

Anjishnu Mukherjee

PhD in Computer Science | George Mason University

iamshnoo.github.io @ amukher6@gmu.edu github.com/iamshnoo  Google Scholar

Research Interests

I work on Responsible AI, specifically in the evaluation and mitigation of cultural biases within multimodal, multilingual applications with a focus on LLMs. I have published recently at EMNLP, NAACL, ECIR and AIES.

Education

Present Aug 2022	George Mason University Ph.D. in Computer Science, GPA: 3.81/4 Advisors: Antonios Anastasopoulos , Ziwei Zhu	Fairfax, USA
Jun 2021 Jul 2017	Indian Institute of Engineering Science and Technology (IIST) Bachelor of Technology (First class) in Computer Science, CGPA: 9.63/10	Kolkata, India

Experience

Aug 2024 May 2024	Amazon <i>Applied Scientist Intern</i> Mentor: Wenqin Hu <u>Project</u> : Grounded explanations for black-box Size Recommendation systems using LLMs	Seattle, USA
Aug 2022 May 2022	Google Summer of Code Google People + AI Research <i>Developer</i> Mentor: Ryan Mullins <u>Project</u> : Tabular Feature Attributions and Salience Methods for Language Interpretability using LIT	Remote
May 2022 July 2021	Wells Fargo <i>Program Associate</i> Mentor: Aswathy P S <u>Project</u> : Interactive visualizations to track adoption metrics for Microsoft Teams and Zoom, Automatic detection and editing of system configurations for video conferencing devices	Bangalore, India
Aug 2020 Jun 2020	Google Summer of Code mlpack <i>Student Developer</i> Mentor: Ryan Curtin , Sreenik Seal <u>Project</u> : Computer vision functionality and loss functions for mlpack v4	Remote
Aug 2020 Mar 2020	University of Bremen <i>Research Intern</i> Advisor: Gabriel Zachmann <u>Project</u> : Heuristic Time of Impact algorithm for Continuous Collision Detection for CollDet .	Bremen, Germany

Selected Publications

- [S.1] **Crossroads of Continents: Automated Artifact Extraction for Cultural Adaptation with Large Multimodal Models** [PDF | Code]
[Anjishnu Mukherjee](#), [Ziwei Zhu](#), [Antonios Anastasopoulos](#)
In Submission [2024]
- [C.6] **BiasDora: Exploring Hidden Biased Associations in Vision-Language Models** [PDF | Code]
[Chahat Raj](#), [Anjishnu Mukherjee](#), [Aylin Caliskan](#), [Antonios Anastasopoulos](#), [Ziwei Zhu](#)
Conference on Empirical Methods in Natural Language Processing [EMNLP Findings'24]
- [C.5] **Breaking Bias, Building Bridges: Evaluation and Mitigation of Social Biases in LLMs via Contact Hypothesis** [PDF | Code]
[Chahat Raj](#), [Anjishnu Mukherjee](#), [Aylin Caliskan](#), [Antonios Anastasopoulos](#), [Ziwei Zhu](#)
AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society [AIES'24]
- [C.4] **Global Gallery: The Fine Art of Painting Culture Portraits through Multilingual Instruction Tuning** [PDF | Code]
[Anjishnu Mukherjee](#), [Aylin Caliskan](#), [Ziwei Zhu](#), [Antonios Anastasopoulos](#)
Conference of the North American Chapter of the Association for Computational Linguistics [NAACL'24]

- [C.3] **SALSA: Saliency-Based Switching Attack for Adversarial Perturbations in Fake News Detection Models** [PDF | Code]
Chahat Raj*, [Anjishnu Mukherjee*](#), Hemant Purohit, Antonios Anastasopoulos, Ziwei Zhu
European Conference on Information Retrieval [ECIR'24]
- [C.2] **Global Voices, Local Biases: Socio-cultural Prejudices across Languages** [PDF | Code]
[Anjishnu Mukherjee*](#), Chahat Raj*, Ziwei Zhu, Antonios Anastasopoulos
Conference on Empirical Methods in Natural Language Processing [EMNLP'23]
- [C.1] **True and Fair: Robust and Unbiased Fake News Detection via Interpretable Machine Learning** [PDF | Code]
Chahat Raj, [Anjishnu Mukherjee](#), Ziwei Zhu
AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society [AIIES'23]

Skills

Tools	Python, PyTorch, HuggingFace, NumPy, Pandas, Scikit-Learn, Git, GitHub, AWS - SageMaker, S3, Athena
Areas of Expertise	Natural Language Processing, Multimodal Large Language Models, Model Training, Evaluation & Validation, Instruction-tuning, Prompt Engineering, LLM-human alignment, Vision, Ethics and Bias Mitigation, Interpretability & Explainability.

Talks

“Multilingual socio-cultural biases”

- › Multilinguality and Language Diversity - Invited Speaker Series @ UToronto Sep 2024 (Remote)

“Global Gallery: The Fine Art of Painting Culture Portraits through Multilingual Instruction Tuning”

- › MASC-SLL @ JHU May 2024 (Baltimore, USA)

“Cross-Lingual Biases and Cultural Understanding in LLMs”

- › NL+, the Natural Language Processing Lunch Seminar @ Notre Dame Feb 2024 (Remote)

Awards

Outstanding GTA Award @ George Mason University, 2024 For my work as a GTA for Advanced NLP, CS 678.

Summer Graduate Research Award @ George Mason University, 2023 For summer research on evaluating socio-cultural biases in multilingual language models.

Scholarship @ DAAD WISE, 2020 Received a scholarship from the German Academic Exchange Service as 1 of 130 recipients from all over India for a summer internship in Germany.

Summer Research Fellowship @ IAS 2020 Received a fellowship from the Indian Academy of Sciences for a summer internship at IISc Bangalore.

Teaching

CS 678 Advanced Natural Language Processing, GMU *Teaching Assistant* Fall'23, Spring'24

- › Responsible for aiding in developing new assignments, 1:1 student tutoring and assignment grading.

CS 112 Introduction to Python Programming, GMU *Teaching Assistant* Fall'22, Spring'23

- › Responsible for teaching labs, 1:1 student tutoring, creating lab modules, assignment grading, monitoring exams.

Academic Service

Reviewer	AfricaNLP @ ICLR'22, ICLR'23, ICLR'24, EMNLP'23, New In ML @ NeurIPS'23, ACL ARR October'23, Feb'24 (ACL'24), Jun'24 (EMNLP'24), LTEDI @ EACL'24
Sub-Reviewer	AIIES'24, AIIES'23, COLM'24