sam588@cornell.edu
(408) 893-8077

Shubham Mohole

https://github.com/Shub500
US Citizen

Education

Cornell University — Ithaca, NY
B.A. in Computer Science, Minor: Mathematics, GPA: 4.0 — *Spring 2026*

- **Coursework Highlights**: Analysis of Algorithms, Machine Learning, AI Reasoning/Decision-Making, Software Testing, Computational Genetics & Genomics, Applied Econometrics, Probability Models and Inference, Computer System Organization, Linear Algebra, Computational Algebra
- **Teaching Assistant**: Data Structures and Functional Programming (CS 3110) - Spring 2025

Publications

[1] **Shubham Mohole, Sainyam Galhotra**: "SIFOTL: A Principled, Statistically-Informed Fidelity-Optimization Method for Tabular Learning," in *KDD*, 2025.

[2] **Shubham Mohole, Sainyam Galhotra**: "VeriMinder: Mitigating Analytical Vulnerabilities in NL2SQL," in *ACL*, 2025.

Experience

Lawrence Livermore National Laboratory — Livermore, CA
Summer Computing Scholar — *2025*

- Developed VeriMed-RAG, a retrieval-augmented generation framework for biomedical claim validation using decision-theoretic veritable checks
- Implemented self-reflection mechanisms and fine-tuned LLM modules for statistical check failures using synthetic examples based on Synthea-generated data
- Built evaluation datasets across multiple knowledge base configurations (TY0-TY10) to test RAG performance under varying document states

Prism Lab, Cornell University — Ithaca, NY
Research Assistant, Supervisor: Professor Sainyam Galhotra — *2024 – Present*

- Created VeriMinder's contextual semantic mapping framework with 53-category cognitive bias taxonomy, implementing multi-candidate LLM prompt engineering with critic feedback and self-reflection to operationalize Hard-to-Vary principle for analytical vulnerability detection, achieving 64% user effectiveness rating and 30% improvement over baseline approaches
- Architected SIFOTL's twin XGBoost models with Pareto-weighted decision trees (α-optimization) and LLM-guided feature synthesis operating exclusively on privacy-safe statistical summaries, achieving F1 scores of 0.85-0.96 on healthcare datasets and outperforming BigQuery Contribution Analysis by 85% on MEPS benchmark
- Built full-stack web application using React frontend and Flask backend with real-time SQL execution via SQLite/MySQL adapters, server-sent events for streaming responses, and modular service architecture supporting both standalone and enterprise plugin deployment modes

Social Dynamics Lab, Cornell University — Ithaca, NY
Research Assistant, Supervisor: Professor Michael Macy — *2024 – Present*

- Co-developed multi-LLM framework for automated persuasive dialogue generation handling complex social scenarios with minimal human oversight
- Fine-tuned BERT and LLaMA-3.2-8B models for persuasion strategy classification with cross-topic validation, achieving robust performance across diverse dialogue contexts

Selected Projects

- **ClamPy (Sparse AutoEncoders for LLM Behavior Control)**: Implemented core functionality for manipulating text generation using activation hooking and dynamic steering. Awarded PrimeIntellect Sponsor Prize at Open Source AGI Hackathon.
- **Rhythm Game**: Developed front-end graphics and backed logic for a rhythm game where, based on the music playing in the background, the player presses on tiles that are synchronized with the timing and beat per minute of the song playing. Written OUnit and inline tests to ensure the functionality.

Technical Skills

Programming Languages: Python, Java, C++, C, OCaml, JavaScript, R, MATLAB
ML/AI Frameworks: PyTorch, TensorFlow, scikit-learn, XGBoost, Transformers, NLTK, NumPy, Pandas
Technologies & Tools: Git, Docker, AWS, Node.js, React, MySQL, HTML/CSS, LaTeX