Shivam Gupta

shivamgupta@utexas.edu

EDUCATION

University of Texas at Austin (UT Austin) Dec 2024 (Expected)
Ph.D. in Computer Science

University of Illinois at Urbana-Champaign (UIUC) May 2018
B.S. in Computer Science, Minor in Mathematics

INTERESTS

Large Language Models, Diffusion Models, Machine Learning, Statistics, related topics

EXPERIENCE

Google DeepMind *San Francisco, CA*
Student Researcher May 2024 – Present
- Working on leveraging bandit optimization techniques for training and fine-tuning Large Language Models (LLMs) using evaluation feedback.
- Built optimized data and evaluation pipeline from scratch for the Gemma series of models for experimentation using Python, JAX and TensorFlow. Reduced training and evaluation times from several hours for the reference implementation to ≤ 5 minutes using 100+ TPUs. Showed result for algorithmically computing data mixture to optimize evaluation. Working on extensive evaluation.
- Showed that naive gradient estimation is too noisy for convergence, and contributed to an improved gradient estimator based on importance scores.
- Showed that the system is poorly conditioned, and proposed use of adaptive conditioning schemes (Adagrad). Showed large performance boost empirically.
- Proposed approach for fine-grained sampling of training datasets to optimize evaluation metrics.

Massachusetts Institute of Technology *Cambridge, MA*
Visiting Student Jun 2023 – Aug 2023
- Proved sharper rates for high-probability mean estimation. Resulted in paper at COLT 2024.

University of California, Berkeley *Berkeley, CA*
Visiting Student Researcher Aug 2022 – May 2023
- Developed theory for diffusion models, and wrote experiments to improve understanding, resulting in a paper (in submission to NeurIPS 2024).
- Developed mean estimation algorithms, resulting in multiple papers at NeurIPS, ICML and COLT.

Sigma Computing, Inc. *San Francisco, CA*
Research Intern May 2022 – Aug 2022
- Worked on designing and implementing anomaly detection algorithms for various datasets

University of Wisconsin, Madison *Madison, WI*
Research Intern May 2020 – Aug 2020
- Studied gradient descent algorithms to robustly estimate the mean of a high-dimensional Gaussian
- Wrote experiments for robust sparse estimation in Python. Resulted in NeurIPS 2022 paper.

Jane Street *New York, NY*
Software Developer Intern May 2016 – Aug 2016
- Wrote server code to employ state machine replication to send and receive data via RPCs
- Developed a market data parser in OCaml

Bloomberg L.P. *New York, NY*
R&D Intern May 2015 – Aug 2015
- Developed real-time system for Bloomberg Terminal to track messages between services in C++

PAPERS

12. Uncovering the Bias in Diffusion-based Posterior Sampling Methods
Shivam Gupta, Brett Levac, Jon Tamir, Eric Price
In submission to NeurIPS 2024

11. Faster Diffusion-based Sampling with Randomized Midpoints: Sequential and Parallel
Shivam Gupta, Linda Cai, Sitan Chen
In submission to NeurIPS 2024

10. Improved Sample Complexity Bounds for Diffusion Model Training
Shivam Gupta, Aditya Parulekar, Eric Price, Zhiyang Xun
In submission to NeurIPS 2024

9. **Diffusion Posterior Sampling is Computationally Intractable**
   Shivam Gupta, Ajil Jalal, Aditya Parulekar, Eric Price, Zhiyang Xun
   International Conference on Machine Learning (ICML) 2024
8. **Beyond Catoni: Sharper Rates for Heavy-Tailed and Robust Mean Estimation**
   Shivam Gupta, Samuel B. Hopkins, Eric Price
   Conference on Learning Theory (COLT) 2024
7. **Minimax-Optimal Location Estimation**
   Shivam Gupta, Jasper C.H. Lee, Eric Price, Paul Valiant
   Neural Information Processing Systems (NeurIPS) 2023
6. **Finite-Sample Symmetric Mean Estimation with Fisher Information Rate**
   Shivam Gupta, Jasper C.H. Lee, Eric Price
   Conference on Learning Theory (COLT) 2023
5. **High-dimensional Location Estimation via Norm Concentration for Subgamma Vectors**
   Shivam Gupta, Jasper C.H. Lee, Eric Price
   International Conference on Machine Learning (ICML) 2023
4. **Finite-Sample Maximum Likelihood Estimation of Location**
   Shivam Gupta, Jasper C.H. Lee, Eric Price, Paul Valiant
   Neural Information Processing Systems (NeurIPS) 2022
3. **Outlier-Robust Sparse Estimation via Non-Convex Optimization**
   Yu Cheng, Ilias Diakonikolas, Rong Ge, Shivam Gupta, Daniel Kane, Mahdi Soltanolkotabi
   Neural Information Processing Systems (NeurIPS) 2022
2. **Sharp Constants in Uniformity Testing via the Huber Statistic**
   Shivam Gupta, Eric Price
   Conference on Learning Theory (COLT) 2022
1. **Nash Equilibrium Computation in Resource Allocation Games**
   Shivam Gupta, Ruta Mehta
   International Conference on Autonomous Agents and Multiagent Systems (AAMAS) 2018

OTHER EXPERIENCE

Research with Prof. Andreas Klockner (UIUC) Aug 2014 –May 2017

- Developed and implemented mesh-refinement algorithms (in Python and using NumPy) to iteratively refine and coarsen meshes while preserving connectivity information.

Eventifier - Software Engineer Intern May 2014–Jul 2014

- Developed system to stream Twitter data related to particular topics and classify them using NLP.

Freelance Programming May 2014–Aug 2014

- Developed Android app for scholarship database company

Game Development Jan 2013-Aug 2014

- Developed a game engine in C++ and SDL with entity management and rendering functionality
- Developed several game prototypes using the engine
- Initiated organization of Global Game Jam in 2013 for the first time in India (still occurring annually as of 2024)

SELECTED AWARDS

- COLT 2024 Travel Award
- C.W. Gear Outstanding Undergraduate Student Award 2018
- Conference Travel Grant 2018 (for travel to AAMAS)
- Horace and Kate King Wu International Undergraduate Scholarship 2018
- Illinois Engineering Achievement Scholarship 2017
- NTT Data, Inc., Scholarship 2015, 2016
- Franz Hohn and J.P. Nash Scholarship 2015 (for research in scientific computing)
- Perfect score in the Indian National Olympiad in Informatics, and selected as one of 26 students in India to attend the International Olympiad in Informatics training camp 2014
- ACM ICPC Mid-Central Regional: Team placed 6th in 2017, 4th in 2016
- Represented India in SEARCC Software Competition 2013, Colombo, Sri Lanka, and placed 3rd

- Placed 2nd in Dropbox Open programming contest 2015 at UIUC
- Won Bloomberg CodeCon Challenge and invited to CppCon 2015 in Bellevue, Washington
- Won 3Red Trading Tech Challenge in 2015 and 2016 and invited to Chicago

SKILLS

- *Languages:* Python, C/C++, Java, OCaml, Haskell, JavaScript
- *Software and Libraries:* PyTorch, JAX, TensorFlow, NumPy, SciPy, Mathematica, LaTeX

SERVICE

- *Program Committee:* ALT 2025, AAAI 2025
- *Reviewing:* SODA 2021, 2024; NeurIPS 2023, 2024; ITCS 2024; ALT 2024; ICLR 2024, 2025; ICML 2024, TF2M@ICML 2024

TALKS

- **Beyond Catoni: Sharper Rates for Heavy-Tailed and Robust Mean Estimation** July 2024
  Conference on Learning Theory 2024 (Online)
- **Sample-Efficient Training for Diffusion** November 2023
  IFML Workshop on Generative AI, UT Austin
- **A Finite-Sample Theory for Mean Estimation with Fisher Information Rate** October 2023
  MIT Algorithms and Complexity Seminar
- **A Finite-Sample Theory for Mean Estimation with Fisher Information Rate** October 2023
  CMU Theory Lunch
- **Finite-Sample Symmetric Mean Estimation with Fisher Information Rate** July 2023
  Conference on Learning Theory 2023, Bangalore, India
- **High-dimensional Estimation via Norm Concentration for Subgamma Vectors** July 2023
  International Conference on Machine Learning 2023 (Online)
- **Sharp Constants in Uniformity Testing via the Huber Statistic** July 2022
  ML Tea, UT Austin

TEACHING

At UT Austin:

- *Teaching Assistant*, Sublinear Algorithms (CS 395T) Fall 2020
- *Teaching Assistant*, Introduction to Algorithms (CS 331) Spring 2019
- *Teaching Assistant*, Machine Learning (CS 395T) Summer 2019, Fall 2019
- *Teaching Assistant*, Elements of Data Analytics (CS 329E) Spring 2019
- *Teaching Assistant*, Elements of Computer Programming (CS 303E) Fall 2018

At UIUC:

- *Course Assistant*, Introduction to Algorithms (CS 374) Spring 2018
- *Grader,* Algorithms II (CS 473) Spring 2018

RELEVANT COURSES

Probability and Stochastic Processes, Learning Theory, Coding Theory, Theoretical Statistics, Randomized Algorithms, Markov Chains and Mixing Time, Approximation Algorithms, Combinatorial Mathematics, Numerical Linear Algebra, Wireless Networking