

Ayan Majumdar

Address: Max Planck Institute for Software Systems (MPI-SWS)

Building E1 5, Campus, Room 538, 66123 Saarbrücken, Germany

✉ [ayanm\[at\]mpi-sws.org](mailto:ayanm[at]mpi-sws.org)

🌐 <https://ayanmaj.netlify.app/>

☎ +49-15252879478

🔗 <https://scholar.google.com/citations?user=HeQLWA4AAAAJ>

🌐 <https://linkedin.com/in/ayan-92/>

🆔 <https://orcid.org/0009-0004-7535-6850>

🐙 <https://github.com/ayanmaj92>

About

I am passionate about bridging the gap between theoretical research in fairness, causality, and explainability and the practical deployment of safety-compliant Automated Decision-Making and Generative AI systems.

Research Interests

- Trustworthy AI 📖 Responsible AI, AI Safety, Fairness, Explainability, Causality
- Deep Learning 📖 Generative AI, Large Language Models, Vision Language Models, Model Alignment

Education


- 2021 – 2026
(expected) 📖 **Ph.D. in Computer Science**
MPI-SWS and Saarland University, Saarbrücken, Germany
Area: *Deep Learning, Generative AI, Fairness, Explainability*
Advisors: Prof. Dr. Krishna P. Gummadi, Prof. Dr. Isabel Valera
Contact: [gummadi\[at\]mpi-sws.org](mailto:gummadi[at]mpi-sws.org), [ivalera\[at\]cs.uni-saarland.de](mailto:ivalera[at]cs.uni-saarland.de)
- 2017 – 2021 📖 **M.Sc. in Computer Science**
Saarland University, Saarbrücken, Germany
Thesis: *Generating Counterfactuals for Causal Fairness*
Outline: *Deep generative models and their implicit assumptions in generating counterfactuals from observed data in the context of fairness.*
Advisors: Prof. Dr. Krishna P. Gummadi, Prof. Dr. Isabel Valera
GPA: 1.2/1.0 (German Scale)
- 2011 – 2015 📖 **B.Tech. in Electronics & Communication**
Heritage Institute of Technology, Kolkata, India
Project: *Automated traffic detection using image processing*
GPA: 8.8/10.0

Publications






Conference Proceedings

- 1 **Majumdar***, A., Kanubala*, D. D., Gupta, K., & Valera, I. (2026). A causal framework to measure and mitigate non-binary treatment discrimination. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 40). 📄 Paper 🔄 Code.
- 2 **Majumdar**, A., & Valera, I. (2024). CARMA: A practical framework to generate recommendations for causal algorithmic recourse at scale. In *2024 ACM Conference on Fairness, Accountability, and Transparency*. 📄 Paper 🔄 Code.
- 3 Nanda, V., **Majumdar**, A., Kolling, C., Dickerson, J. P., Gummadi, K. P., Love, B. C., & Weller, A. (2023). Do invariances in deep neural networks align with human perception? In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 37). 📄 Paper 🔄 Code.
- 4 Rateike*, M., **Majumdar***, A., Mineeva, O., Gummadi, K. P., & Valera, I. (2022). Dont Throw it Away! The Utility of Unlabeled Data in Fair Decision Making. In *2022 ACM Conference on Fairness, Accountability, and Transparency*. 📄 Paper 🔄 Code.

Preprints & Under Review



- 5 Budde, L. M., **Majumdar, A.**, Uth, R., Langer, M., & Valera, I. (2026). From universal to individualized actionability: Revisiting personalization in algorithmic recourse. Under review.
- 6 Singh, P., Jarollahi, S., **Majumdar, A.**, Pahari, V., Chakraborty, A., Dash, A., ... Weber, I. (2026). From policy to practice: The anatomy of bluesky moderation service. Under review.
- 7 **Majumdar, A.**, Chen, F., Li, J., & Wang, X. (2025). Evaluating LLMs for demographic-targeted social bias detection: A comprehensive benchmark study. *arXiv preprint arXiv:2510.04641*.  Paper.

Work Experience



- Apr. 2021 – ongoing  **Doctoral Researcher**
Max Planck Institute for Software Systems, Saarbrücken, Germany
Role: Leading multiple projects on improving the practicality of Responsible AI and Safety in the context of Automated Decision-Making and Generative AI.
Supervisor: Prof. Dr. Krishna P. Gummadi, Prof. Dr. Isabel Valera
- Aug. 2024 – Feb. 2025  **Ph.D. Research Intern**
Huawei Research TTE-DE Lab, München, Germany
Project: Scaling-up Data Bias Detection & Survey on AI Governance
Role: Developed a benchmark using LLMs to detect demographic-targeted social bias in large datasets (preprint [7]). Conducted research surveys to support AI governance and alignment with regulations like the EU AI Act.
- Oct. 2019 – Mar. 2021  **Research Assistant**
Max Planck Institute for Software Systems, Saarbrücken, Germany
Project: *Exploring bias and fairness with deep generative models*
Role: Led project regarding exploration of bias in deep generative models for facial image data; designed core methodologies and evaluations.
Supervisor: Prof. Dr. Krishna P. Gummadi
- Apr. 2018 – Mar. 2019  **Research Assistant**
SFB1102, Saarland University, Saarbrücken, Germany
Project: *Mutual Intelligibility in Slavic Languages*
Role: Engineered data processing to collect large-scale corpora, developed web-based user studies to evaluate mutual intelligibility in Slavic languages.
Supervisor: Prof. Dr. Dietrich Klakow
- Jul. 2015 – Aug. 2017  **Systems Engineer**
Infosys Ltd., Bengaluru, India
Role: Optimized and validated Session Initiation Protocol (SIP) and VoIP functionalities within session border controllers. Recognized with the Infy Insta Award (2017) and Spot Award (2016) for critical project contributions.

Technical Skills

Generative AI & LLM Engineering



- Frameworks  Transformers, vLLM, Sentence-Transformers, FAISS
- Techniques  Prompt Optimization, In-Context Learning, Multi-modal Inference, RAG, Fine-tuning, Benchmarking, LLM-as-a-Judge

Responsible AI & AI Safety




- Frameworks  AIF360, Fairlearn, CleverHans, Foolbox, robustness
- Techniques  Causal Inference, Counterfactual Explanations, Statistical & Causal Fairness

Technical Skills (continued)



Programming & Core Infrastructure

Core ML  PyTorch, Lightning, Scikit-learn, NumPy, Pandas, Scipy, Optuna
Languages & Tools  Python, Git, L^AT_EX, Weights & Biases, Matplotlib, Seaborn

Teaching Assistance




Summer 2024  **Machine Learning (Core Lecture)**, Saarland University
Summer 2021  **Machine-Assisted Decision Making (Seminar)**, Saarland University
Summer 2019  **Statistical Natural Language Processing (Advanced Lecture)**, Saarland University

Mentoring Experience




Sept. 2025 – Jan. 2026  Co-mentored Lena Marie Budde (M.Sc., Saarland University) during thesis and project on **personalization in algorithmic recourse**.
May 2024 – July 2024  Co-mentored Pranav Nyati (B.Tech.+M.Tech., IIT-Kharagpur) during DAAD summer internship on **algorithmic recourse using swarm algorithms**.

Talks and Posters


Invited Talks

2025  **A Causal Framework to Measure and Mitigate Non-binary Treatment Discrimination**
LMU-München, University of Amsterdam
2024  **Scaling Up Causal Algorithmic Recourse with CARMA**
Center for Perspicuous Computing 2024 Meeting
2022  **Don't Throw it Away! The Utility of Unlabeled Data in Fair Decision Making**
SAP Innovation Sessions, MPI for Intelligent Systems, MILA Quebec AI Institute







Conference Talks

2026  **A Causal Framework to Measure and Mitigate Non-binary Treatment Discrimination**
AAAI 2026 (Poster)
2024  **Scaling Up Causal Algorithmic Recourse with CARMA**
ACM FAccT 2024, European Workshop on Algorithmic Fairness 2024 (Poster)
2022  **Don't Throw it Away! The Utility of Unlabeled Data in Fair Decision Making**
ACM FAccT 2022


Academic Activities

Program Committee  FAccT 2024-25, ICLR 2023-25, ICML 2023-24, NeurIPS 2023/25, WebConf-Web4Good 2026, AAAI 2025-26, AIES 2025

Language Skills

Native/Fluent  English  Bengali 
Proficient/Intermediate  German  Hindi 

Achievements & Awards

 Infy Insta award (2017) and Spot award (2016) for significant contributions to the project, Infosys, India.