

SASWAT DAS

222 Rice Hall, 85 Engineer's Way, Charlottesville, VA 22903

📞 +1 434-760-7428 📩 duh6ae@virginia.edu 🌐 saswatdas.com 🐾 github.com/SaswatD27

Responsible AI, (Differential) Privacy, Fairness, Security, Agentic LLMs

Education

University of Virginia (UVA) <i>Ph.D. in Computer Science</i> 4.0 GPA	Aug. 2023 – Present <i>Charlottesville, Virginia</i>
National Institute of Science Education and Research (NISER) <i>Integrated M.Sc. (BS + MS) in Mathematics (Minor in Computer Science)</i> 8.39/10.0 CGPA; 9.44/10.0 in Computer Science	Jul. 2018 – May 2023 <i>Bhubaneswar, India</i>

Experience

Department of Computer Science, University of Virginia <i>Graduate Research Assistant (PhD)</i> <ul style="list-style-type: none">Mentor: Dr. Ferdinando FiorettoFocus Areas: Differential Privacy, Fairness, Responsible AI, Agentic LLMs	Aug. 2023 – Present <i>Charlottesville, VA</i>
Pacific Northwest National Laboratory <i>PhD Intern (Summer and Fall)</i> <ul style="list-style-type: none">Mentors: Dr. Joseph Manzano Franco, Dr. Andres MarquezFocus Areas: Data Privacy and Security	July 2025 – December 2025 <i>Remote (from Charlottesville, VA)</i>
EECS, Syracuse University <i>Visiting Research Scholar</i> <ul style="list-style-type: none">Mentor: Dr. Ferdinando Fioretto (now at UVA)Focus Areas: Differential Privacy, Fairness, Responsible AI	June 2022 – Aug 2022 <i>Syracuse, NY</i>
School of Computer Sciences, NISER <i>MS Research Scholar</i> <ul style="list-style-type: none">Mentor: Dr. Subhankar MishraFocus Areas: Differential Privacy, Security, Responsible AI	Aug. 2021 – May 2023 <i>Bhubaneswar, India</i>
School of Computer Sciences, NISER <i>BS Summer and Winter Research Intern</i> <ul style="list-style-type: none">Mentor: Dr. Rishiraj Bhattacharyya (Now at the University of Birmingham)Focus Areas: Blockchain, P2P Networks, Cryptography, Zero-Knowledge Proofs	May 2019 – Jul. 2019, Dec 2019 – Jan. 2020 <i>Bhubaneswar, India</i>

Publications

Conference Publications

Fairness Issues and Mitigations in (Differentially Private) Socio-demographic Data Processes Joonhyuk Ko, Juba Ziani, <i>Saswat Das</i> , Matt Williams, Ferdinando Fioretto Accepted at AAAI-25 (Oral Presentation) and at AAAI-PPAI-25 (Oral Presentation) ; <i>arXiv:2408.08471</i>	2025
Disparate Impact on Group Accuracy of Linearization for Private Inference <i>Saswat Das</i> , Marco Romanelli, Ferdinando Fioretto Accepted at ICML-24 ; <i>arXiv:2402.03629</i>	2024
Finding ε and δ of Traditional Disclosure Control Systems <i>Saswat Das</i> , Keyu Zhu, Pascal van Hentenryck, Christine Task, Ferdinando Fioretto Accepted at AAAI-24 ; <i>DOI:10.1609/aaai.v38i20.30204</i>	2024

Workshop Publications

💡 Low-rank Finetuning for LLMs is Inherently Unfair	2025
<i>Saswat Das</i> , Marco Romanelli, Cuong Tran, Zarreen Reza, Bhavya Kailkhura, Ferdinando Fioretto	
CoLoRAI @ AAAI-25; Best Paper Award; arXiv:2405.18572	
Beyond Jailbreaking: Auditing Contextual Privacy in LLM Agents	2025
<i>Saswat Das</i> , Jameson Sandler, Ferdinando Fioretto	
MTI-LLM Workshop @ NeurIPS-25; arXiv:2506.10171	
Conversational Privacy Attacks Against Agentic LLMs	2025
<i>Saswat Das</i> , Joseph Moretto, David Evans, Ferdinando Fioretto	
Accepted at PPAI-25 (Extended Abstract)	
Examining Deidentified Data Quality using NIST Datasets and Tools	2023
<i>Saswat Das</i> , Razane Tajeddine, Ferdinando Fioretto	
Accepted at NIST CRC 2023	
Fair Context-Aware Privacy Threat Modelling	2023
<i>Saswat Das</i> , Rakshit Naidu	
Accepted at WPTM at USENIX SOUPS 2022; arXiv:2207.09750	

Book Chapters

Advances in Differential Privacy and Differentially Private Machine Learning: A Survey	2023
<i>Saswat Das</i> , Subhankar Mishra	

Published in **Information Technology Security** (Springer Nature)

Preprints

NeuroFilter: Privacy Guardrails for Conversational LLM Agents	2026
<i>Saswat Das</i> , Ferdinando Fioretto	

arXiv:2601.14660

-
- ## Selected Honours
- **Best Paper Award** for “Low-rank Finetuning for LLMs is Inherently Unfair” at CoLoRAI Workshop at AAAI-25
 - **AAAI-25 - Oral Presentation** (Paper Title: Fairness Issues and Mitigations in (Differentially Private) Socio-demographic Data Processes)
 - **AAAI - Student Travel Award** (2024 and 2025)
 - **UVA Computer Science Scholarship** (2023-24)
 - Selected for the **ELLIS Europe PhD Programme** (2023)
 - **DISHA Research Fellowship, Department of Atomic Energy, Government of India** (2018-2023)
 - Selected for **TIFR – Visiting Students’ Research Programme (VSRP), 2022** in **Technology & Computer Science**
 - **National Talent Search (NTS) Scholarship, NCERT, India** (2016)

Invited Talks

- **Google Privacy Seminar (May 2025)** - With Ferdinando Fioretto on “Auditing Privacy Leakage in LLM Agents”
- **TOC4Fairness (March 2025)** - With Joonhyuk Ko on “Fairness Issues and Mitigations in (Differentially Private) Socio-demographic Data Processes”
- **NIST CRC Workshop (December 2023)** - On “Examining Deidentified Data Quality using NIST Datasets and Tools”

Students Advised

- Joonhyuk Ko - BS, University of Virginia - CRA Outstanding Undergraduate Researcher (Honorable Mention)
- Kyle Briggs - BS, University of Virginia
- Lauren LaPorta - BS, University of Virginia
- Aarav Lodha - BS, University of Virginia
- Eric Nguyen - BA, University of Virginia

Teaching

- CS 6501 - Responsible AI: Privacy, Fairness, and Robustness (UVA, Spring 2025)
- CS 4710 - Artificial Intelligence (UVA, Fall 2024)

Service/Reviewing

- Workshop Workflow Co-Chair - **AAAI-PPAI 2025**
- Reviewer - **JAIR** (2025), **NeurIPS 2024**, **ICML 2025**, **NeurIPS 2025**, **ICLR 2025**
- Program Committee Member - **AAAI 2024**, **AAAI-PPAI 2024**, **AAAI 2025**
- Emergency Reviewer - **AAMAS 2023**, **AAMAS 2025**
- External Reviewer - National Institute of Standards and Technology (NIST) (2023)

Other Associations

- External Collaborator - Lawrence Livermore National Laboratory (2024 - Present)

Volunteering/Organising

Reading Group on Responsible Generative AI, University of Virginia

Oct. 2023 - Present

Organiser

at UVA

- Organising the reading group on responsible generative AI at the University of Virginia where members of various research groups that work on responsible AI/generative AI come together to discuss recent advancements and future directions of work.

CodeInPlace, Stanford University

Apr. 2021 - May 2021

Section Leader/Teaching Assistant

Online; Organised by Stanford University

- Taught coding to a section of students from a global pool of applicants using material from the first half of Stanford University's popular course, CS106A.

School of Mathematical Sciences, NISER

Feb. 2021 - May 2021

Department Mathematics Tutor

At NISER

- Served as a mathematics tutor to freshmen in the spring of 2021 as part of an initiative by mathematics majors at the School of Mathematical Sciences, NISER.

Aveti Learning

Apr. 2020 - Sep 2022

Volunteer/Tutor/Mentor

Online

- Served as an online mathematics tutor and mentor to high school students in rural Eastern India to make quality education and college advice more accessible.

Avanti Fellows

Sep. 2018 - Aug. 2019

Section Leader/Teaching Assistant

Dhenkanal, India

- Served as a mentor and mathematics tutor to high school students at JNV Dhenkanal as part of the acclaimed Avanti Fellows programme.

Workshops and Summer/Winter Schools

AI-SCORE: The AI School for CS and OR Education

Attendee; organised on the UMD College Park campus.

May-Jun 2024

INFORMS/ACM SIGAI

Technion: Summer School on Computer & Cyber Security

Attendee

September 2020

Organised by Technion, Israel

IIT-D: Winter School on Theoretical Computer Science

Accepted Attendee

December 2022

Organised by IIT-Delhi

- Did not attend due to academic schedule conflicts.

USENIX SOUPS '23: Workshop on Privacy Threat Modeling

Presenter/Attendee

August 2022

Organised by MITRE at USENIX SOUPS 2023

- Presented a short paper on context-aware fair privacy threat modeling.

Leadership / Extracurricular

Undergraduate Committee of the School

Dec. 2020 - Dec. 2022

Student Representative

School of Mathematical Sciences, NISER

- Served as one of two student representatives to the Undergraduate Committee of the School (UGCS) of Mathematical Sciences, NISER, an honour bestowed upon well-performing students by the school.
- Helped professors design and develop the academic course structure, pedagogy, and evaluation schemes of the undergraduate programme.

Inventa Magazine

July 2021 - Dec. 2021

Editor

IISERs, NISER, CeBS, and IISc (India)

- Worked on the editorial board for Inventa, an annual national science outreach magazine published jointly by the premier public research institutes of India: NISER, CeBS, IISc, and the IISERs.

Quizone (Campus Quiz Club), NISER

May. 2019 - Nov. 2021

President

NISER

- Served as the president of Quizone, the quiz club of NISER, HBNI, for two consecutive terms.
- Organised and hosted quizzes (online and offline), which involved planning and running logistics, making questions, maintaining a social media presence for the club, engaging with partners and sponsors etc.
- Helped codify club rules and regulations by drafting a club constitution.

Cultural Committee, NISER

Aug. 2019 - Nov. 2021

Editor

NISER

- Served as a member of the cultural committee of NISER, which is responsible for planning and organising cultural events in the college.

Delhi Public School Kalinga

Sep. 2016 - July 2018

Head Boy (Students' Council President)

Odisha, India

- Served as the Head Boy (2017-18) and Vice Head Boy (2016-17) of Delhi Public School Kalinga, in a role which involved representing the school on an interschool level, presiding over the students' council, organising events, and looking after the welfare and interests of the students of the school.

Skills (Not Exhaustive)

Programming/Scripting Languages: Python, R, C++, HTML/CSS, SQL, L^AT_EX

Developer Tools: VS Code, SLURM, AWS Cloud, R Studio, etc.

Selected Libraries: Opacus, OpenDP/SmartNoise, PyTorch, HuggingFace Transformers, Captum, sdcMicro + other common libraries

Languages: English, Hindi, Odia, (familiarity with) French