# Falaah Arif Khan

COMPUTER SCIENTIST · ARTIST

*567 Ocean Parkway, Apt 6E, Brooklyn, New York, 11218.*

+1 646-203 8253  |  fa2161@nyu.edu  |  https://falaaharifkhan.github.io/research/  |  @FalaahArifKhan

## Education

**New York University**                                                                          *New York, USA*
PHD IN DATA SCIENCE                                                                              *2021 - present*
- CGPA: 3.892/4
- My research broadly focuses on 'AI fairness' and 'AI safety'.

**Shiv Nadar University**                                                                 *Uttar Pradesh, India*
B.TECH IN ELECTRONICS AND COMMUNICATION ENGINEERING (WITH DISTINCTION)                              *2014-2018*
- CGPA: 8.87/10
- Minor in Mathematics
- Thesis: Behavioral Biometrics and Machine Learning to secure Web Logins
- Supervisors: Sajin Kunhambu (Dell) and Madhur Deo Upadhayay (Shiv Nadar University)

## Professional Experience

**Center for Responsible AI, New York University**                                           *New York City, USA*
ARTIST IN RESIDENCE                                                                          *Oct 2020 - Dec 2023*
- I explored the use of comic books as a novel medium of scientific scholarship, and as a vehicle for public education.
- I created the 'Data Responsibly' comic series in collaboration with Prof Stoyanovich, targeted at data science students and practitioners. Volume 1 titled 'Mirror, Mirror' is a primer on responsible data science and the impact of poorly designed systems on marginalized social groups, and ties together these ideas with broader issues in the ML landscape, including problems with exclusionary discourse, misguided incentive structures in scholarship, and questions of culpability when things go wrong. Volume 2 titled 'Fairness and Friends' is our attempt to bridge 'fairness' ideas from different disciplines, notably, computer science and political philosophy using doctrines of equality of opportunity.
- I co-created the 'We Are AI' comic series, with Prof Stoyanovich, as part of the center's effort to develop educational materials and methodologies to teach members of the public about responsible AI. It is a 5-part series and accompanies our free public education course (of the same name), with one volume for each module.

**Montreal AI Ethics Institute**                                                                     *Remote*
ARTIST IN RESIDENCE                                                                          *July 2020 - Feb 2021*
- I explored into the socio-political underpinnings of data-driven technology, with an emphasis on the role of power and how it influences the design of AI. My visual essay 'Decoded Reality' is an artistic depiction of how algorithmic interventions manifest in society, and was an exhibit at the Resistance AI Workshop at NeurIPS 2020

**International Institute of Information Technology, Hyderabad**                                        *Remote*
RESEARCH FELLOW                                                                                *Aug - Dec 2020*
- I worked on the neural machine translation of Indic languages, focusing on the robustness of translations between different information domains and different source and target languages.

**Dell EMC**                                                                                  *Bangalore, India*
RESEARCH SOFTWARE ENGINEER II                                                               *Aug 2019 - May 2020*
- I explored the use of machine learning to enhance the security of Dell's Identity and Access Management product.
- I implemented a bot detection filter that uses a game theory to identify malicious login attempts performed from headless devices (without browser activity), and is used on Dell login pages, globally.
- I designed and implemented a novel adaptively-robust CAPTCHA that learns on the fly and does not require periodic manual re-design under evolving threat landscapes. The model uses Bayesian inference to learn the preferences of human users and custom deep adversaries, and is used on Dell login pages, globally. We were also granted a patent on this product.

- I built a behavior-based authentication model that classifies browser activity as being done by a human or by a bot. It was built using an auto-associative neural network, trained on browser activity biometrics such as keystrokes, click patterns and mouse dynamics, and is used on Dell login pages, globally. We were also granted a patent on this product.
- I created a traffic forecaster using a Recurrent Neural Network trained on activity logs, used to monitor product traffic and alert anomalous behavior, such as adversarial action (too many requests) or deteriorating product health (too few requests).

## Publications

### JOURNAL PAPERS

Shubha Guha, Falaah Arif Khan, Julia Stoyanovich and Sebastian Schelter, "Automated Data Cleaning Can Hurt Fairness in Machine Learning-based Decision Making," in IEEE Transactions on Knowledge and Data Engineering (2024)

Alene Kellogg Rhea, Kelsey Markey, Lauren D'Arinzo, Hilke Schellmann, Mona Sloane, Paul Squires, Falaah Arif Khan, and Julia Stoyanovich, "An External Stability Audit Framework to Test the Validity of Personality Prediction in AI Hiring." Data Mining and Knowledge Discovery, Special Issue on Bias and Fairness (2022)

### CONFERENCE PROCEEDINGS

Denys Herasymuk, Falaah Arif Khan, and Julia Stoyanovich. "Responsible Model Selection with Virny and VirnyView", International Conference on Management of Data (SIGMOD/PODS '24), Demo Track.

Ribeiro, Vitor, Eduardo HM Pena, Raphael Saldanha, Reza Akbarinia, Patrick Valduriez, Falaah Arif Khan, Julia Stoyanovich, and Fabio Porto. "Subset Modelling: A Domain Partitioning Strategy for Data-efficient Machine-Learning." Brazilian Symposium on Databases. (SBBSC 2023)

Shubha Guha, Falaah Arif Khan, Julia Stoyanovich, and Sebastian Schelter, "Automated Data Cleaning Can Hurt Fairness in Machine Learning-based Decision Making", 39th IEEE International Conference on Data Engineering (ICDE 2023)

Falaah Arif Khan, Eleni Manis, and Julia Stoyanovich, "Towards Substantive Conceptions of Algorithmic Fairness: Normative Guidance from Equal Opportunity Doctrines", 2nd ACM conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO 2022)

Falaah Arif Khan, Eleni Manis, and Julia Stoyanovich, "Translational tutorial: Fairness and Friends", 4th ACM Conference on Fairness, Accountability, and Transparency (FAccT 2021)

Falaah Arif Khan, Sajin Kunhambu and Kalyan Chakravarthy G. "Behavioral Biometrics and Machine Learning to Secure Website Logins". In: Security in Computing and Communications (SSCC 2018). Communications in Computer and Information Science, vol 969

### PEER-REVIEWED WORKSHOP PUBLICATIONS

Lucius E.J. Bynum, Falaah Arif Khan, Oleksandra Konopatska, Joshua Loftus and Julia Stoyanovich, "An Interactive Introduction to Causal Inference". VISxAI: Workshop on Visualization for AI Explainability, 2022

Falaah Arif Khan, Eleni Manis, and Julia Stoyanovich, "Fairness and Friends". Data, Responsibly Comic Series, Volume 2, 2021. Rethinking ML Papers Workshop, ICLR 2021 [Online]. Available: https://dataresponsibly.github.io/comics

Falaah Arif Khan and Abhishek Gupta. "Decoded Reality". Resistance AI Workshop, NeurIPS 2020 [Online]. Available: https://ai-ethics.github.io/decoded-reality/intro.html

Falaah Arif Khan and Julia Stoyanovich. "Mirror, Mirror". Data, Responsibly Comics, Volume 1, 2020. Resistance AI Workshop, NeurIPS 2020 [Online]. Available: https://dataresponsibly.github.io/comics/

### WORKING PAPERS

Falaah Arif Khan, Denys Herasymuk, Nazar Protsiv, and Julia Stoyanovich. "Still More Shades of Null: A Benchmark for Responsible Missing Value Imputation." arXiv preprint arXiv:2409.07510 (2024). (Under submission at VLDB 2025)

Falaah Arif Khan, Denys Herasymuk, and Julia Stoyanovich. "On Fairness and Stability: Is Estimator Variance a Friend or a Foe?." arXiv preprint arXiv:2302.04525 (2023).

Falaah Arif Khan, and Julia Stoyanovich. "The unbearable weight of massive privilege: Revisiting bias-variance trade-offs in the context of fair prediction." arXiv preprint arXiv:2302.08704 (2023).

PATENTS

Falaah Arif Khan and Hari Surender Sharma. Framework to Design Completely Automated Reverse Turing Tests. US Patent 16/828520, filed March 24, 2020 and US Patent 62/979500, filed February 21, 2020. Assigned to Dell Technologies Pvt Ltd

Falaah Arif Khan, Tousif Mohammed, Subham Gupta, Hung Dinh, Ramu Kannapan. Event-Based Search Engine, US Patent 16/752775, filed January 27, 2020. Assigned to Dell Technologies Pvt Ltd

Falaah Arif Khan, Sajin Kunhambu, Kalyan Chakravarthy G. Behavioral Biometrics and Machine Learning to secure Website Logins. US Patent 16/257650, filed January 25, 2019. Assigned to Dell Technologies Pvt Ltd

SCIENTIFIC COMICS AND POPULAR PRESS

Falaah Arif Khan and Zachary C. Lipton. "Superheroes of Deep Learning, Volume 2: Machine Learning for Healthcare" (2021) [Online]. Available: https://github.com/acmi-lab/superheroes-deep-learning/

Julia Stoyanovich and Falaah Arif Khan. We are AI Comic Series, Volume 1-5 (2021) [Online]. Available: https://dataresponsibly.github.io/we-are-ai/comics/

Falaah Arif Khan and Zachary C. Lipton. "Superheroes of Deep Learning, Volume 1: Machine Learning Yearning" (2020) [Online]. Available: https://github.com/acmi-lab/superheroes-deep-learning/

Falaah Arif Khan. "Meet AI" (2020), in AAAI Interactive Magazine [Online]. Available: https://interactiveaimag.org/columns/articles/comic/meet-ai-comic-vol-1-ed-2/

## Teaching Experience

| | |
|---|---|
| Spring'24 | **DS-UA 301: Independent Research in Data Science (Grader)**, New York University |
| Summer'23 | **We Are AI**, New York University |
| Spring '17 | **Mathematical Methods**, Shiv Nadar University |
| Fall '17 | **Digital Electronics**, Shiv Nadar University |
| Fall '17 | **Basics of Electrical and Electronic Circuits**, Shiv Nadar University |

## Awards & Fellowships

| | |
|---|---|
| 2024-26 | **Apple Scholars in AIML PhD Fellowship**, Apple Inc |
| 2021-26 | **Data Science PhD Fellowship**, New York University |
| Sept 2020 | **Innovation of the Month**, Government Technology and MetroLab Network. "Comic Book Bridges Gap Around Education in AI, Ethics" https://metrolabnetwork.org/projects/innovation-of-the-month/ |
| Aug 2019 | **Game Changer Award**, Office of the CIO, Dell EMC For "outstanding innovation in the security features of the Dell Access and Identity Management Product" |
| March 2019 | **Long Term Investment Award**, Dell EMC |
| 2014-18 | **Category 'A' Scholarship**, Shiv Nadar University 100% Scholarship/ complete waiver of tuition fees |

## Presentations, Workshops, Talks

INVITED TALKS

June 2022. *"It's funny because it's true - confronting scientific catechisms"*. D&I Talk, ACM DEBS 2022.

February 2022. *"Fairness as Equal Opportunity"*. Guest Lecture: Responsible Data Science, New York University.

April 2021. *"It's funny because it's true - confronting ML catechisms"*. Invited Talk: Rethinking ML Papers Workshop, ICLR'21.

March 2021. *"Fairness and Friends"*. Guest Lecture: Responsible Data Science, New York University.

Nov 2020. *"Ethics in AI"*. Panelist : Ethics in AI Panel by the McGill AI Society

Nov 2020. *"Decoded Reality: Ethics in AI"*. Invited talk: TechAide Montreal AI4Good Conference and Hackathon.

Nov 2020. *"Data, Responsibly"*. Invited speaker: Rutgers IIPL Algorithmic Justice Webinar.

## FACILITATED WORKSHOPS

Summer 2023. Co-facilitated the "We are AI" course to library staff at NYU.

June 2022. Ran the D&I Workshop at *IEEE-MDM 2022.* titled "Demystifying AI: Comics as a medium of scientific dissemination"

Oct 2021. Co-facilitated "Demystifying AI: Comics as a vehicle for public education" at *A Better Tech*.

June 2021. Co-facilitated "Demystifying AI: Comics as a vehicle for public education" at *RightsCon*.

March 2021. Ran a Community Lab at *MozFest* titled "Decoded Reality".

Dec 2020. Co-facilitated the "Perspectives on the Future of Responsible AI" workshop by *Montreal AI Ethics Institute and RAIN-Africa*.

Nov 2020. Co-facilitated the "Privacy in AI" workshop by *AI4Good and Montreal AI Ethics Institute*.

Nov 2020. Co-facilitated an invited session for the Future of Work and Disability Study group at *Inclusive Design Research Center @ OCAD University*.