

Avijit Ghosh

☎ (+1) 857-337-0180 | ✉ ghosh.a@northeastern.edu | 🌐 evijit.github.io | in evijit | 📷 evijit

Algorithmic Fairness

Ethical AI

Machine Learning

AI Explainability

Computational Social Science

Education

Northeastern University

Ph.D. candidate in Computer Science (Advised by Dr. Christo Wilson)

Boston, MA

2019 - 2023 (Expected)

Indian Institute of Technology (IIT) Kharagpur

B.Tech. in Chemical Engineering, M.Tech in Financial Engineering, Minor in Computer Science

Kharagpur, India

2014 - 2019

Doctoral Thesis

Algorithmic Fairness in the Real World: Challenges and Considerations

Expected defense 2023

- Social bias in machine learning algorithms is a widespread problem that has been addressed through various measures, but implementing fair machine learning systems in the real world is challenging due to issues like noisy demographic information, adversarial vulnerabilities, policy restrictions and complex interactions between humans and algorithms.
- In my thesis, I attempt to outline these problems in fair ML systems, with the aim to gain a more complete understanding of the challenges involved and to be able to provide technical and policy recommendations to overcome their real world implementation challenges.

Publications

Peer-Reviewed Conference

Subverting Fair Image Search with Generative Adversarial Perturbations

Avijit Ghosh, Matthew Jagielski, Christo Wilson

FACCT '22

Seoul, South Korea

FairCanary: Rapid Continuous Explainable Fairness

Avijit Ghosh*, Aalok Shanbhag*, Christo Wilson

AIES '22

Oxford, United Kingdom

Algorithms that "Don't See Color": Comparing Biases in Lookalike and Special Ad Audiences

Piotr Sapiezynski, Avijit Ghosh, Levi Kaplan, Alan Mislove, Aaron Rieke

AIES '22

Oxford, United Kingdom

When Fair Ranking Meets Uncertain Inference

Avijit Ghosh, Ritam Dutt, Christo Wilson

SIGIR '21

Montreal, Canada / Virtual

Building and Auditing Fair Algorithms: A Case Study in Candidate Screening

Christo Wilson, Avijit Ghosh, Shan Jiang, Alan Mislove, Lewis Baker, Janelle Szary, Kelly Trindel, Frida Polli

FACCT '21

Toronto, Canada / Virtual

Public Sphere 2.0: Targeted Commenting in Online News Media

Ankan Mullick, Sayan Ghosh*, Ritam Dutt*, Avijit Ghosh*, Abhijnan Chakrabarty

ECIR '19

Cologne, Germany

Peer-Reviewed Workshop

Can There be Art Without an Artist?

Avijit Ghosh, Genoveva Fossas

HEGM@NeurIPS '22

New Orleans, USA

Characterizing Intersectional Group Fairness with Worst-Case Comparisons

Avijit Ghosh, Lea Genuit, Mary Reagan

AIDBEI@AAAI '21

Vancouver, Canada / Virtual

Analyzing Political Advertisers' Use of Facebook's Targeting Features

Avijit Ghosh, Giridhari Venkatadri, Alan Mislove

Conpro@S&P '19

San Francisco, USA

SAVITR: A System for Real-time Location Extraction from Microblogs during Emergencies

Ritam Dutt, Kaustubh Hiware, Avijit Ghosh, Rameshwar Bhaskaran

SMERP@WWW '18

Lyon, France

WebSelect: A Research Prototype for Optimizing Ad Exposures based on Network Structure

Avijit Ghosh, Agam Gupta, Divya Sharma, Uttam Sarkar

WITS'19

Dublin, Ireland

Peer-Reviewed Journal

Connectedness of Markets with Heterogeneous Agents and the Information Cascades

Avijit Ghosh, Aditya Chourasiya, Lakshay Bansal, Abhijeet Chandra

AAA'21

Journal

Preprints and Working Manuscripts

UComparing Fair Classifiers With and Without Demographic Attributes

Avijit Ghosh, Pablo Kvitca, Christo Wilson

Working Manuscript

Dangers of using improper datasets for fair ranking

Avijit Ghosh, Amifa Raj, Christo Wilson, Kristian Lum, Michael Ekstrand

Working Manuscript

Unified Shapley Framework to Explain Prediction Drift

Aalok Shanbhag*, Avijit Ghosh*, Josh Rubin*

Preprint

Supervised extraction of catchphrases from legal documents

Avijit Ghosh*, Prerit Gupta*, Ritam Dutt, Kaustubh Hiware, Arpan Mandal, Kripabandhu Ghosh, Sap-tarshi Ghosh

Term paper

* Equal contribution

Awards and Grants

2022	Winner Best Paper - Runner Up	Conpro'22
2019	Winner Best Poster Award	ECIR'19
2019	Dean's Fellowship for first Year PhD students (USD 72K)	Northeastern University
2019	Winner Institute Order of Merit - Technology	IIT Kharagpur
2018	Winner SGSIS Institute Challenge Grant (INR 1M)	IIT Kharagpur
2017	Silver Medal Stock Market Analysis	Inter IIT Tech Meet, Kanpur
2016	Gold Medal Software Development	Inter IIT Tech Meet, Mandi
2012	Governor's Medal National Rank 5, ICSE Board	Government of West Bengal
2010	NTSE Scholar National Talent Search Examination	NCERT

Teaching

Algorithmic Auditing

Teaching Assistant for Dr. Piotr Sapiezynski

Northeastern University

Spring 2022

- Designing audits that measure the effects of interests and control noise sources
- Minimize potential harms of audits to all stakeholders
- Legal bounds of algorithm audits
- Beyond audits: potential harms that cannot be measured through audits

Academic Experience

Northeastern University

Boston, MA

Research Assistant at Khoury College of Computer Sciences

Sep. 2019 – Present

- Analyzing Fair ranking systems and showing how they fail in the presence of noisy protected attribute data. Also investigated adversarial attacks on the guaranteed fairness of retrieval systems.
- A cooperative fairness audit of the recommendation algorithm of [PyMetrics](#), a talent matching software. [Press Release](#).
- Investigated Facebook's Special Audiences system for opportunity advertisements and showed that the audience creation algorithm was still biased against women, seniors and minorities.
- Analyzed the ad reach and spend information obtained from Facebook's ad transparency feature and the personal targeting dataset from Propublica's Facebook ad dataset and showed that advertisers with higher budgets use more privacy sensitive targeting techniques like PII or Lookalike audiences. Findings published and presented at [IEEE ConPro 2019](#).

LIG, University of Grenoble Alps

Grenoble, France

Visiting Researcher

May 2019 – July 2019

- Study of how news companies promote different items on social media, investigating possible patterns of differential information spreading using both posts and ads.
- We also discovered and reported an exposed access token bug to [Facebook Bug Bounty](#).

IIT Kharagpur

Kharagpur, India

Undergraduate Researcher - Complex Networks Research Group

2014 – 2019

- Automated Extraction of Catchwords from Legal Documents using a novel NER tagger to help categorize lengthy legal texts.
- Automatically position user comments against relevant news article paragraphs. Presented at [ECIR 2019](#).
- Savitr - A real-time location extraction system for disaster management using twitter. Presented at [WWW-SMERP 2018](#).
- Classification and Summarization of tweets during a disaster event, presented at [IBM Day 2016](#).

Industry Experience

Twitter

San Francisco, CA

Research Intern

Sep – Dec 2021 and Jun – Aug 2022

- Worked with the META (Machine Ethics, Transparency and Accountability) team at Twitter, to investigate the relationship between demography agnostic and demography dependent author impression fairness metrics at scale.
- Developed home timeline diversity metrics based on user feedback, to find balance between recommendation efficiency and fairness.

Fiddler Labs

Palo Alto, CA

Research Intern

Oct 2020 – Apr 2021

- Explain distributional shifts in Machine Learning model outputs by unifying Shapley based methods.
- Using optimal transport theory, proposed a threshold independent fairness metric that allows for real time explanations.
- Worked with the product team and civil rights lawyers in the deployment of Fiddler's Machine Learning model fairness dashboard. Introduced and incorporated intersectional fairness metrics in the product.

Xerox Research Centre

Bangalore, India

Research Intern

May 2017 – July 2017

- Implemented XTrack, a Smart Vehicle Tracking and Battery usage minimizing Algorithm, using BLE to relay GPS information.
- Proposed a method for Uber-like Surge Price Prediction using Spatio-Temporal techniques like the Neural Hawkes and Recurrent Marked Temporal Point Process. Awarded the title of [Best Internship Project](#).

Google Summer of Code

Remote

GSoc Student at OpenMRS

Apr 2016 – Aug 2016

- Replaced the HTML XForms system used with native generated forms using the Forms REST Api in the android client of the Opensource Medical Record System. Added offline form saving. Configured Travis CI to automatically build and push the apk to the play store.
- Overall, contributed 100K lines of code and became the top code contributor in the project repository.

Academic Service

2023	The Web Conference	<i>Reviewer</i>
2022	ACM Conference on Fairness, Accountability, and Transparency	<i>Reviewer</i>
2022	AAAI/ACM Conference on AI, Ethics, and Society	<i>Reviewer</i>
2022	Conference on Neural Information Processing Systems	<i>Reviewer</i>
2022	Conference on Empirical Methods in Natural Language Processing	<i>Reviewer</i>
2021	Conference on Neural Information Processing Systems	<i>Reviewer</i>

Outreach and Leadership

2022	FAccT CRAFT Workshop on Identifying Queer Harms as a bias bounty with Queer in AI	<i>Organizer</i>
2021	SIGIR Queer in AI social with speaker panel on queer stereotypes in web search	<i>Organizer</i>
2017	Kharagpur Winter Of Code	<i>Founder</i>
2016	Kharagpur Open Source Society	<i>Founder</i>

Speaking Engagements

2023	Can There be AI Art Without an Artist?	<i>South By Southwest (SXSW)</i>
2022	Proxies for bias monitoring: Ethics workshop	<i>Centre for Data Ethics and Innovation, UK</i>
	Subverting Fair Image Search with Generative Adversarial	
2022	Perturbations as part of the 'Celebrating Young Researchers' event	<i>Trustworthy ML Initiative</i>

Press Mentions

2021	NYC aims to be first to rein in AI hiring tools	<i>Associated Press</i>
2021	Auditors are testing hiring algorithms for bias, but there's no easy fix	<i>MIT Technology Review</i>
2021	New York City Proposes Regulating Algorithms Used in Hiring	<i>Wired</i>
2021	Supporting Responsible Use of AI and Equitable Outcomes in Financial Services	<i>The Federal Reserve</i>
2019	Facebook Ads Can Still Discriminate Against Women and Older Workers, Despite a Civil Rights Settlement	<i>Propublica</i>
2019	Facebook Agreed Not to Let Its Ads Discriminate. But They Still Can.	<i>Mother Jones</i>

Technical Skills

Languages: Python, Java, C, R, Bash, SQL, HTML/CSS, JavaScript, Matlab

Tools & Frameworks: Git, TravisCI, Pytorch, Keras, TensorFlow, Docker, AWS, Sagemaker, Google Cloud ML, Android

Machine Learning: Transformers/BERT, ResNets, Adversarial examples, Fairness, Explainability