Sharan Sahu

www.sharansahu.com · ss4329@cornell.edu · Google Scholar

Last updated: May 22, 2025

Research interests

High-dimensional statistics, robust and stochastic optimization, reinforcement learning, deep learning, language and diffusion models, differential privacy

Education

2024 – Present **Cornell University** – Ithaca, NY
PhD in Statistics and Machine Learning
GPA: 4.105 / 4.000

2020 – 2024 **University of California, Berkeley** – Berkeley, CA
BA in Computer Science
GPA: 3.998 / 4.000
Advisors: Iain Carmichael, Ryan Tibshirani

Awards

2025 G-Research Doctoral Fellowship
2024 Cornell University Graduate Fellowship
2023 UC Berkeley Department of Data, Society, and Computing Data Science Insights Award Winner
2022 Science, Mathematics, and Research For Transformation DoD Scholarship (SMART)
2020 Northrop Grumman Engineering Scholarship
2020 Lockheed Martin Engineering Scholarship
2019 Math Olympiad Program (MOP) Invitee
2019 USA Physics Olympiad (USAPhO) Qualifier
2018 USA Mathematics Olympiad (USAMO) Qualifier
2017 USA Junior Mathematics Olympiad (USAJMO) Qualifier

Publications and preprints

* denotes first author(s)

2025 **DRO–REBEL: Fast and Robust Policy Optimization for RLHF via Distributional Regression**
Sharan Sahu*
In Preparation

2025 **Mixed Supervision Improves Performance As A Function Of Human Annotation Time For Nuclear Instance Segmentation**
Van Hovenga, Sharan Sahu, Iain Carmichael
Submitted (NeurIPS)

2025 **Towards Optimal Differentially Private Regret Bounds in Linear MDPs**
Sharan Sahu*
Submitted (NeurIPS)

2024 **WSIC: A Python Package To Facilitate Running Nuclear/Cellular Segmentation On Whole Slide Images**
Sharan Sahu*, Jerry Li, Van Hovenga, Kaitlin Smith, Neo Yin, Richard J. Chen, Iain Carmichael
In Preparation (JOSS)

2023 **Developing Multi-Dimensional Metrics for Precision, Recall, Fidelity, Diversity, and Authenticity in Evaluating Generative Networks Performance using Deep Perceptual Embeddings**
Sharan Sahu*, Daniel Flaherty, Abhishek Vinchure, Jonny Pei, Suya You
SPIE DCS

Invited Talks and Guest Lectures

Towards Optimal Differentially Private Regret Bounds in Linear MDPs
May 2025 Cornell University (Statistical Graduate Society)

The Machine Learning Problems Behind Large Language Models: Self-Supervision, Fine-Tuning, and Reinforcement Learning
February 2025 University of North Carolina, Chapel Hill (Computational Pathology Labs)

Unlocking the Power of Databases: The Crucial Role of Theory and Indices in Scalable Vector Databases for Machine Learning
August 2024 Naval Postgraduate School (NPS)

How Do Neural Networks Learn
December 2023 Naval Postgraduate School (NPS)

Anchored Intelligence: Navigating the Waters of Machine Learning And Charting the Course to Augmented Decision-Making
October 2023 Naval Postgraduate School (NPS)

Conference Posters and Oral Presentations

System Requirements Clustering with Machine Learning and Architecture Design
February 2025 Naval Applications of Machine Learning (NAML)
March 2024 Naval Applications of Machine Learning (NAML)
March 2023 Neptune Office of Naval Research (ONR) Conference

Developing Multi-Dimensional Metrics for Precision, Recall, Fidelity, Diversity, and Authenticity in Evaluating Generative Networks Performance using Deep Perceptual Embeddings
September 2022 DoD 6.1 Research Conference

Advancing Procedural Scene Synthesis through Enhanced Grammars and Gradient Policies in MetaSim

September 2021 UC Berkeley Data Science Conference

Using Machine Learning to Model and Discover New Catalysts To Address The Energy Challenges Posed by Climate Change

May 2021 UC Berkeley Data Science Conference

Teaching experience

Fall 2021 **EECS 16A: Foundations of Signals, Dynamical Systems, and Information Processing**
Served as an undergraduate course assistant for the EECS 16AB sequence, delivering foundational instruction in signal processing, control, circuit design, and machine learning with a strong emphasis on practical linear algebra applications. Facilitated both individual and group tutoring sessions, led EECS 16A review workshops, and supported over 50 students in lab settings by connecting theoretical concepts to hands-on implementation. Consistently recognized for teaching excellence, earning an average student rating of 4.9/5.0 for instructional quality, subject expertise, and proactive support.

Spring 2021 **CS 61A: Structure and Interpretation of Computer Programs**
Served as a undergraduate course assistant for the CS 61A series, covering software construction, machine operations, and programming abstraction using Python, Scheme, and SQL. Led weekly tutorials and bi-weekly labs focused on fundamental concepts such as recursion, induction, and data structures including trees and binary search trees. Prepared and delivered review sessions before exams, collaborated on the development of problem sets and assessments, and provided individualized support during office hours, contributing to measurable improvements in student performance.

Industry experience

2025-Present **Google Deepmind**, Student Researcher – London, UK
RLHF and LLM Reasoning With RL.

2023 - Present **167 Labs**, Cofounder and Lead Research Scientist – San Francisco, CA
Employee Self-Service AI Framework — Building an LLM-powered HR assistant using Multi-Modal RAG, API integrations, and personalized function calling for real-time policy, bonus, and paycheck insights.

Summer 2023 **US Department of Defense (DoD)**, Data Science Intern – San Diego, CA
System Requirements Automation — Built ML models and full-stack tools for classifying military system requirements, saving $1M and presenting to Pentagon and Marine Corps leadership.

Summer 2022

US Department of Defense (DoD), Software Engineering Intern – San Diego, CA
Defense ML & Cloud Systems — Built ML and cloud tools for EMF waveform classification, automated document parsing, and scalable resource management, boosting detection and efficiency across key pipelines.

Summer 2021 **Novartis**, Product Data Science Intern – Chicago, IL
Operations & Forecasting Analytics — Delivered ML and statistical solutions for inventory forecasting, database optimization, and secure access control, driving $1.5M in savings and boosting system efficiency and security.

2018 - 2020 **Northrop Grumman**, Software Engineering and Data Science Intern – Chicago, IL
Autonomous Systems Research — Developed ML models for sensor fusion and defect detection, combining filtering techniques and CNNs; published findings in a 20-page paper presented at Northrop Grumman's research conference.

Mentorship and service

2024 – Present **Cornell Statistics Graduate Society**, Vice President
Represented Ph.D. students in academic and departmental affairs, while organizing initiatives to support professional, social, and community development.

Skills

Programming
Proficient: Python, C++, C, Java
Familiar: R, Rust, C#, Go, Dart

Software
LaTeX, Git, Docker, PowerBI, Tableau, Spotfire, Qlik

Languages
English (native), Oriya (Advanced)

Coursework

Statistics: High dimensional statistics, asymptotic statistics, mean field asymptotics, mathematical statistics, nonparametric estimation, statistical computing, generalized linear models, empirical process theory, optimal transport, statistical learning theory

Mathematics: Real analysis, functional analysis, measure theoretic probability and martingale theory, robust and stochastic optimization, convex and variational analysis, differential manifolds, ring and group theory, random matrix theory, partial differential equations, linear algebra, topology and metric spaces

Machine Learning: Machine learning theory, reinforcement learning theory, deep learning and generative models

Computer Science: Data structures and algorithms, operating systems, compilers, computer security, computer architecture, database systems, networking and internet protocols, parallel computing,

Other interests

Basketball, Guitar, Reading, Cooking