. Texas A&M University, USA, Computer Science 2016.08 – 2022.08
Advisor: James Caverlee

B.S. Wuhan University, China, Computer Science 2012.09 – 2016.06

RESEARCH INTERESTS

I am broadly interested in **data mining**, **machine learning**, and **natural language processing**, with a specific focus on enhancing fairness, interpretability, safety, and robustness to achieve **responsible AI**. Core topics include:

- Social bias in large (vision) language models
- Interpretable machine learning
- Robust machine learning against spurious features
- Fairness and bias in recommendation and learning-to-rank systems

WORK EXPERIENCE

Tenure-Track Assistant Professor, George Mason University, VA 2022.08 – Now
Research Intern, Netflix, CA 2020.05 – 2020.08
Research Intern, Comcast Applied AI Lab, DC 2019.05 – 2019.08

STUDENTS

PhD Students

Chahat Raj (co-advised by Dr. Antonios Anastasopoulos) Fall 2022 -
Anjishnu Mukherjee (co-advised by Dr. Antonios Anastasopoulos) Fall 2022 -
Yuqing Zhou Fall 2023 -
Bowen Wei Fall 2023 -
Mehrdad Fazli Fall 2023 -
Jinhao Pan Fall 2024 -
Fardin Ahsan Sakib (co-advised by Dr. Özlem Uzuner) Fall 2021 -

MS, Undergrad, and High School Students

Mamnuya Rinki, MS
Balassubramanian Srinivasan, MS
Diwita Banerjee, MS
Rishi Pania, Undergrad

Tanvi Pedireddi, Thomas Jefferson High School
Amrit Singh, Thomas Jefferson High School
Mahika Banerjee, Thomas Jefferson High School

TEACHING EXPERIENCE

CS 484, Data Mining, George Mason University	2025 Spring
CS 584, Data Mining, George Mason University	2024 Fall
CS 584, Data Mining, George Mason University	2024 Spring
CS 782, Advanced Machine Learning, George Mason University	2023 Fall
CS 484, Data Mining, George Mason University	2023 Spring
CS 584, Data Mining, George Mason University	2022 Fall

PUBLICATIONS (**BOLD** INDICATES STUDENT UNDER MY SUPERVISION)

Refereed Publications

WACV 2025 **Anjishnu Mukherjee**, Ziwei Zhu, and Antonios Anastasopoulos. Crossroads of Continents: Automated Artifact Extraction for Cultural Adaptation with Large Multimodal Models. IEEE/CVF Winter Conference on Applications of Computer Vision, 2025.

WSDM 2025 **Jinhao Pan**, James Caverlee, and Ziwei Zhu. Combating Heterogeneous Model Biases in Recommendations via Boosting. The 18th ACM International Conference on Web Search and Data Mining, 2025.

BigData 2024 Rahul Pandey, Ziwei Zhu, and Hemant Purohit. ORIS: Online Active Learning Using Reinforcement Learning-based Inclusive Sampling for Robust Streaming Analytics System. 2024 IEEE International Conference on Big Data.

EMNLP 2024 Findings **Yuqing Zhou**, Ruixiang Tang, Ziyu Yao, Ziwei Zhu. Navigating the Shortcut Maze: A Comprehensive Analysis of Shortcut Learning in Text Classification by Language Models. The 2024 Conference on Empirical Methods in Natural Language Processing.

EMNLP 2024 Findings **Chahat Raj**, **Anjishnu Mukherjee**, Aylin Caliskan, Antonios Anastasopoulos, and Ziwei Zhu. BiasDora: Exploring Hidden Biased Associations in Vision-Language Models. The 2024 Conference on Empirical Methods in Natural Language Processing.

AIES 2024 **Chahat Raj**, **Anjishnu Mukherjee**, Aylin Caliskan, Antonios Anastasopoulos, and Ziwei Zhu. Breaking Bias, Building Bridges: Evaluation and Mitigation of Social Biases in LLMs via Contact Hypothesis. AAAI/ACM conference on AI, Ethics, and Society, 2024.

IJCNN 2024 Ajay Vajjala, Arun Vajjala, Ziwei Zhu, and David Rosenblum. Analyzing the Impact of Domain Similarity: A New Perspective in Cross-Domain Recommendation. The IEEE International Joint Conference on Neural Networks, 2024.

NAACL 2024 **Anjishnu Mukherjee**, Aylin Caliskan, Ziwei Zhu, Antonios Anastasopoulos. Global Gallery: The Fine Art of Painting Culture Portraits through Multilingual Instruction Tuning. The North American Chapter of the Association for Computational Linguistics, 2024.

WWW 2024 Zheyuan Liu, Guangyao Dou, Eli Chien, Chunhui Zhang, Yijun Tian, Ziwei Zhu. Breaking the Trilemma of Privacy, Utility, Efficiency via Controllable Machine Unlearning. The 2024 ACM Web Conference.

ECIR IR4Good 2024 **Chahat Raj**, Anjishnu Mukherjee, Hemant Purohit, Antonios Anastasopoulos, Ziwei Zhu. SALSA: Salience-Based Switching Attack for Adversarial Perturbations in Fake News Detection Models. 46th European Conference on Information Retrieval, 2024.

ECIR IR4Good 2024 **Jinhao Pan**, Ziwei Zhu, Jianling Wang, Allen Lin, James Caverlee. End-to-End Adaptive Local Learning for Alleviating Mainstream Bias in Collaborative Filtering. 46th European Conference on Information Retrieval, 2024.

ECIR IR4Good 2024 Allen Lin, Jianling Wang, Ziwei Zhu, James Caverlee. Federated Conversational Recommender Systems. 46th European Conference on Information Retrieval, 2024.

SDM 2024 Ajay Krishna Vajjala, Dipak Falgun Meher, Shrunal Pothagoni, Ziwei Zhu, David S. Rosenblum. Vietoris-Rips Complex: A New Direction for Cross-Domain Cold-Start Recommendation. The 2024 SIAM International Conference on Data Mining.

EMNLP 2023 **Anjishnu Mukherjee**, **Chahat Raj**, Ziwei Zhu, and Antonios Anastasopoulos. Global Voices, Local Biases: Socio-Cultural Prejudices across Languages. The 2023 Conference on Empirical Methods in Natural Language Processing.

EMNLP 2023 Findings Xiangjue Dong, Ziwei Zhu, Zhuoer Wang, Maria Teleki, and James Caverlee. Co^2PT: Mitigating Bias in Pre-trained Language Models through Counterfactual Contrastive Prompt Tuning. The 2023 Conference on Empirical Methods in Natural Language Processing.

EMNLP 2023 Findings Zhuoer Wang, Yicheng Wang, Ziwei Zhu, and James Caverlee. Unsupervised Candidate Answer Extraction through Differentiable Masker-Reconstructor Model. The 2023 Conference on Empirical Methods in Natural Language Processing.

CIKM 23 **Yuqing Zhou**, Tianshu Feng, Mingrui Liu, and Ziwei Zhu. A Generalized Propensity Learning Framework for Unbiased Post-Click Conversion Rate Estimation. The 32nd ACM International Conference on Information and Knowledge Management, 2023.

ACL 23 Findings Xiangjue Dong, Yun He, Ziwei Zhu, and James Caverlee. PromptAttack: Probing Dialogue State Trackers with Adversarial Prompts. Findings of the Association for Computational Linguistics 2023.

WWW 23 Allen Lin, Ziwei Zhu, Jianling Wang, and James Caverlee. Enhancing User Personalization in Conversational Recommenders. The 2023 ACM Web Conference, 2023.

MobiSys 23 Liuyi Jin, Tian Liu, Amran Haroon, Radu Stoleru, Michael Middleton, Ziwei Zhu, Theodora Chaspari. EMSAssist: An End-to-End Mobile Voice Assistant at the Edge for Emergency Medical Services. The 21st ACM International Conference on Mobile Systems, Applications, and Services, 2023.

ECIR 23 Han Zhang, Ziwei Zhu, and James Caverlee. Evolution of Filter Bubbles and Polarization in News Recommendation. The 45th European Conference on Information Retrieval, 2023. (short paper)

CIKM 22 Shuo Lin, Jianling Wang, Ziwei Zhu, and James Caverlee. Quantifying and Mitigating Popularity Bias in Conversational Recommender Systems. The 31st ACM International Conference on Information and Knowledge Management, 2022.

WWW 22 James Kotary, Ferdinando Fioretto, Pascal Van Hentenryck, and Ziwei Zhu. End-to-end Learning for Fair Ranking Systems. The Web4Good special track in 33rd ACM International Conference on World Wide Web, 2022.

WSDM 22 Ziwei Zhu and James Caverlee. Fighting Mainstream Bias in Recommender Systems via Local Fine Tuning. The 15th ACM International Conference on Web Search and Data Mining, 2022.

KDD 21 Ziwei Zhu, Yun He, Xing Zhao, and James Caverlee. Popularity Bias in Dynamic Recommendation. The 27th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2021.

SIGIR 21 Ziwei Zhu, Jingu Kim, Trung Nguyen, Aish Fenton, and James Caverlee. Fairness among New Items in Cold Start Recommender Systems. The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2021.

WSDM 21 Ziwei Zhu, Yun He, Xing Zhao, Yin Zhang, Jianling Wang, and James Caverlee. Popularity-Opportunity Bias in Collaborative Filtering. The 14th ACM International Conference on Web Search and Data Mining, 2021.

WWW 21 Xing Zhao, Ziwei Zhu, and James Caverlee. Rabbit Holes and Taste Distortion: Distribution-Aware Recommendation with Evolving Interests. The 32th International Conference on World Wide Web, 2021.

SDM 21 Jianling Wang, Kaize Ding, Ziwei Zhu, and James Caverlee. Session-based Recommendation with Hypergraph Attention Networks. The 2021 SIAM International Conference on Data Mining.

SIGIR 20 Ziwei Zhu, Shahin Sefati, Parsa Saadatpanah, and James Caverlee. Recommendation for New Users and New Items via Randomized Training and Mixture-of-Experts Transformation. The 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2020.

SIGIR 20 Ziwei Zhu, Jianling Wang, and James Caverlee. Measuring and Mitigating Item Under Recommendation Bias in Personalized Ranking Systems. The 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2020.

RecSys 20 Ziwei Zhu, Yun He, Yin Zhang, and James Caverlee. Unbiased Implicit Recommendation and Propensity Estimation via Combinational Joint Learning. The 14th ACM Conference on Recommender Systems, 2020. (short paper)

RecSys 20 Yin Zhang, Ziwei Zhu, Yun He, and James Caverlee. Content-Collaborative Disentanglement Representation Learning for Enhanced Recommendation. The 14th ACM Conference on Recommender Systems, 2020.

EMNLP 20 Yun He, Ziwei Zhu, Yin Zhang, Qin Chen, and James Caverlee. Infusing Disease Knowledge into BERT for Health Question Answering, Medical Inference and Disease Name Recognition. The 2020 Conference on Empirical Methods in Natural Language Processing.

WWW 20 Xing Zhao, Ziwei Zhu, Majid Alfifi, and James Caverlee. Addressing the Target Customer Distortion Problem in Recommender Systems. The 31st International Conference on World Wide Web, 2020. (short paper)

WSDM 20 Xing Zhao, Ziwei Zhu, Yin Zhang, and James Caverlee. Improving the Estimation of Tail Ratings in Recommender System with Multi-Latent Representations. The 13th ACM International Conference on Web Search and Data Mining, 2020.

WSDM 20 Jianling Wang, Kaize Ding, Ziwei Zhu, Yin Zhang, and James Caverlee. Elite Opinion Leaders in Recommendation Systems: Elicitation and Diffusion. The 13th ACM International Conference on Web Search and Data Mining, 2020

WSDM 20 Jianling Wang, Ziwei Zhu, and James Caverlee. User Recommendation in Content Curation Platforms. The 13th ACM International Conference on Web Search and Data Mining, 2020.

WWW 19 Ziwei Zhu, Jianling Wang, and James Caverlee. Improving Top-K Recommendation via Joint Collaborative Autoencoders. The 30th International Conference on World Wide Web, 2019. (short paper)

CIKM 18 Ziwei Zhu, Xia Hu, and James Caverlee. Fairness-Aware Tensor-Based Recommendation. The 27th ACM International Conference on Information and Knowledge Management, 2018.

ICDM 18 Yun He, Haochen Chen, Ziwei Zhu, and James Caverlee. Pseudo-Implicit Feedback for Alleviating Data Sparsity in Top-K Recommendation. The 2018 IEEE International Conference on Data Mining, 2018. (short paper)

BSN 17 Ziwei Zhu, Sebastian Ober, and Roozbeh Jafari. Modeling and Detecting Student Attention and Interest Level Using Wearable Computers. The 14th IEEE International Conference on Wearable and Implantable Body Sensor Networks, 2017.

Pre-print Papers

arxiv **Yuqing Zhou** and Ziwei Zhu. Towards Robust Text Classification: Mitigating Spurious Correlations with Causal Learning.

arxiv **Bowen Wei** and Ziwei Zhu. Neural Symbolic Logical Rule Learner For Interpretable Learning.

arxiv **Bowen Wei** and Ziwei Zhu. Advancing Interpretability in Text Classification through Prototype Learning.

arxiv Beidi Dong, Jin R. Lee, Ziwei Zhu, and **Balassubramanian Srinivasan**. Assessing Large Language Models for Online Extremism Research: Identification, Explanation, and New Knowledge.

Workshop Papers

PaRiS 24 **Jinhao Pan**, **Bowen Wei**, and Ziwei Zhu. Reducing User Mainstream Bias in Personalized Recommendation via End-to-End Adaptive Local Learning. Third Workshop on Personalization and Recommendations in Search (PaRiS) at the SIGIR 2024.

EAI 23 Han Zhang, Ziwei Zhu, and James Caverlee. Alleviating Filter Bubbles and Polarization in News Recommendation via Dynamic Calibration. 2nd ACM SIGKDD Workshop on Ethical Artificial Intelligence: Methods and Applications, 2023.

FAccTRec 22 Allen Lin, Ziwei Zhu, Jianling Wang, and James Caverlee. Towards Fair Conversational Recommender Systems. The 5th FAccTRec Workshop on Responsible Recommendation at RecSys 2022.

DSAI4RRS Ziwei Zhu, Yun He, Xing Zhao, and James Caverlee. Evolution of Popularity Bias: Empirical Study and Debiasing. KDD 2022 Workshop on Data Science and Artificial Intelligence for Responsible Recommendations, 2022. (Best Paper Award)

SSL 21 Yun He, Ziwei Zhu, Yin Zhang, Qin Chen and James Caverlee. Infusing disease knowledge into BERT for Health Question Answering, Medical Inference and Disease Name Recognition. The Workshop on Self-Supervised Learning for the Web at WWW, 2021.

FatRec 18 Ziwei Zhu, Jianling Wang, Yin Zhang, and James Caverlee. Fairness-Aware Recommendation of Information Curators. The 2nd FATREC Workshop on Responsible Recommendation, 2018.

FUNDED PROJECTS

Title: Break the Dilemmas between Model Performance and Fairness: A Holistic Solution for Fairness Learning on Graphs
Sponsor: 4-VA
PIs: Ziwei Zhu (GMU PI), Dawei Zhou (VT PI)
Total: $25,000; My Share: $5,000
Time Period: 10/2022 – 10/2023

Title: Towards Holistic and Dynamic Debiasing for Online Search and Recommendation
Sponsor: 4-VA
PIs: Ziwei Zhu (GMU PI), Dawei Zhou (VT PI)
Total: $24,900; My Share: $19,900
Time Period: 6/2023 – 6/2024

Title: Data-Centric Social Bias Mitigation for Large Language Model-based Cyberharassment Detection
Sponsor: Commonwealth Cyber Initiative
PIs: Ziwei Zhu (PI), Jin Lee (Co-PI)
Total: $50,000; My Share: $43,000
Time Perio0: 6/2024 – 6/2025

HONORS AND AWARDS

- TAMU CSE Graduate Research Excellence Award 2022.03
- WSDM Travel Grant 2022.03
- SIGIR Travel Grant 2021.07
- WSDM Travel Grant 2021.03
- SIGIR Travel Grant 2020.07
- CIKM Travel Grant 2018.09
- First Class Scholarship at Wuhan University (top 5%) 2015.10
- National Scholarship, Wuhan University (top 1%) 2014.10
- Third Class Scholarship, Wuhan University (top 30%) 2013.10

PROFESSIONAL SERVICE

Editorial board:

- ACM Transactions on Intelligent Systems and Technology

NSF Panelist: 2024 spring, 2024 fall, 2025 spring

Conference Program Committees

- KDD: 2022, 2023, 2024, 2025
- WSDM: 2022, 2023, 2024, 2025
- NeurIPS: 2024
- ICLR: 2025
- ICML 2025
- CIKM: 2023, 2024
- RecSys: 2023, 2024
- SDM: 2023, 2024, 2025
- AAAI: 2023, 2024, 2025
- SIGIR: 2022
- FAccT: 2023
- ECML: 2023, 2024
- PAKDD: 2024
- ASONAM: 2024
- ARR (since 2023)

Journal Reviewer

- IEEE Transactions on Knowledge and Data Engineering
- IEEE Transactions on Services Computing
- IEEE Intelligent Systems
- ACM Transactions on Information Systems
- ACM Transactions on Recommender Systems
- ACM Transactions on Intelligent Systems and Technology
- Information Processing and Management
- Big Data Journal
- The Electronic Library
- Knowledge-based Systems
- Machine Learning Journal
- Heliyon Journal
- Neurocomputing Journal
- International Journal of Human-Computer Interaction
- IEEE Transactions on Big Data
- IEEE Transactions on Mobile Computing
- ACM Computing Surveys

DEPARTMENT AND COLLEGE SERVICE

CS PhD Committee, 9/2022 - Now

INVITED TALKS

- Responsible AI for Education Panel at AI for Education Policy and Equity event, GMU, 10/2024
- Fairness in RecSys, VT, 10/2023
- Fairness in RecSys, American University, 10/2023
- Fairness in Artificial Intelligence, MPI ReConEx 2023, 04/2023
- Toward Fairness-aware Recommender Systems, DEFirst Seminar, 03/2023
- Fairness among New Items in Cold Start Recommender Systems, research seminar at Netflix, 07/2021
- Toward Fairness-aware Recommender Systems, University of North Texas, 04/2021
- Item Fairness in Recommender Systems, research seminar at Netflix, 08/2020