# SURBHI GOEL

https://www.surbhigoel.com
[first name][last initial]@cis.upenn.edu

## EDUCATION

**The University of Texas at Austin** August 2015 - June 2020
M.S. and Ph.D. in Computer Science
Advisor: Adam R. Klivans
Committee: Alex Dimakis, Raghu Meka, Eric Price
Dissertation: Towards Provably Efficient Algorithms for Learning Neural Networks
***Received the Bert Kay dissertation award***

**Indian Institute of Technology, Delhi** July 2011 - May 2015
B.Tech. in Computer Science and Engineering

## APPOINTMENTS

**University of Pennsylvania, Philadelphia, PA** January 2023 - Present
*Magerman Term Assistant Professor, Computer and Information Science*

**Microsoft Research, New York, NY** July 2020 - December 2022
*Postdoctoral Researcher, Machine Learning Group*

**Institute for Advanced Study, Princeton, NJ** January - May 2020
*Visiting Graduate Student, Theoretical Machine Learning Program*

**Simons Institute for Theory of Computing, Berkeley, CA** May - August 2019
*Research Fellow, Foundations of Deep Learning Program*

## RESEARCH INTERESTS

My research is on the theoretical aspects of the modern practice of machine learning, where my goal is to develop the next generation of principled machine learning methods. In the pursuit of this goal, my work focuses on quantifying the computational and statistical aspects of state-of-the-art deep learning methods, and expanding the toolbox of current algorithms using new theoretically grounded insights.

## AWARDS AND FELLOWSHIPS

2023 Microsoft Accelerate Foundation Models Research Award
2020 Bert Kay Dissertation Award for best dissertation in CS at UT Austin
2019 Rising Stars in ML by University of Maryland
2019 Rising Stars in EECS by UIUC
2019 J.P. Morgan AI PhD Fellowship
2019 Simons-Berkeley Research Fellowship for Foundations of Deep Learning program
2018 The University of Texas at Austin Graduate Continuing Bruton Fellowship
2017 The University of Texas at Austin Graduate School Summer Fellowship
2015 ICIM Stay Ahead Award and Suresh Chandra Memorial Trust Award for Undergraduate Thesis
2011 Aditya Birla Scholarship
2011 OPJEM Scholarship
2011 All India Rank 37 (*Rank 2 amongst women applicants*) in IITJEE among 450,000 students
2010 Indian National Mathematics Olympiad Top 30

## PUBLICATIONS

* *indicates $\alpha$-$\beta$ (alphabetical) ordering.*

### PREPRINTS

GuanWen Qiu, Da Kuang, **Surbhi Goel**
**Complexity Matters: Feature Learning in the Presence of Spurious Correlations**
In submission, 2024

Benjamin L. Edelman*, Ezra Edelman*, **Surbhi Goel***, Eran Malach*, Nikolaos Tsilivis*
**The Evolution of Statistical Induction Heads: In-Context Learning Markov Chains**
In submission, 2024

### CONFERENCE PAPERS

**Surbhi Goel***, Steve Hanneke*, Shay Moran*, Abhishek Shetty*
**Adversarial Resilience in Sequential Prediction via Abstention**
Neural Information Processing Systems (NeurIPS) 2023

Benjamin L. Edelman*, **Surbhi Goel***, Sham M. Kakade*, Eran Malach*, Cyril Zhang*
**Pareto Frontiers in Neural Feature Learning: Data, Compute, Width, and Luck**
Neural Information Processing Systems (NeurIPS) 2023
***Selected as a spotlight presentation***

Bingbin Liu, Jordan T. Ash, **Surbhi Goel**, Akshay Krishnamurthy, Cyril Zhang
**Exposing Attention Glitches with Flip-Flop Language Modeling**
Neural Information Processing Systems (NeurIPS) 2023
***Selected as a spotlight presentation***

Sitan Chen*, Zehao Dou*, **Surbhi Goel***, Adam R. Klivans*, Raghu Meka*
**Learning Narrow One-Hidden-Layer ReLU Networks**
Conference on Learning Theory (COLT) 2023

Bingbin Liu, Jordan T. Ash, **Surbhi Goel**, Akshay Krishnamurthy, Cyril Zhang
**Transformers Learn Shortcuts to Automata**
International Conference on Learning Representations (ICLR) 2023
***Selected as a notable top-5% paper***

**Surbhi Goel***, Sham M. Kakade*, Adam T. Kalai*, Cyril Zhang*
**Recurrent Convolutional Neural Networks Learn Succinct Learning Algorithms**
Neural Information Processing Systems (NeurIPS) 2022

Boaz Barak*, Benjamin L. Edelman*, **Surbhi Goel***, Sham M. Kakade*, Eran Malach*, Cyril Zhang*
**Hidden Progress in Deep Learning: SGD Learns Parities Near the Computational Limit**
Neural Information Processing Systems (NeurIPS) 2022

Benjamin L. Edelman*, **Surbhi Goel***, Sham M. Kakade*, Cyril Zhang *
**Inductive Biases and Variable Creation in Self-Attention Mechanisms**
International Conference on Machine Learning (ICML) 2022

Nikunj Saunshi, Jordan T. Ash, **Surbhi Goel**, Dipendra Misra, Cyril Zhang, Sanjeev Arora, Sham M. Kakade, Akshay Krishnamurthy
**Understanding Contrastive Learning Requires Incorporating Inductive Biases**
International Conference on Machine Learning (ICML) 2022

Jordan T. Ash, Cyril Zhang, **Surbhi Goel**, Akshay Krishnamurthy, Sham M. Kakade
**Anti-Concentrated Confidence Bonuses For Scalable Exploration**
International Conference on Learning Representations (ICLR) 2022

Jordan T. Ash*, **Surbhi Goel***, Akshay Krishnamurthy*, Dipendra Misra*
**Investigating the Role of Negatives in Contrastive Representation Learning**
International Conference on Artificial Intelligence and Statistics (AISTATS) 2022

Jordan T. Ash, **Surbhi Goel**, Akshay Krishnamurthy, Sham M. Kakade
**Gone Fishing: Neural Active Learning with Fisher Embeddings**
Neural Information Processing Systems (NeurIPS) 2021

Naman Agarwal*, **Surbhi Goel***, Cyril Zhang*
**Acceleration via Fractal Learning Rate Schedules**
International Conference on Machine Learning (ICML) 2021

Anthimos-Vardis Kandiros, Yuval Dagan, Nishanth Dikkala, **Surbhi Goel**, Constantinos Daskalakis
**Statistical Estimation from Dependent Data**
International Conference on Machine Learning (ICML) 2021

**Surbhi Goel***, Adam R. Klivans*, Pasin Manurangsi*, Daniel Reichman*
**Tight Hardness Results for Learning One-Layer ReLU Networks**
Innovations in Theoretical Computer Science (ITCS) 2021

**Surbhi Goel***, Adam R. Klivans*, Frederic Koehler*
**From Boltzmann Machines to Neural Networks and Back Again**
Neural Information Processing Systems (NeurIPS) 2020

**Surbhi Goel***, Aravind Gollakota*, Adam R., Klivans*
**Statistical-Query Lower Bounds via Functional Gradients**
Neural Information Processing Systems (NeurIPS) 2020

**Surbhi Goel***, Aravind Gollakota*, Zhihan Jin*, Sushrut Karmalkar*, Adam R. Klivans*
**Superpolynomial Lower Bounds for Learning One-Layer Neural Networks using Gradient Descent**
International Conference on Machine Learning (ICML) 2020

Omar Montasser, **Surbhi Goel**, Ilias Diakonikolas, Nathan Srebro
**Efficiently Learning Adversarially Robust Halfspaces with Noise**
International Conference on Machine Learning (ICML) 2020

Jessica Hoffmann, Soumya Basu, **Surbhi Goel**, Constantine Caramanis
**Learning Mixtures of Graphs from Epidemic Cascades**
International Conference on Machine Learning (ICML) 2020

Ilias Diakonikolas*, **Surbhi Goel***, Sushrut Karmalkar*, Adam R. Klivans*, Mahdi Soltanolkotabi*
**Approximation Schemes for ReLU Regression**
Conference on Learning Theory (COLT) 2020

**Surbhi Goel**
**Learning Ising and Potts Models with Latent Variables**
International Conference on Artificial Intelligence and Statistics (AISTATS) 2020

**Surbhi Goel***, Sushrut Karmalkar*, Adam R. Klivans*
**Time/Accuracy Trade-offs for Learning a ReLU with respect to Gaussian Marginals**
Neural Information Processing Systems (NeurIPS) 2019
***Selected for a spotlight presentation***

**Surbhi Goel***, Daniel Kane*, Adam R. Klivans*
**Learning Ising Models with Independent Failures**
Conference on Learning Theory (COLT) 2019

**Surbhi Goel***, Adam R. Klivans*
**Learning Neural Networks with Two Nonlinear Layers in Polynomial Time**
Conference on Learning Theory (COLT) 2019

**Surbhi Goel***, Adam R. Klivans*, Raghu Meka*
**Learning One Convolutional Layer with Overlapping Patches**
International Conference on Machine Learning (ICML) 2018
***Selected for a full oral presentation***

**Surbhi Goel***, Adam R. Klivans*
**Eigenvalue Decay Implies Polynomial-Time Learnability for Neural Networks**
Neural Information Processing Systems (NeurIPS) 2017

**Surbhi Goel***, Varun Kanade*, Adam R. Klivans*, Justin Thaler*
**Reliably Learning ReLU in Polynomial Time**
Conference on Learning Theory (COLT) 2017

## WORKSHOP PAPERS

GuanWen Qiu, Da Kuang, **Surbhi Goel**
**Complexity Matters: Feature Learning in the Presence of Spurious Correlations**
Mathematics of Modern Machine Learning, Neural Information Processing Systems (NeurIPS) 2023

Bingbin Liu, Jordan T. Ash, **Surbhi Goel**, Akshay Krishnamurthy, Cyril Zhang
**Exposing Attention Glitches with Flip-Flop Language Modeling**
Challenges of Deploying Generative AI, International Conference on Machine Learning (ICML) 2023
Knowledge and Logical Reasoning in the Era of Data-driven Learning, International Conference on Machine Learning (ICML) 2023

Jessica Hoffmann, Soumya Basu, **Surbhi Goel**, Constantine Caramanis
**Disentangling Mixtures of Epidemics on Graphs**
Graph Representation Learning, Neural Information Processing Systems (NeurIPS) 2019

**Surbhi Goel***, Adam R. Klivans*
**Learning Depth-Three Neural Networks in Polynomial Time**
Deep Learning: Bridging Theory and Practice, Neural Information Processing Systems (NeurIPS) 2017

**Surbhi Goel***, Varun Kanade*, Adam R. Klivans*, Justin Thaler*
**Reliably Learning ReLU in Polynomial Time**
Optimization for Machine Learning (OPT), Neural Information Processing Systems (NeurIPS) 2016
***Selected for an oral presentation***

## UNPUBLISHED MANUSCRIPTS

**Surbhi Goel***, Rina Panigrahy*
**Learning Two layer Networks with Multinomial Activation and High Thresholds**
Manuscript, 2019

Matthew Jordan, Naren Manoj, **Surbhi Goel**, Alexandros Dimakis
**Quantifying Perceptual Distortion of Adversarial Examples**
Manuscript, 2019

Simon Du*, **Surbhi Goel***
**Improved Learning of One-hidden-layer Convolutional Neural Networks with Overlaps**
Manuscript, 2018.

## INVITED TALKS

**Understanding Training Dynamics in Deep Learning using Simplified Models**
*Optimization Seminar at UPenn* — *March 2024*

**How do Large Language Models Think?**
*Women in Data Science at UPenn* — *February 2024*

**Beyond Worst-case Sequential Prediction: Adversarial Robustness via Abstention**
*CIS and MINDS Seminar at JHU* — *March 2024*
*EnCORE Workshop at IPAM, UCLA* — *March 2024*
*Theory Seminar at UPenn* — *November 2023*
*Alg-ML Seminar at Princeton* — *November 2023*
*BLISS Seminar at UC Berkeley* — *October 2023*
*Math Machine Learning seminar at MPI MIS + UCLA* — *August 2023*
*FODSI Workshop on Computational Complexity of Statistical Problems at MIT* — *June 2023*

**Thinking fast with Transformers - Algorithmic Reasoning via Shortcuts**
*Deep Learning Down Under Workshop, Lorne, Australia* — *January 2024*
*IFML Workshop on Generative AI at UT Austin* — *November 2023*
*Youth in High Dimensions, Trieste, Italy* — *May 2023*
*MaD Seminar at NYU* — *April 2023*
*ASSET Seminar at UPenn* — *April 2023*

**Sparse Feature Emergence in Deep Learning**
*Symposium on New Directions in Theoretical Machine Learning [slides]* — *September 2022*

**What Functions do Self-attention Blocks Prefer to Represent?**
**Demystifying Attention-based Architectures in Deep Learning**
*Joint IFML/Data-Driven Decision Processes Workshop at Simons Institute* — *October 2022*
*ML Foundations Seminar at MSR Redmond* — *August 2022*

*Workshop on Algorithms for Learning and Economics (WALE) in Greece* — *June 2022*
*ML Symposium at USC* — *December 2021*
*ELLIS Talk Series at IST Austria* — *December 2021*
*Learning Theory Workshop at Google* — *October 2021*

**The Hidden Progress Behind Loss Curves**
*Workshop on Learning: Optimization and Stochastics at EPFL* — *July 2022*

**Principled Algorithm Design in the Era of Deep Learning**
*CS/CSE Colloquium at NYU Courant/Tandon* — *April 2022*
*CS Colloquium at UW-Madison* — *March 2022*
*CS Colloquium at Halicioglu Data Science Institute, UCSD* — *March 2022*
*CS Colloquium at UMD* — *February 2022*
*SCS Talk at CMU* — *February 2022*
*CS Colloquium at Duke* — *February 2022*
*CIS Colloquium at UPenn* — *February 2022*
*CS Colloquium at Cornell* — *February 2022*
*Talks at TTIC* — *February 2022*

**Computational Barriers For Learning Some Generalized Linear Models**
*Information-Computation Trade-offs Workshop at Simons Institute [video][slides]* — *September 2021*

**Computational Complexity of ReLU Regression**
*The Multifaceted Complexity of Machine Learning Workshop at IMSI [video]* — *April 2021*

**Computational Complexity of Learning Neural Networks over Gaussian Marginals**
*MIC Seminar at NYU* — *May 2020*
*Algorithms Seminar at Duke University* — *October 2020*
*ML Theory Seminar at Harvard University [video]* — *October 2020*
*ARC Colloquium at Georgia Tech* — *November 2020*
*IDEAL Seminar at TTIC* — *November 2020*
*TOC Colloquium at MIT* — *December 2020*
*SILO Seminar at UW-Madison* — *January 2020*
*Statistics Seminar at Stanford University* — *July 2021*

**Approximation Schemes for ReLU Regression**
*Deep Learning Program Reunion at Simons Institute* — *August 2020*

**Provably Efficient Algorithms for Learning Neural Networks**
*Microsoft Research New York* — *February 2020*
*Microsoft Research New England* — *February 2020*
*Microsoft Research Redmond* — *February 2020*

**Time/Accuracy Tradeoffs for Learning a ReLU wrt Gaussian Marginals**
*Spotlight Talk at Neural Information Processing Systems (NeurIPS)* — *December 2019*

**Exploring Surrogate Losses for Learning Neural Networks**
*TTIC Young Researcher Seminar Series* — *December 2019*

**Efficiently Learning Simple Neural Networks**
*Rising Star in ML Talk at University of Maryland* — *September 2019*

**Learning Ising Models with Independent Failures**
*Research Fellows Talk at Simons Institute* — *July 2019*

**Efficiently Learning Simple Convolutional Networks**
*China Theory Week at Tsinghua University* *September 2019*

**Learning One Convolutional Layer with Overlapping Patches**
*Google Research Theory Reading Group* *June 2018*

**Reliably Learning the ReLU in Polynomial Time**
*OPT-ML Workshop at Neural Information Processing Systems (NeurIPS)* *December 2016*

## WORK EXPERIENCE

**Google, Mountain View CA** May - August 2018
*Research Intern* *Supervisor: Rina Panigrahy*

**Dell, Round Rock TX** June - August 2017
*Research Intern*

**Google, New York, NY** May - August 2016
*Research Intern* *Supervisor: Natalia Ponomareva*

**Google, Mountain View CA** May - August 2014
*Software Engineering Intern* *Supervisor: Neha Jha*

**University of Michigan, Ann Arbor MI** May - July 2013
*Research Scholar* *Supervisor: Atul Prakash*

## TEACHING

**CIS 5200: Machine Learning** Spring 2024
*Co-instructor with Eric Wong* *University of Pennsylvania*

**CIS 7000: Foundations of Modern ML - Theory and Empirics** Fall 2023
*Instructor* *University of Pennsylvania*

**CIS 5200: Machine Learning** Spring 2023
*Co-instructor with Eric Wong* *University of Pennsylvania*

## OUTREACH

**Co-founder** 2020-Present
*Learning Theory Alliance (LeT-All)*

Co-organized the Fall 2023 Mentoring Workshop
Co-organized the Fall 2022 Mentoring Workshop in collaboration with FODSI
Co-organized the COLT 2022 Mentoring Panel
Co-organized the ALT 2022 Mentoring Workshop
Co-organized the Graduate Applications Support Program in collaboration with WiML-T
Co-organized the COLT 2021 Mentoring Workshop
Co-organized the ALT 2021 Mentoring Workshop

**Mentor**
*Women in Machine Learning Theory (WiML-T) Mentoring Program* *2021-Present*
*UT Austin's Women in CS (GWC-WiCS) Mentoring Program* *2018-19*

**Panelist**

*New in ML Workshop, NeurIPS 2023* — *Decemeber 2023*
*WiML Un-Workshop, ICML 2022* — *July 2022*
*New Horizons in Theoretical Computer Science* — *June 2022*
*VMware Nirman for Women in Tech* — *January 2021*

## SERVICE ROLES

**Co-treasurer** 2024-Present
*Association for Computational Learning*

**Workshop Co-Organizer** 2024-Present
*Transformers as a Computational Model, Simons Institute's Special Year on LLMs and Transformers*

**Program Co-Organizer** 2023-Present
*Simons Institute's Special Year on LLMs and Transformers*

**Office Hours Chair** 2023-Present
*International Conference on Learning Representations (ICLR) 2024*

**Workshop Reviewing Committee** 2024
*International Conference on Machine Learning (ICML)*

**Workshop Co-organizer** 2023
*Mathematics of Modern Machine Learning (M3L) at NeurIPS 2023*

**Virtual Experience Chair** 2023
*Conference on Learning Theory (COLT) 2023*

**Online Experience Chair** 2021
*Conference on Learning Theory (COLT) 2021*

Co-organized the virtual part of the hybrid conference, including the 2-day virtual-only program

**Seminar Co-organizer** 2020-21
*One World Machine Learning Seminar Series*

**Treasurer** 2016-17
*Graduate Representative Association of Computer Sciences (GRACS) 2024*

**Program Committee**

*International Conference on Algorithmic Learning Theory (ALT)* — *2021/22/23/24*
*Conference on Learning Theory (COLT)* — *2021/22/24*
*International Conference on Artificial Intelligence and Statistics (AISTATS) (area chair)* — *2023*
*Neural Information Processing Systems (NeurIPS) (area chair)* — *2023*
*International Conference on Algorithmic Learning Theory (ALT) (senior program committee)* — *2024*
*Conference on Learning Theory (COLT) (senior program committee)* — *2024*

**Conference Reviewing**

*Symposium on Theory of Computing (STOC)* — *2019/20/21*
*Neural Information Processing Systems (NeurIPS)* — *2018 (top 30%)/20/21*
*Conference on Learning Theory (COLT)* — *2018/19/20*
*International Conference on Learning Representations (ICLR)* — *2019/20/23*

*Symposium on Discrete Algorithms (SODA)* *2020/23*
*Foundations of Computer Science (FOCS)* *2020/22*
*International Conference on Machine Learning (ICML)* *2019 (top 5%)*

**Journal Reviewing**

*TheoretiCS* *2024*
*Journal of Machine Learning Research* *2021/22*
*IEEE Transactions on Information Theory* *2020*