

Bhavya Ghai

Applied Scientist, Amazon

🏠 Webpage: cs.stonybrook.edu/~bghai/

🐦 Twitter: [@Bhavya_Ghai](https://twitter.com/Bhavya_Ghai)

✉ Email: bghai@cs.stonybrook.edu

🌐 LinkedIn: in.linkedin.com/in/bhavyaghai/

🐙 Github: [bhavyaghai](https://github.com/bhavyaghai)

RESEARCH SUMMARY

My PhD research, at the intersection of ML, HCI and Data Visualization, focused on human-in-the-loop AI approach to tackle algorithmic bias. I have conducted empirical studies and developed visual interactive systems to identify and mitigate algorithmic bias in its different manifestations like classification models and word embeddings. The broader objective is to tackle rising concerns of fairness, accountability, interpretability & trust prevalent with fully automated approaches. More recently, I have primarily focused on large language models like Claude, Mixtral, etc. to address fairness and privacy concerns using natural language techniques like prompting, RAGs, fine-tuning, knowledge distillation, etc.

EDUCATION

Stony Brook University, State University of New York, NY

PhD in Computer Science, GPA 3.81/4

Aug'2016-May'23

MS in Computer Science

Adviser: Prof. Klaus Mueller

Dissertation: Tackling Algorithmic Bias through Human-in-the-loop AI

Indian Institute of Information Technology, Gwalior, India

2011-16

B. Tech + M. Tech (Integrated 5-year program), Computer Science (IT), GPA 8.20/10

Adviser: Prof. Anupam Shukla

Master's Thesis: Multi-level Ensemble Learning based Recommender System

RESEARCH INTERESTS

Algorithmic Fairness, Human-in-the-loop AI, Data Visualization, Machine Learning, HCI, Explainable AI, NLP

HONORS & FELLOWSHIPS

SIGIR Student Travel Award, National Science Foundation

2022

IACS Travel Award, Institute of Advanced Computational Sciences, Stony Brook University (\$2,000)

2022

Young Writer's Award, Institute of Advanced Computational Sciences, Stony Brook University

2020

Best Research Talk Award, Graduate Research Day, Stony Brook

2019

Winner, 3 Minute Thesis Competition, Stony Brook University

2019

Junior Researcher Award, Institute of Advanced Computational Sciences, Stony Brook University (\$34,000)

2018, 2019

Full Silver Scholarship, Open Data Science Conference, Boston

2018

Data Science for Social Good Fellowship, University of Washington (Declined)

2018

Bloomberg Immersion Fellow, Bloomberg

2017

Conference Scholarship, Fairness, Accountability & Transparency Conference (FAccT), NY

2017

Data Science for Social Good Fellow, Georgia Tech, Atlanta

2017

Travel Grant Award, Data Science for Social Good Conference, Chicago

2017

Chairman's Fellowship Award, CS Dept. Stony Brook University (\$3,000)

2016

Post Graduate Scholarship, All India Council for Technical Education

2015, 2016

Oracle Certified Java Professional for Java Standard Edition 6

2014

CURRENT RESEARCH

WorDebias: A Visual tool for Exploration & Mitigation of Word Embedding Bias through Human-in-the-Loop AI

Social biases & stereotypes recorded in text are reflected in word embeddings. To prevent propagation of such biases, we present a visual framework for detecting different kinds of biases and mitigating them using a novel active learning-based debiasing approach. Our framework shows promising results for word embedding models trained over languages like english and hindi.

FairWriter: An AI-Powered Writing Assistant for Gender-Inclusive Writing

FairWriter can be loosely understood as a spell checker for gender bias. It employs a bunch of NLP techniques to facilitate gender inclusive writing by promoting the use of gender-neutral pronouns and discouraging gendered nouns and sexist phrases.

Towards Inclusive Group Fairness

Existing interventions aimed at enhancing group fairness solely try to reduce disparity between groups without any consideration of how that is achieved. This can lead to worse performance for the privileged and/or the unprivileged group(s). We propose a novel model-based intervention that enhances group fairness while ensuring that no population group is harmed in the process.

RESEARCH EXPERIENCE

Amazon

Feb'23-present

Applied Scientist

New York, NY

Manager: Dr. Srinivasan Sengamedu

Working with the People eXperience and Technology Central Science Team (PXTCS) on natural language processing, privacy and algorithmic fairness. We leverage large language models to solve business problems in the HR space, including detecting and mitigating social bias and anonymizing text using techniques like prompting, RAG, fine-tuning, knowledge distillation, etc.

Twitter

May'22-Aug'22

ML Engineering Intern

Remote

Adviser: Dr. Yunfeng Zhang

Worked with the META (ML Ethics, Transparency and Accountability) team, a part of Twitter Cortex, on integrating a popular fairness auditing tool into their ecosystem and customizing it to meet their unique needs.

IBM T. J. Watson Research Center

May'19-Aug'19

Research Intern

Yorktown Heights, NY

Adviser: Dr. Q. Vera Liao

Devised a new paradigm called Explainable Active Learning (XAL). Our empirical study showed that adding local explanations for a model's prediction at every step of the active learning improves annotator's experience without compromising performance.

Nokia Bell labs

May'18-Aug'18

Research Intern

Murray Hill, NJ

Adviser: Buvana Ramanan

Worked with the APSS Lab to build robust Automatic Speech Recognition systems using speech enhancement of publicly available data. Our approach reduces the Word error rate for the DeepSpeech model by ~9.5% compared to training with noisy data.

Georgia Institute of Technology (GaTech)

May'17-July'17

Research Intern

Atlanta, GA

Adviser: Prof. Ellen W. Zegura

We tried to counter housing injustice by analyzing a proposed anti-displacement program in the Westside region of Atlanta (GA). We estimated the total number of households which will benefit from this policy and forecasted the total cost of this program over a 7-year period. We found that only 1 in 8 households will be covered & only a quarter of the estimated budget is allocated.

Indian Institute of Technology, Delhi (IIT)

May'15-June'15

Research Intern

New Delhi, India

Adviser: Prof. G. N. Tiwari

Analyzed 35 years of Delhi weather data for predicting solar radiation to ensure a power-supply equilibrium. Used time series analysis to predict solar radiation 6 months into the future with more than 80% accuracy.

PUBLICATIONS

Peer-reviewed Conference and Journal papers

- The HaLLMark Effect: Supporting Provenance and Transparent Use of Large Language Models in Writing through Interactive Visualization
Md Naimul Hoque, Tasfia Mashiat, **Bhavya Ghai**, Cecilia Shelton, Fanny Chevalier, Kari Kraus, Niklas Elmqvist
ACM Conference on Human Factors in Computing Systems, ACM CHI 2024

- **Portrayal: Leveraging NLP and Visualization for Analyzing Fictional Characters**
Md Naimul Hoque, **Bhavya Ghai**, Kari Kraus, Niklas Elmqvist
ACM Conference on Designing Interactive Systems, ACM DIS 2023
- **D-BIAS: A Causality-Based Human-in-the-Loop System for Tackling Algorithmic Bias**
Bhavya Ghai and Klaus Mueller
IEEE Visualization Conference, to appear, Oklahoma City, IEEE VIS 2022
- **DramatVis Personae: Visual Text Analytics for Identifying Social Biases in Creative Writing**
Md Naimul Hoque, **Bhavya Ghai**, Niklas Elmqvist
ACM Conference on Designing Interactive Systems, ACM DIS 2022
- **Cascaded Debiasing: Studying the Cumulative Effect of Multiple Fairness-Enhancing Interventions**
Bhavya Ghai, Mihir Mishra, Klaus Mueller
ACM Conference on Information and Knowledge Management, Atlanta, ACM CIKM 2022
- **Fluent: An AI Augmented Writing Tool for People who Stutter**
Bhavya Ghai and Klaus Mueller
ACM SIGACCESS Conference on Computers and Accessibility, ACM ASSETS 2021
- **Explainable Active Learning (XAL): Toward AI Explanations as Interfaces for Machine Teachers**
Bhavya Ghai, Q. Vera Liao, Yunfeng Zhang, Rachel K. E. Bellamy, Klaus Mueller
ACM Conference on Computer-Supported Cooperative Work and Social Computing, ACM CSCW 2020
- **Wave Front Method Based Path Planning Algorithm for Mobile Robots**
Bhavya Ghai and Anupam Shukla
International Conference on ICT for Intelligent Systems, Ahmedabad, ICTIS 2016
- **Energy Efficient Dynamic Nearest Node Election for Localizations of Mobile Node in Wireless Sensor Networks**
Bhavya Ghai, Girish Pradeep Bindalkar, Sanjeev Sharma, Anupam Shukla
IEEE Conference on Computational Intelligence & Computing Research, IEEE ICCIC 2015

Workshop papers, Extended Abstracts, Non-archival papers

- **WordBias: An Interactive Visual Tool for Exploring Intersectional Social Biases Encoded in Word Embeddings**
Bhavya Ghai, Md Naimul Hoque, Klaus Mueller
ACM Conference on Human Factors in Computing Systems (LBW), ACM CHI 2021
- **Active Learning++: Incorporating Annotator's Rationale using Local Model Explanation**
Bhavya Ghai, Q. Vera Liao, Yunfeng Zhang, Klaus Mueller
Workshop on Data Science with Human in the Loop, ACM KDD 2020
- **Measuring Social Biases of Crowd Workers using Counterfactual Queries**
Bhavya Ghai, Q. Vera Liao, Yunfeng Zhang, Klaus Mueller
Workshop on Fair & Responsible AI, ACM CHI, Honolulu, 2020
- **Does Speech enhancement of publicly available data help build robust Speech Recognition Systems?**
Bhavya Ghai, Buvana Ramanan, Klaus Mueller
AAAI Conference on Artificial Intelligence (extended abstract), New York, NY, 2020
- **Visualization of Multivariate Data with Network Constraints using Multi-Objective Optimization**
Bhavya Ghai, Alok Mishra, Klaus Mueller
IEEE VIS Conference (extended abstract), Phoenix, AZ, 2017
- **Coupling data science with participatory planning for equity in urban renewal programs: An Analysis of Atlanta's Anti-Displacement Tax Fund**
Jeremy Auerbach, Hayley Barton, Takeria Blunt, Vishwamitra Chaganti, **Bhavya Ghai**, Amanda Meng, Christopher Blackburn, Ellen Zegura
Bloomberg Data for Good exchange, New York, NY, 2017

TECHNICAL SKILLS

Languages	<i>Python, Javascript, Java, SQL, R</i>
Data Analytics	<i>Jupyter Notebook, pandas, numpy, sklearn, pytorch</i>
Visualization	<i>D3.js, Plotly.js, Cytoscape, Matplotlib</i>

TEACHING EXPERIENCE

- **Teaching Assistant** Aug'16-Dec'19
Stony Brook University Stony Brook, NY
CSE 303 Theory of Computation, CSE 214 Data Structures, CSE 337 Scripting languages, CSE 564 Data Visualization
- **Guest Lecturer** Aug'17-Dec'19
Stony Brook University Stony Brook, NY
Instructor: Prof. Robert Harrison
Gave lectures on identifying and mitigating social biases encoded in static word embeddings for AMS 561 Introduction to Computational and Data Science

PATENTS

- **Optimizing a Machine Learning System**
Yunfeng Zhang, Qingzi Liao, Bhavya Ghai, Klaus Mueller
Publication #US20220237504A1. Filing date: 01/26/2021. Publication date: 07/28/2022.
- **Active Learning using Causal Network Feedback**
Qingzi Liao, Bhavya Ghai, Yunfeng Zhang, Tian Gao
Publication #US20220245508A1. Filing date: 02/02/2021. Publication date: 08/04/2022.

PROFESSIONAL ACTIVITIES AND SERVICE

- **Program Committee**
ACM FAccT 2023, ACM AIES 2023, ACM EAAMO 2023, ACM CIKM 2023
- **Reviewer**
IEEE PacificVis 2021, ACM SIGCHI 2022, ACM DIS 2022, IEEE VIS 2022, ACM SIGCHI 2023, ACM TiS 2022, ACM DIS 2023, IEEE VIS 2023, IEEE TVCG
- **Student Volunteer**
ACM FAccT 2021, ACM IUI 2021, NeurIPS 2020, ACM CSCW 2020, IEEE VIS 2020, ACM FAT* 2019, IEEE VIS 2022
- **Presentations**
Symposium on Biases in Human Computation and Crowdsourcing (BHCC), Sheffield, 2019
Doctoral Colloquium, IEEE VIS Conference, Vancouver, 2019
Doctoral Colloquium, ACM FAT* Conference, Atlanta, 2019
Panel Speaker, Bloomberg Data for Good Exchange, New York City, 2017

REFERENCES

Klaus Mueller, Ph.D.
Professor
Stony Brook University, NY
mueller@cs.stonybrook.edu

Q. Vera Liao, Ph.D.
Principal Researcher
Microsoft Research, Montreal
veraliao@microsoft.com

Additional references available upon request