

# Sanjana Gautam

RESEARCH SCIENTIST

☎ (+1) 623-308-7723 | ✉ sanjanagautam96@gmail.com | 🏠 sanjanagautam.com | 🌐 sanjanagautam | 📄 Sanjana Gautam

## Education

### Pennsylvania State University

PHD IN INFORMATICS, UNDER THE ADVISEMENT OF DR. MARY BETH ROSSON

University Park, PA.

2019 - Present

### Shiv Nadar University

B.TECH. IN COMPUTER SCIENCE AND ENGINEERING

Greater Noida, UP.

2015 - 2019

## Research Experience

### HLT Lab, Penn State

RESEARCH ASSISTANT

University Park, PA.

July 2022 - Present

- Designed studies focused on ethical AI practices, **investigating biases in generative large language models** (LLM) such as GPT-4, Llama.
- Developed methods to identify and mitigate nationality bias using zero-shot, few-shot, and chain-of-thought prompting and instruction tuning approaches.
- Performed quantitative and qualitative evaluation of nationality bias in LLMs.

### Tech Policy Lab, University of Washington

RESEARCH ASSISTANT

Seattle, WA.

Oct 2023 - July 2024

- Created **novel AI harms framework identifying cultural harms** within language and vision GenAI models.
- Designed and thematically analyzed focus group and annotation study with Global South population to identify perceived bias and harms.
- Evaluated quantitatively and qualitatively various LLM cultural bias alignment approaches.

### PennHCI Lab, University of Pennsylvania

RESEARCH ASSISTANT

Philadelphia, PA.

June 2023 - Jan 2024

- Performed **LLM auditing for bias and discrimination identification** in script generation using various prompting techniques.
- Implemented data pipeline to collect over 17,000+ user labelled TikTok videos
- Performed data cleaning, topic modelling, sentiment analysis on the data collected
- Statistically analysed implicit and explicit nature of political data on TikTok using the data collected

### Dataminr

RESEARCH INTERN, HUMAN AI INNOVATION

New York City, NY (Remote)

Sept 2022 - Dec 2022

- Designed and executed three **studies for human-in-the-loop AI design**: exploratory card sorting, dairy study and focus group study
- Communicated insights and findings with the stakeholders, aligning with strategic business objectives
- Developed comprehensive needs assessment reports to inform key product development decisions, after collaboration with designers

### Dataminr

RESEARCH INTERN, AI ENGINEERING

New York City, NY (Remote)

Sept 2021 - Dec 2021

- Qualitative researcher in **the AI Engineering team, driving critical research initiatives and delivering valuable insights**
- Designed, executed, and analyzed semi-structured interviews with diverse stakeholders
- Designed survey, collected data and employing advanced analytical techniques to summarize results and uncover meaningful patterns
- Extracted performance insights to optimize workflow

### Crowd AI Lab, Penn State

RESEARCH ASSISTANT

University Park, PA.

May 2023 - Present

- Investigated a AI tool to support creative writing
- Analysed user interviews to capture user feedback on creative writing assistive tool
- Performed thematic analysis on the interview data and extracted design insights

### Collaboration and Innovation Lab, Penn State

RESEARCHER

University Park, PA.

Aug 2019 - Present

- Conceived and executed end-to-end research project for peer-based adaptive learning from scoping phase, participant recruitment and designing study plans to method creation and analysis
- Lead user initiative to understand user requirements and inform product development strategies for a cross-functional team
- Evaluated design implications that originate from a geographically distributed community

### Goldman Sachs

TECHNOLOGY STRATS INTERN, CORPORATE CASH MANAGEMENT

Bengaluru, India

Jan 2019 - May 2019

- Spearheaded the development of a cutting-edge, web-based microservices-driven payment control portal to replace legacy platform.
- Designed and developed an intuitive user interface for the web-based microservice, ensuring seamless functionality and an exceptional user experience.

## Skills

---

- **Programming Languages/Tools** : Python; R; sklearn; PyTorch; TensorFlow; JAVA;
- **Statistical Softwares** : SPSS; Stata
- **Research Methods** : qualitative research methods; quantitative research methods; prompting technique; topic modelling; model training; usability testing; qualitative research; interview; focus group; diary study; survey design; statistical analysis; adobe suite

## Honors & Awards

---

2024	<b>Logic(s) Liberatory Tech Fellowship</b> , Pennsylvania State University	<i>State College</i>
2023	<b>IST Graduate Student Travel Grant</b> , Pennsylvania State University	<i>State College</i>
2023	<b>Nittany AI Challenge 2023 Semi-Finalist (ChronoNews)</b> , Pennsylvania State University	<i>State College</i>
2023	<b>Nittany AI Challenge 2023 Semi-Finalist (ClassCollab)</b> , Pennsylvania State University	<i>State College</i>
2020	<b>Robert Graham Endowment Graduate Fellow</b> , Pennsylvania State University	<i>State College</i>
2019	<b>Robert Graham Endowment Graduate Fellow</b> , Pennsylvania State University	<i>State College</i>
2015	<b>Undergraduate Scholar (Tuition Waiver)</b> , Shiv Nadar University	<i>Greater Noida</i>
2013	<b>National Talent Search Scholar</b> , Government of India	<i>New Delhi</i>

## Professional Services

---

### Millennium Scholars Program

*Pennsylvania State University*

MEMBER OF THE SELECTION COMMITTEE

2024

- Evaluate candidates' qualifications, including academic achievements, leadership potential, and commitment to increasing diversity in STEM fields.
- Collaborate with fellow interviewers and program coordinators to select scholars who demonstrate exceptional promise and dedication to making a positive impact in their fields and communities.

### ACM Conferences

*Pennsylvania State University*

STUDENT VOLUNTEER AND REVIEWER

2020-2023

- Student volunteer at CSCW 2020, CSCW 2021, CSCW 2022, UbiComp 2021, CHI 2023 and FAccT 2023
- It was an enriching experience coordinating and organizing paper sessions item Reviewed papers for various conferences and journals including CHI, CSCW, ISCRAM, IUI and Elsevier.

### Graduate Student Organization College of Information Sciences and Technology

*Pennsylvania State University*

PRESIDENT

2020-2021

- Spearheaded the organization's activities, including organizing talks, social events, and hackathons.
- Led efforts to foster both social and intellectual peer-to-peer connections within the IST community.
- Acted as a liaison between faculty and students, bridging communication and collaboration.

### HackData- The Data Science Hackathon

*Shiv Nadar University*

CO-ORGANIZER

Feb 2018

- Successfully co-organized HackData at Shiv Nadar University, Greater Noida, orchestrating a pan-India participation of over 200 students.
- Established key sponsorships, showcasing adept networking skills.
- Designed and implemented the comprehensive hackathon program, ensuring a dynamic and engaging experience for participants.

### SNU ACM-W Chapter

*Shiv Nadar University*

CHAIR

Nov. 2017 - Nov. 2018

- (Association for Computing Machinery - Women) Organized tutorials, talks and workshops to promote the presence of women in technology, as part as the founding committee of the SNU ACM-W Chapter

## Publications

---

- **Gautam, S.**, Zhang, X., and Rosson, M.B.. Towards Dynamic Learning: A Framework for Simulating Adaptive Learning Systems. Companion of the 2024 Computer-Supported Cooperative Work and Social Computing (CSCW 2024). **(Accepted)**
- Ghosh, S.\*, Venkit, P.\*, **Gautam, S.\***, Wilson, S., and Caliskan, A. Do Generative AI Models Output Harm while Representing Non-Western Cultures: Evidence from A Community-Centered Approach. AI, Ethics and Society (2024) **(Accepted)**
- **Gautam, S.** and Srinath, M. What the Annotator Does Not Know, Can Harm Us. Bridging Human-Computer Interaction and Natural Language Processing. NAACL 2024

- **Gautam, S.** and Lakshminarayanan, R. Balancing Act: Improving Privacy in AI through Explainability. ACM CHI Workshop on Human-Centered Explainable AI at CHI '24.
- **Gautam, S.**, Venkit, P., and Ghosh, S. From Melting Pots to Misrepresentations: Exploring Harms in Generative AI. CHI '24: Generative AI and HCI workshop in CHI '24.
- Mahomed, Y., Crawford, C. M., **Gautam, S.**, Friedler, S. A., & Metaxa, D. (2024, June). Auditing GPT's Content Moderation Guardrails: Can ChatGPT Write Your Favorite TV Show?. In The 2024 ACM Conference on Fairness, Accountability, and Transparency (pp. 660-686).
- Grace, R., **Gautam, S.**, & Tapia, A. (2023). Continuity planning for public health crises: Designing workplace redundancies for organizational resilience. *Journal of Emergency Management*, 21(6), 523-537.
- Dwaraghanath, R., Majethia, R., & **Gautam, S.** (2023). ECHO: An Automated Contextual Inquiry Framework for Anonymous Qualitative Studies using Conversational Assistants. arXiv preprint arXiv:2312.07576.
- Huang, C. Y., **Gautam, S.**, Brooks, S. M., Lin, Y. F., & Huang, T. H. K. (2023). Inspo: Writing Stories with a Flock of AIs and Humans. arXiv preprint arXiv:2311.16521.
- Pranav Venkit, Mukund Srinath, **Sanjana Gautam**, Saranya Venkatraman, Vipul Gupta, Rebecca Passonneau, and Shomir Wilson. 2023. The Sentiment Problem: A Critical Survey towards Deconstructing Sentiment Analysis. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, pages 13743–13763, Singapore. Association for Computational Linguistics.
- **Gautam, S.** What the Annotator Does Not Know, Can Harm Us. Workshop at CSCW' 23 : Understanding and Mitigating Cognitive Biases in Human-AI Collaboration. **(Workshop Position Paper)**
- **Gautam, S.\***, Narayanan Venkit, P.\*, Panchanadikar, R.\*, Huang, T. H., and Wilson, S. (2023, August). Unmasking Nationality Bias: A Study of Human Perception of Nationalities in AI-Generated Articles. In Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society (pp. 554-565).
- **Gautam, S.**, Rosson, M.B. and Akgun, M. Designing a Research Test-bed for Adaptive Learning Systems that Integrates Metacognition and Social Constructivism. In Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (pp. 1-6).
- Venkit, P. N., **Gautam, S.**, Panchanadikar, R., Huang, T. and Wilson, S. (2023). Nationality Bias in Text Generation. Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics.
- **Gautam, S.**, Akgun, M., and Mitra, P. (2022, March). Exploring the challenges of AI experts to inform AI curriculum. In Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 2 (pp. 1338-1338).
- Romera, G., **Gautam, S.** & Tapia, A. Understanding Twitter's behavior during the pandemic: Fake News and Fear, Poster at ISCRAM 2021
- **Gautam, S.**, & Rosson, M.B. (2020).. Exploring Feelings of Student Community across a Geographically Distributed University. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW1), 1-16.
- Saurabh, S., & **Gautam, S.** (2019). Modelling and statistical analysis of YouTube's educational videos: A channel Owner's perspective. *Computers & Education*, 128, 145-158.