

NEW YORK UNIVERSITY
PH.D. IN DATA SCIENCE
2019 - 2023 (4 YRS.)
GPA: 3.8/4

MUMBAI UNIVERSITY
BENG IN COMPUTER SCIENCE
2014 - 2018
GPA: 9.23/10

COURSEWORK

DOCTORAL

Machine Learning
Causal Inference
Inference and Representation
Math Tools for Data Science
Unicode ML Summer School
(Instructor, funded by Google Research)

UNDERGRADUATE

Artificial Intelligence
Big Data Analysis
Operating Systems
Database Management Systems
Computer Networks
Unicode
(Co-founder, FOSS Training Program)

TRAVEL AWARDS

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

SKILLS

Python, Pytorch • MySQL • Bash • R
GCP • BigQuery • Unix/Linux • AWS
Docker • git • LAMP • SLURM
HTML/CSS • Javascript • C++ • Java

LINKS

Github:// [swapneelm](#)
Twitter:// [@swapneel_mehta](#)
YouTube:// [swapneel_mehta](#)

BOSTON UNIVERSITY AND MIT | POSTDOCTORAL ASSOCIATE
Oct 2023 - Present | Cambridge, MA

- Researching AI safety in **multi-turn** environments. Developed **human-AI marketplaces** to test economic interventions to limit advertising deception in simultaneous controlled online experiments with ≥ 4000 human participants.
- Helped found and grow the Platform Governance Lab to 28 members.
- Publications: ICIS, ACR, SCP, WISE, IC2S2, Stanford TSRC, Columbia MarkTech, Yale AI/ML Conf. and tech companies Bluesky, Google

SIMPPL | CO-FOUNDER AND PRESIDENT
Dec 2021 - Present | New York, NY

- Built a social sensemaking platform to **improve trust on the social internet** with the UN, New York Public Radio, Deutsche Welle, Jagran News leading team of 12 engg. staff, winning over \$4 million in grants and revenue from Harvard, AWS, Google, Wikimedia Foundation, Mozilla, Goethe Institut.
- **16+ workshop and conference publications** at ICML, NeurIPS, AAAI, ICWSM, IC2S2, dg.o, Stanford TSRC.
- Selected Google Research Innovator, Founder Fellow, Belfer Fellow .

SAKHI | CO-FOUNDER AND CHIEF SCIENTIST
July 2023 - May 2025 | New York, NY

- MIT and UNICEF-incubated tech venture **improving women's health literacy** through conversational agents for **health literacy** on WhatsApp.
- Launched **pilots** with 500 families in India and Bangladesh.
- Won \$62.5K in grants from AWS, MIT IDEAS, Delta-V, UNICEF, UNDP.
- Partnering with Indian Council of Medical Research, Cohere AI, for health LLM evaluations and field deployment.

NYU CENTER FOR SOCIAL MEDIA & POLITICS | PH.D. RESEARCH
Feb 2021 – 2024 | New York, NY

- Thesis: **Towards Informed Interventions to Limit the Effects of Misleading Information on Social Networks.**
- Identify the causal effect of Reddit's outages on low quality news sharing.
- Identify the causal impact of Twitter's warning labels on Trump's tweets.
- COVID infection modeling with probabilistic programs at **MIT PROBPROG**.
- Lead: **NYU AI School**, 2020-24; **AI, Misinfo., & Policy Seminar** 2023-24.

CERN, CMS EXPERIMENT | TECHNICAL STUDENT (ML RESEARCH)
March 2018 - July 2019 | Geneva, Switzerland

- Developed graph neural networks for particle track reconstruction at the LHC
- Productionized the **DeepJet** ML framework; awarded by **Google, Oxford**.

INTERNSHIPS

SLACK, CENTRAL DS | PRODUCT DATA SCIENCE INTERN
May 2023 – Aug 2023 | New York, NY

- Identified 'social' learning drivers of feature adoption at Slack (NDA).

TWITTER, CIVIC INTEGRITY | MACHINE LEARNING ENGINEER
June 2022 – Aug 2022 | New York, NY

- Improved the precision of their civic and health misinfo. classifier (BigQuery ML) by **20%** scoring 1M+ tweets / week. NLP; recommender systems.

REVIEWER

2025	Nature Patterns Journal
2025	IC2S2
2024	NSF
2024	IEEE DSAA MLJ
2024	Springer AI and Ethics
2024	ACM CHI
2024	EPJ Data Science
2024	Springer Nature ML
2023	NeurIPS Workshops
2023	AABI
2023	ICLR, AI4ABM Workshop
2023	AI+Local News Challenge
2022	ICML, AI4ABM Workshop
2022	NeurIPS, WBRC
2019	Elsevier JPDC

TALKS

2025	Aethos Lunch Seminar
2025	Global Media Forum
2025	Harvard (BKC Spring Synth.)
2025	IC2S2 (AI and Economics)
2025	TrustCon (Global Majority)
2025	Microsoft NERD (LLMs)
2025	MIT Decentralized AI
2025	Prince Mahidol Awards Conf.
2025	Google (Monetization Policy)
2024	WISE Conf.
2024	Yale AI/ML Conf.
2024	MarkTech Columbia
2024	MIT Decentralized AI Summit
2024	Stanford (TSRC)
2024	TrustCon (AI + Elections)
2024	IC2S2 (Oral)
2024	ATIH, Embassy of Finland
2024	UNESCO, DW Akademie
2024	DW (Online Trust)
2023	DW Innovation Days
2023	IIT Madras Data Science
2023	SimPPL Fellowships
2023	Mandiant mWise (Google)
2023	McKinsey Guest Talk
2023	TrustCon (Community Building)
2023	MIDAS Future Leaders Summit (Responsible AI)
2023	CCNY SASE "TED Talks"
2023	MozFest by Mozilla
2023	NYU Data Science Seminar
2023	Deutsche Welle (multiple)
2023	NYC Media Lab
2022	Graphika
2022	Truth and Trust Online Conf.
2022	Twitter, Health ML
2022	MIT Media Lab
2022	Stanford, T&S Research Conf.
2022	NYU Graduate Seminar
2022	Misinfo. Village, DEFCON

OXFORD, TORR VISION GROUP | VISITING RESEARCHER

Aug 2021 – June 2022 | Oxford, UK

- Generative modeling of adversarial agent behaviors on Reddit recommendations presented in workshops at ICML and DEFCON, 2022.

ADOBE RESEARCH | DATA SCIENCE RESEARCH INTERN

May 2020 – Dec 2020 | New York, NY

- Trending Hashtag Recommendation for Videos; large language models, data mining, weakly-supervised learning, multimodal graph neural networks.
- First-author paper accepted at IEEE ISM; U.S. Patent 12050647.

MICROSOFT RESEARCH (IIT-B) | RESEARCH INTERN

May 2016 – May 2017 | New York, NY

- Prof. S. Sahasrabudhe, Automated evaluation of Blender 3D Modeling assignments (ICALT 2018) for 5000 learners; edX Prize Finalist, 2019.
- Microsoft Research: Google Analytics for Lokacart D2C farmer app.

SMOKESCREEN (ACQ. ZSCALER) | ML INTERN

Nov 2016 – March 2017 | New York, NY

- Reducing the footprint of network scans and building Markov-chain based hostname generation for intelligent decoys.
- Automated deployment of NASSCOM-DSCI awarded software.

FELLOWSHIPS

2025	ITS Rio Fellow (Brazil)
2025	ISPI Future Leader (Italy)
2024	Community Advisory Board, Integrity Institute
2024	Responsible AI Affiliate, All Tech is Human
2023	Future Founders Fellowship (declined)
2023	Google Research Innovator
2023	Future Founders Startup Bootcamp
2023	CTS Belfer Fellow

AWARDS AND GRANTS

2025	(final review) AI x Mental Health, with Prof. V. Patel (Harvard)	\$ 1.96M
2025	Ford Fdn., Omidyar, with Prof. H. Kim (INSEAD, Harvard)	\$ 150,000
2025	BW Stiftung, with Dr. E. A. Rahman (Harvard)	€ 1.8M
2024	Google (exploreCSR), with Prof. Pranit Bari	\$ 75,000
2024	MIT Delta V Accelerator	\$ 20,000
2024	MIT PKG IDEAS Challenge	\$ 13,500
2024	FTI: People to Watch in AI and Local News	
2024	Google PaliGemma Award	\$ 5,000
2023	Google (exploreCSR), with Prof. Pranit Bari	\$ 32,000
2023	Mozilla (Responsible Computing Challenge)	\$ 25,000
2023	Algovera AI	\$ 3,000
2023	NYUTech Venture Workshop	\$ 1,000
2023	Belfer Fellowship	\$ 40,000
2023	Goethe Institut	€ 14,000
2023	Wikimedia Foundation 'WikiCred' Grant	\$ 10,000
2023	Deutsche Welle	€ 4,950
2022	Google Cloud Research Credit Award	\$ 9,000
2022	Amazon (AWS) Research Credit Award	\$ 5,000
2022	JournalismAI, The Sunday Times	\$ 4,000
2021	UK Research & Innovation Grant (w/ Oxford TVG)	£ 23,000
2021	NYC Media Lab AI + Local News Challenge	\$ 7,500
2021	Google Research India	\$ 1,100

TALKS

- 2021 Twitter, Cortex
- 2021 Facebook, Probability
- 2021 Flatiron Comp. Bio. Group
- 2020 NYC AI Workshop
- 2020 NYU Data Science
- 2018 Mumbai University (multiple)

ORGANIZING

- 2024 Responsible Innovation Day
- 2024 SimPPL Fellowships
- 2023 NeurIPS MASec Workshop
- 2023 AI, Misinfo. & Policy Seminar
- 2023 ICLR AI4ABM Workshop
- 2023 Misinfo. Village, DEFCON
- 2023 NYU AI School 2023
- 2022 NeurIPS WBRC
- 2022 NYU AI School 2022
- 2022 AI for Everyone, C+J 2022
- 2021 NYU AI School 2021
- 2021 Unicode ML Summer School

MEDIA

- 2024 UNESCO: Tackling Disinfo.
- 2024 DW: Media Literacy
- 2024 Rest of World
- 2024 Boston University
- 2024 Boston University
- 2024 Prothom Alo (Bangladesh)
- 2024 NYU: AI School
- 2023 Jagran News (India)
- 2023 Google Research
- 2018 Open Source for U

PUBLICATIONS

- [1] A. Nichols, N. Mažar, S. Mehta, T. Parker, G. Pennycook, D. Rand, and M. Van Alstyne. '**Certifiably True: The Impact of Self-Certification on Misinformation**'. *Major Revision at Nature Communications*, 2025
- [2] S. Mehta, A. Nichols, N. Mazar, M. van Alstyne. '**Market Design Interventions for Safer Agentic AI**'. *Conditional Accept, ICIS 2025*; presented at IC2S2 2025; Columbia MarkTech 2024; Yale AI, ML, and Business Analytics Conf., 2024.

WORKING PAPERS

- [3] S. Mehta, Z. Yang, C. Fan, D. Liu, N. Mazar, M. van Alstyne. '**Truth Warrants Increase Welfare and Accelerate Sales in Digital Markets**'. *Working Paper; Lightning Talk, Platform Strategy Conf. 2025; Poster at AMA Marketing and Public Policy Conf., 2025; Workshop on Information Systems and Economics, 2024.*
- [4] G. Malpani, S. Mehta. '**Influence Outshines Credibility: Unraveling the Amplification of News on Truth Social**'. *under review, 2025; Invited Talk at Stanford T&S Research Conf. 2025.*
- [5] A.V. Singh, V. Dalal, S. Mehta. '**Multi-agent Debates for Deliberative Content Moderation on X**'. *under review, 2025; Poster at Stanford T&S Research Conf. 2025.*
- [6] D. Shah, S. Shetty, S. Mehta. '**Estimating the Effects of Blocking on User Behavior**'. *under review, 2025; Poster at Stanford T&S Research Conf. 2024 and ICWSM Data Challenge, 2023.*
- [7] S. Mehta, J. Bisbee, Z. Sanderson, R. Bonneau, J. Tucker, and J. Nagler. '**Identifying the Causal Effects of Twitter's Interventions on Trump's Tweets**'. *Working Paper; Invited Talk, Stanford T&S Research Conf.*
- [8] S. Mehta, K. Aslett, Z. Sanderson, R. Bonneau, J. Tucker, and J. Nagler. '**Platform Outages Increase the Sharing of High-quality News on Reddit**'. *Working Paper, 2024.*

CONFERENCE PAPERS

- [9] C. Vergara, J. Lalwani, H. Ranka, N. Kothari, and S. Mehta. '**A Framework for Social Media Safety: Comparing Platform Interventions**'. *Research Conference on Communications, Information, and Internet Policy (TPRC), 2025..*
- [10] C. Vergara, R. Jain, and S. Mehta. '**A history of transparency regulations: Interdisciplinary strategies for shaping social media regulation and self-governance**'. *Intl. Conference on Digital Governance (dg.O), 2024.*
- [11] S. Mehta, S. Sarkhel, X. Chen, S. Mitra, V. Swaminathan, R. Rossi, A. Aminian, H. Guo, and K. Garg. '**Open-Domain Trending Hashtag Recommendation for Videos**'. *In IEEE International Symposium on Multimedia (ISM), 2021.*
- [12] S. Mehta, and N. Kasmanoff. '**Covid-19 Modeling and Control via Policy Interventions**'. *In International Conference on Probabilistic Programming (PROBPROG), 2021.*
- [13] S. Mehta, C. Raman, N. Ayer, and S. Sahasrabudhe. '**Auto-Grading for 3D Modeling Assignments in MOOCs**'. *Oral, International Conference on Advanced Learning Technologies (ICALT), 51–53. IEEE, 2018.*
- [14] S. Mehta, P. Kothuri, and D.L. Garcia. '**A Big Data Architecture for Log Data Storage and Analysis**'. *Best Paper, ML Track. Integrated Intelligent Computing, Communication and Security, 201–9. Springer, Singapore, 2018.*

WORKSHOP PAPERS

- [17] H. Ranka, M. Surana, N. Kothari, V. Pariawala, P. Banerjee, A. Surve, S.R. Sankepally, R. Jain, J. Lalwani, S. Mehta. ‘**Examining the Implications of Deepfakes for Election Integrity**’. *Oral, AAAI Workshop on AI for Credible Elections*, 2024.
- [18] A. Surve, A. Shamraj, and S. Mehta. ‘**How Decentralization Affects User Agency on Social Platforms**’. *Oral, Data Challenge, AAAI Intl. Conference on Web and Social Media (ICWSM)*, 2024.
- [19] C. Shah, R. Konka, G. Malpani, S. Mehta, and L. Ng. ‘**Can Social Media Platforms Transcend Political Labels? An Analysis of Neutral Conservations on Truth Social**’. *Oral, DARE Workshop, ICWSM*, 2024.
- [20] A. Surve, A. Rathod, M. Surana, G. Malpani, A. Shamraj, S.R. Sankepally, R. Jain, and S. Mehta. ‘**Multiagent Simulators for Social Networks**’. *Oral, Multi-Agent Security Workshop, NeurIPS 2023*, 2023.
- [21] S. Mehta, A.G. Baydin, R. Bonneau, J. Nagler, and P. Torr. ‘**Estimating the Impact of Coordinated Inauthentic Behavior on Content Recommendations in Social Networks**’. *Oral, Workshop AI for Agent-Based Modelling, ICML 2022*.
- [22] D. Gandhi, R. Jain, J. Gala, J. Lalwani, and S. Mehta. ‘**Expanding Access to ML Research through Student-Led Collaboratives**’. *Poster, Workshop on Broadening Research Collaborations, NeurIPS 2022*.
- [23] S. Mehta, P. Kothuri, and D.L. Garcia. ‘**Anomaly Detection for Network Connection Logs**’. arXiv Preprint arXiv:1812.01941, 2018.

TECH POLICY REPORTS

- [25] A. Sharma, A. Shetty, M. Surana, and S. Mehta. ‘**Harassment in the Digital Arena: Analyzing Online Abuse Against U.S. Politicians**’. *Report, ADL Center for Tech and Society*, 2024.
- [26] S.W. Baksh, D. Mungra, N. Ahsan, V. Pariawala, R. Jain, A. Surve, P. Banerjee, S.M. Eron, S. Mehta, S.R. Diya. ‘**How Facebook Has Become A Political Battleground In Bangladesh**’. *Report (Tech Global Institute)*, 2024.
- [27] S.W. Baksh, D. Mungra, V. Pariawala, R. Jain, A. Surve, A. Das, M. Biswas, F. Afroz, S.M. Eron, S. Mahmood, S. Mehta, S.R. Diya. ‘**From homophobia to assault: The gendered landscape of Bangladesh’s political disinformation**’. *Report (Tech Global Institute)*, 2024; Meta issued takedown.
- [28] H. Ross, B. Silverman, J. Allen, S. Mehta, et al. ‘**Shining a Light on Platform Transparency Best Practices**’. *Report (Integrity Institute)*, 2023.
- [29] G. Ellingson, K. Harbath, J. Allen, Eric Davis, A. Leavitt, S. Mehta, et al. ‘**Defining and Achieving Success in Elections Integrity**’. *Report (Integrity Institute)*, 2023.