

Swapneel Mehta

Curriculum Vitae

Jersey City, NJ — swapneel@bu.edu — (551) 328-7074 — mehtaver.se — LinkedIn — U.S. Permanent Resident

EMPLOYMENT

Postdoctoral Associate

2023–present

Boston University & Massachusetts Institute of Technology (Cambridge, MA)

Research on digital platforms and AI safety in **multi-turn** environments; developed **human-AI marketplaces** to test economic interventions in two-sided markets, limiting advertising deception. Deployed controlled online experiments with **4,000+** participants. Helped found and grow the Platform Governance Lab to 25+ members. Papers at Nature Comms. (R&R), ICIS; Invited Talks: ACR, SCP, WISE, IC2S2, Stanford TSRC, Columbia MarkTech, Yale AI/ML Conf., and industry (Bluesky, Google).

Co-founder and President

2021–present

SimPPL (New York, NY)

Built an applied research lab with 7 full-time engineers, supporting 60+ student researchers; secured \$2.5 million in grants/revenue from funders listed below. Built a platform used by 20+ partners to trace **civic discourse on the social internet, called Arbiter**. Built partnerships in 7 countries including UN agencies, with work resulting in 16+ workshop publications at ICML, NeurIPS, AAAI, ICWSM, IC2S2, dg.o, TPRC, Stanford TSRC. Selected as Google Research Innovator, Founder Fellow, Atlantic Dialogues Emerging Leader, ISPI Future Leader, ITS Rio Fellow and Belfer Fellow.

Co-founder and Chief Scientist

2023–2025

Sakhi (New York, NY)

MIT-incubated GenAI platform deploying IRB-approved **RCTs to improve reproductive health literacy piloted with 500 families (India, Bangladesh)**. Partnerships with Indian Council of Medical Research (Maternal and Child Health); conducting research with Cohere Labs, (frontier AI company with \$7B valuation) on **evaluation benchmarks for medical LLMs**, and **AI-generated answer rating platform for medical professionals**. Awards: AWS, MIT IDEAS, Delta-V, UNDP, invited to submit \$2 million Gates Foundation proposal.

Ph.D. Research (Graduate Researcher)

2021–2024

NYU Center for Social Media & Politics (New York, NY)

Research: Estimating the Causal effect of Reddit outages on low-quality news sharing; Identifying the Causal Effects of Twitter's warning labels on Trump's tweets. COVID infection modeling with probabilistic programs (MIT PROBPROG). Lead: NYU AI School (2020–2024); AI, Misinformation & Policy Seminar (2023–2024).

Technical Student (ML Research)

2018–2019

CERN, CMS Experiment (Geneva, Switzerland)

Research: Graph neural networks for particle track reconstruction at the LHC; productionized DeepJet ML framework; awarded by Google & Oxford MLHEP (2nd place in graph ML model training competition at a Ph.D. Summer School as an undergraduate).

EDUCATION

Ph.D. in Data Science

Years: 2019 - 2023

New York University, Center for Data Science

Thesis: *Towards Informed Interventions to Limit the Effects of Misleading Information on Social Networks*

Advisors: Richard Bonneau and Jonathan Nagler (Center for Social Media and Politics), Rajesh Ranganath.

Honors: Google Research Innovator; Google Research India and Wikimedia 'WikiCred' Grantee

B.Eng. in Computer Science

Years: 2014 - 2018

Mumbai University

Department of Computer Engineering

PUBLICATIONS

Under Editorial Review

- A. Nichols, N. Mažar, **S. Mehta**, T. Parker, G. Pennycook, D. Rand, M. Van Alstyne. *Certifiably True: The Impact of Self-Certification on Misinformation*. R&R at **Nature Communications** (2025).
- S. Mehta**, J. Bisbee, Z. Sanderson, R. Bonneau, J. Tucker, J. Nagler. *Identifying the Causal Effects of Twitter's Interventions on Trump's Tweets*. Working paper (target venue: **PNAS**, November 2025).
- G. Malpani, **S. Mehta**. *Who Trumps What: News 'Republication' on Truth Social*. Working Paper (2025).
- A.V. Singh, V. Dalal, **S. Mehta**. *Multi-agent Debates for Deliberative Content Moderation on X*. Working Paper (2025).
- D. Shah, S. Shetty, **S. Mehta**. *Estimating the Effects of Blocking on User Behavior*. Working Paper (2025).

Conference Proceedings

- S. Mehta**, A. Nichols, N. Mazar, M. van Alstyne. *Market Design Interventions for Safer Agentic AI*. **International Conference on Information Systems ICIS**(2025).
- C. Vergara, J. Lalwani, H. Ranka, N. Kothari, **S. Mehta**. *A Framework for Social Media Safety: Comparing Platform Interventions*. **Research Conference on Communications, Information and Internet Policy (TPRC)** (2025).
- C. Vergara, R. Jain, **S. Mehta**. *A History of Transparency Regulations*. **Annual International Conference on Digital Government Research (dg.O)** (2024).
- S. Mehta**, S. Sarkhel, X. Chen, S. Mitra, V. Swaminathan, R. Rossi, A. Aminian, H. Guo, K. Garg. *Open-Domain Trending Hashtag Recommendation for Videos*. **IEEE International Symposium on Multimedia (ISM)** (2021).
- S. Mehta**, N. Kasmanoff. *Covid-19 Modeling and Control via Policy Interventions*. International Conference on Probabilistic Programming (PROBPROG) (2021).
- S. Mehta**, C. Raman, N. Ayer, S. Sahasrabudhe. *Auto-Grading for 3D Modeling Assignments in MOOCs*. IEEE International Conference on Advanced Learning Technologies (ICALT) (2018).
- S. Mehta**, P. Kothuri, D.L. Garcia. *A Big Data Architecture for Log Data Storage and Analysis*. **Best Paper (ML Track)**. International Conference on Integrated Intelligence and Communication Systems (2018).

CONFERENCE PRESENTATIONS

Market Design Interventions for Safer Agentic AI (2025)

Invited Talks: Workshop on Information Systems and Economics (WISE), International Conference on Computational Social Science (IC2S2), Columbia MarkTech Conference, Yale AI/ML & BA Conference.

Multi-agent Debates for Deliberative Content Moderation on X (2025)

Poster: Stanford Trust and Safety Research (TSR) Conference.

Who Trumps What: News 'Republication' on Truth Social (2025)

Invited Talk: Stanford TSR Conference.

Certifiably True: The Impact of Self-Certification on Misinformation (2025)

Invited Talk: Association for Consumer Research, Society for Consumer Psychology.

Truth Warrants Increase Welfare and Accelerate Sales in Digital Markets. (2024-25)

Invited Talks: Platform Strategy Conference, Workshop on Information Systems and Economics (WISE). Posters: IC2S2, AMA Marketing and Public Policy Conference.

Examining the Implications of Deepfakes for Election Integrity (2024)

Invited Talk, AAAI AI4CE.

How Decentralization Affects User Agency on Social Platforms (2024)

Invited Talk, ICWSM Data Challenge.

Can Social Media Platforms Transcend Political Labels? An Analysis of Neutral Conservations on Truth Social (2024)

Invited Talk, DARE Workshop, ICWSM.

Identifying the Causal Effects of Twitter's Interventions on Trump's Tweets (2023)

Invited: Stanford TSR Conference, Twitter (Cortex), Meta (Probability), McKinsey (Quantum Black).

Estimating the Impact of Coordinated Inauthentic Behavior on Recommendations (2022)

Invited Talk, ICML AI4ABM.

Expanding Access to ML Research through Student-led Collaboratives (2022)

Poster: NeurIPS Workshop on Broadening Research Collaborations.

AWARDS AND HONORS

Travel & Recognition

2018–2025

TrustCon (2025); ICML (2025); WSDM Smart City Challenge (2023); AI2Amplify AI4Good (2022); MIDAS Future Leaders Forum (2022); IRIS-HEP Fellowship (2020); Oxford MLHEP Challenge (2018); **Economic Times: India's Top 33 Engineers** (2018).

Media Features

2018–2024

UNESCO: Tackling Disinfo. (2024); Deutsche Welle: Media Literacy (2024); Rest of World (2024); Boston University news (2023/24); Prothom Alo (2024); NYU: AI School (2024); Jagran (2023); Google Research (2023); Open Source for U (2018).

GRANTS AND PROJECT FUNDING (\$2.5 million)

- **Deploying SimPPL's research platform Arbiter, and AI Training for Journalists in Kenyan Newsrooms** (2025–2026). Funder: Deutsche Welle; Institutional Location: Germany/USA; Amount: **\$25,200**.
- **How do Investors Value Responsible Tech Practices Developed at AI Startups?** (with Prof. H. Kim, INSEAD/Harvard AI Venture Lab) (2025–2027). Funders: Ford Foundation; Omidyar Network; Institutional Location: USA/Europe; Amount: **\$150,000**.
- **Designing Decision-support Systems for Child Trafficking Prevention using Multimodal AI Systems** (with Dr. E.A. Rahman, Harvard, University of Mannheim) (2025–2031). Funder: Baden-Württemberg Stiftung; Institutional Location: Germany/USA; Amount: **\$2.1 million**.
- **Google (exploreCSR)** (with Prof. P. Bari) (2024). Funder: Google Research; Institutional Location: India/USA; Amount: **\$75,000**.
- **MIT DeltaV (Flagship Accelerator at MIT)** (2024). Funder: MIT; Institutional Location: Cambridge, MA; Amount: **\$20,000**.
- **MIT PKG IDEAS Challenge** (2024). Funder: MIT PKG Center; Institutional Location: Cambridge, MA; Amount: **\$13,500**.
- **Google PaliGemma Award** (2024). Funder: Google; Institutional Location: USA; Amount: **\$5,000**.
- **Google (exploreCSR)** (with Prof. P. Bari) (2023). Funder: Google Research; Institutional Location: India/USA; Amount: **\$32,000**.
- **Mozilla Responsible Computing Challenge** (2023). Funder: Mozilla Foundation; Institutional Location: India; Amount: **\$25,000**.
- **Belfer Fellowship** (2023). Funder: ADL Center for Technology & Society; Institutional Location: USA; Amount: **\$40,000**.
- **AI2Amplify Fellowship: WhatsApp based Health Literacy Interventions for Rural Audiences** (2023). Funder: Goethe Institut; Institutional Location: Germany; Amount: **€14,000**.
- **WikiCred: Arbiter, a platform for Tracing Digital Narratives** (2023). Funder: Wikimedia Foundation; Institutional Location: USA; Amount: **\$10,000**.
- **Google Cloud Research Credits** (2022). Funder: Google Cloud; Institutional Location: USA; Amount: **\$9,000**.
- **Amazon (AWS) Research Credits** (2022). Funder: Amazon Web Services; Institutional Location: USA; Amount: **\$5,000**.
- **Detecting Coordinated Inauthentic Networks Promoting Russian State-sponsored Media Ar-**

ticles on X (subgrant from The Sunday Times) (2022). Funder: JournalismAI; Institutional Location: UK; Amount: \$4,000.

- **Towards Probabilistic Programming Methods to Simulate Social Networks (PI Prof. Philip Torr)** (2021). Funder: UK Research & Innovation; Institutional Location: UK; Amount: £23,000.
- **NYC Media Lab AI + Local News Challenge** (2021). Funder: NYC Media Lab; Institutional Location: New York, NY; Amount: \$7,500.
- **Machine Learning Summer School for Undergraduate Students** (2021). Funder: Google Research India; Institutional Location: India; Amount: \$1,100.

FELLOWSHIPS (Selected)

ITS Rio Global Policy Fellow (2025, deferred); ISPI Future Leader (2025, deferred); Community Advisory Board, Integrity Institute (2024); Atlantic Dialogues Emerging Leader, Policy Center for the New South (2025); Responsible AI Affiliate, All Tech Is Human (2024); Google Research Innovator (2023); CTS Belfer Fellow (2023); Future Founders (Startup Bootcamp; Fellowship—declined).

INVITED TALKS

Guest Lectures: University of Arkansas with Prof. Alex Rieger in Information Systems (Rebuilding Digital Trust); Meredith College with Prof. Aaron Nichols (Real talk about Artificial Intelligence); (upcoming) Columbia University with Prof. Alan Zhang (Social Entrepreneurship)

NYU Data Science; Aethos Lunch Seminar; Global Media Forum; Harvard BKC Spring Synthesizer; IC2S2 (AI & Economics); TrustCon (Global Majority); Microsoft NERD (LLMs); MIT Decentralized AI; Prince Mahidol Awards Conference; Google (Monetization Policy)

2025

Institutional Location: Various (USA, Europe, Asia)

WISE; Yale AI/ML; MarkTech (Columbia); MIT Decentralized AI Summit; Stanford TSRC; TrustCon (AI + Elections); IC2S2 (Oral); ATIH @ Embassy of Finland; UNESCO World Press Freedom Day; DW (Online Trust)

2024

Institutional Location: Various

DW Innovation Days; IIT Madras Data Science; SimPPL Fellowships; Mandiant mWise (Google); McKinsey Guest Talk; TrustCon (Community); MIDAS Future Leaders (Responsible AI); CCNY SASE “TED Talks”; MozFest; NYU Data Science Seminar; Deutsche Welle (multiple); NYC Media Lab

2023

Institutional Location: Various

Graphika; Truth & Trust Online Conf.; Twitter Health ML; MIT Media Lab; Stanford T&S Research Conf.; NYU Graduate Seminar; DEFCON Misinfo. Village; Twitter Cortex; Facebook Probability; Flatiron Computational Biology Group; NYC AI Workshop; NYU Data Science; Mumbai University

2021–2022

Institutional Location: Various

CONFERENCE ACTIVITY

Panels Organized

Global Majority Perspectives on Big Tech Trust and Safety Issues

2025

Conference: TrustCon 2025

Responsible Innovation Day (Mozilla RCC)

2024

Local event at DJSCE College of Engineering, India

NeurIPS MASEC Workshop

2023

Name of Conference: Conference on Neural Information Processing Systems

Misinformation Village

2023

Conference: DEFCON

AI, Misinformation & Policy Seminar

2023

Institute: NYU Center for Data Science

AI for Everyone Workshop

2022

Conference: Columbia University Computation + Journalism

ICLR AI4ABM Workshop

2023

Name of Conference: International Conference on Learning Representations

TEACHING EXPERIENCE

Recent teaching evaluations available on SimPPL Testimonials. Others available via anonymous forms for NYU AI School, Unicode ML Summer Course on request.

SimPPL NextGenAI Program (Lead Instructor; funded by Mozilla RCC) 2024-2025
12-month responsible AI product development program for 34 CS/IT undergraduates resulting in 6 product launches, 3 commercial contracts entirely bootstrapped.

SimPPL Fellowships Research Program (Instructor; funded by Google Research) 2023-2024
8-month research mentorship program for 35 students resulting in students submitting ICWSM, AAAI, NeurIPS, ICML workshop papers.

Unicode ML Summer School (Instructor; funded by Google Research) Summer 2021
Unicode (student-led FOSS collaborative)
Curriculum design and lectures on ML fundamentals and applications.

NYU AI School (Lead Organizer and Instructor; funded by Deepmind and Genentech) 2020-2024
New York University
Annual week-long program democratizing AI education for STEM and non-STEM undergraduate students belonging to historically marginalized groups in the greater NYC area; conceptualization, speaker curation, leading tutorials, logistics, fundraising.

RESEARCH INTERNSHIPS

Product Data Science Intern

Summer 2023

Slack, Central Data Science (New York, NY)

Identified *social* learning drivers of collaborative feature adoption for Slack.

Machine Learning Engineer (Intern)

Summer 2022

Twitter (X), Civic Integrity (New York, NY)

Improved precision of civic/health misinformation classifier by **20%**; scaled to 1M+ tweets/week (BigQuery ML; NLP; recommender systems).

Visiting Researcher

Fall 2021-2022

University of Oxford, Torr Vision Group (Oxford, UK)

Generative modeling of adversarial agent behaviors on Reddit recommendations; presented at ICML (AI4ABM)

and DEFCON Misinfo. Village (2022).

Data Science Research Intern

Summer, Fall 2020

Adobe Research (New York, NY)

Open-domain trending hashtag recommendation for videos; multimodal GNNs, weak supervision, LLMs. First-author research paper IEEE ISM; **U.S. Patent 12050647**.

Research Intern

Fall 2016–2017

Microsoft Research (IIT-B)

Prof. S. Sahasrabudhe: Automated grading for Blender 3D modeling assignments for **5,000** learners (ICALT 2018); edX Prize Finalist (2019). Microsoft Research analytics for Lokacart (D2C farmer app).

Machine Learning Software Engineer (Intern)

Fall 2016–Spring 2017

Smokescreen (acq. Zscaler) (New York, NY)

Reduced footprint of network scans using OS hardening cybersecurity practices. Conducted Markov-chain host-name generation for decoys. Automated deployment of NASSCOM-DSCI awarded software.

SERVICE TO PROFESSION

Journal Manuscript & Program Review

2019–2025

Nature Patterns Journal (2025); IC2S2 (2025); NSF (2024); IEEE DSAA MLJ (2024); Springer AI & Ethics (2024); ACM CHI (2024); EPJ Data Science (2024); Springer Nature ML (2024); NeurIPS Workshops (2023); AABI (2023); ICLR AI4ABM (2023); AI+Local News Challenge (2023); ICML AI4ABM (2022); Elsevier JPDC (2019).

Leadership / Organizing

2021–2025

Responsible Innovation Day (Mozilla RCC); SimPPL Fellowships; NeurIPS MASec; AI, Misinformation & Policy Seminar; ICLR AI4ABM; Misinfo. Village @ DEFCON; NYU AI School (2021–2024); Unicode ML Summer School (2021).

SERVICE TO DEPARTMENT

Platform Governance Lab (featured by Digital Business Institute, Boston University)

truthmarket.com — Community building, mentorship of 36+ undergraduate students after interviewing 77 applicants over 2 years at BU. Software infrastructure, tutorials, and documentation development for creating an open-source behavioral experimentation platform to enable the department to evangelize and scale our research output. Developed collaborations with top technology companies to enable institutional data access for Meta and receive letter of support for research deployment from Bluesky.

Board of Studies Member at DJSCE, for Depts. of Data Science, AI, and Computer Engineering

D.J. Sanghvi College of Engineering Determine the syllabus and scope of educational programs including course credit allocation, electives to be offered, and alignment of curriculum with industry requirements for increasing student employability. Invited to Boards of three different departments, first alumnus to be included in this capacity in the history of my alma mater. Helped fundraise to win grants from Google and Mozilla and integrate responsible computing into course curriculum through SimPPL's NextGenAI Program, for accredited course titled "Innovative Product Development".

PROFESSIONAL SKILLS

Programming/ML: Python, PyTorch, R, Java, C++, SQL/MySQL.

Data/Infra: GCP (BigQuery), AWS, Unix/Linux, Bash, Docker, Git, SLURM, LAMP.

Web: HTML, CSS, JavaScript.

PROFESSIONAL MEMBERSHIPS

Integrity Institute

Community Advisory Board (2024).

All Tech Is Human

Responsible AI Affiliate (2024).

REFERENCES

Available upon request.