

# Guneet S. Dhillon

 [guneet.dhillon@stats.ox.ac.uk](mailto:guneet.dhillon@stats.ox.ac.uk)  
 [guneet-dhillon.github.io](https://github.com/guneet-dhillon)

## Education

- 2021– **D.Phil. in Statistics**, *University of Oxford*, Oxford, UK
- *Topic* : Robust Machine Learning: A Predictive and Generative Perspective
  - *Supervisors* : Prof. Yee Whye Teh, Prof. Arnaud Doucet, Prof. George Deligiannidis, Prof. Tom Rainforth
  - *Funding* : Clarendon Fund Scholarship, University of Oxford (only 1 awardee from the Department of Statistics)
  - *College* : St. John's College
- 2014–18 **B.Sc. in Computer Science with Honors – Turing Scholars Honors & B.Sc. in Mathematics with Honors**, *University of Texas at Austin*, Austin, TX, USA, *GPA: 3.89/4.00*
- *Certificate in Applied Statistical Modeling*
  - *Minor in Economics*
  - *Honors Thesis* : Training Ising Models on Images using Sparsitron (Supervisor: Prof. Adam Klivans)
  - *Research Programs* : Freshman Research Initiative, College of Natural Sciences  
Directed Reading Program, Department of Mathematics

## Work Experience

- 2025 **Researcher Intern**, *Microsoft Research*, Cambridge, UK
- *Preprint* : E-Scores for (In)Correctness Assessment of Generative Model Outputs
- 2023–24 **Machine Learning Intern**, *Boson AI USA Inc.*, Santa Clara, CA, US
- *Publication* : L3Ms — Lagrange Large Language Models @ ICLR 2025
- 2018–21 **Applied Scientist II (2019-21) & I (2018-19)**, *Amazon Web Services, Inc.*, Pasadena, CA, USA
- *Publications* : Uniform Sampling over Episode Difficulty @ NeurIPS (Spotlight) 2021  
A Baseline for Few-Shot Image Classification @ ICLR 2020
  - *Patent* : Structured Document Analyzer
  - *Products* : Amazon Rekognition Custom Labels [[web](#)]  
Amazon Textract [[web](#)]
- 2017 **Software Development Engineer Intern**, *Amazon Web Services, Inc.*, Palo Alto, CA, USA
- *Publication* : Stochastic Activation Pruning for Robust Adversarial Defense @ ICLR 2018
- 2016 **Machine Learning Intern**, *CognitiveScale*, Austin, TX, USA
- 2015 **Software Technology Engineering Intern**, *Dell Technologies Inc.*, Austin, TX, USA

## Publications

- 2025 **L3Ms — Lagrange Large Language Models** [[proc](#)][[arXiv](#)][[code](#)] **ICLR**  
G. S. Dhillon, X. Shi, Y. W. Teh, A. Smola
- 2024 **On the Expected Size of Conformal Prediction Sets** [[proc](#)][[arXiv](#)][[code](#)] **AISTATS**  
G. S. Dhillon, G. Deligiannidis, T. Rainforth
- 2021 **Uniform Sampling over Episode Difficulty** [[proc](#)][[arXiv](#)][[code](#)] **NeurIPS (Spotlight)**  
S. M. R. Arnold\*, G. S. Dhillon\*, A. Ravichandran, S. Soatto (\*Equal contribution)
- 2020 **A Baseline for Few-Shot Image Classification** [[proc](#)][[arXiv](#)][[code](#)] **ICLR**  
G. S. Dhillon, P. Chaudhari, A. Ravichandran, S. Soatto
- 2018 **Stochastic Activation Pruning for Robust Adversarial Defense** [[proc](#)][[arXiv](#)][[code](#)] **ICLR**  
G. S. Dhillon, K. Azizzadenesheli, Z. C. Lipton, J. Bernstein, J. Kossaifi, A. Khanna, A. Anandkumar

## Preprints

- 2025 **E-Scores for (In)Correctness Assessment of Generative Model Outputs** [[arXiv](#)]  
G. S. Dhillon, J. González, T. Pandeva, A. Curth

## Theses

- 2018 **Training Ising Models on Images using Sparsitron** [[web](#)]  
*Undergraduate Honors Thesis*  
Supervisor: Prof. Adam Klivans, Co-Supervisor: Dr. Philipp Krähenbühl

## Other Articles

- 2020 **Erratum Concerning the Obfuscated Gradients Attack on Stochastic Activation Pruning** [[arXiv](#)]  
G. S. Dhillon\*, N. Carlini\* (\*Equal contribution)

## Patents

- 2020 **Structured Document Analyzer** [[pdf](#)][[text](#)] **US Patent 10839245**  
G. S. Dhillon, V. Mahadevan, Y. Zhang, M. Wang, G. Payyavula, V. C. Nguyen, R. Bhotika, S. Soatto

## Talks and Presentations

- E-Scores for (In)Correctness Assessment of Generative Model Outputs  
2025 Algorithmic Statistics Workshop, Doctoral Training Centre, University of Oxford [[web](#)]  
On the Expected Size of Conformal Prediction Sets  
2023 ELLIS Robust Machine Learning Workshop [[web](#)]  
Uniform Sampling over Episode Difficulty  
2022 Seminar on Continual Learning, Meta-Learning, Transfer Learning and Related Subjects, Google DeepMind  
A Baseline for Few-Shot Image Classification  
2021 Workshop on Computer Vision with Limited Labels, Amazon Computer Vision Conference  
2019 Workshop on Meta-Learning, Conference on Neural Information Processing Systems (NeurIPS) [[web](#)] (*Spotlight*)  
2019 Workshop on Computer Vision Services/Systems in Amazon, Amazon Machine Learning Conference  
2019 Workshop on Data-Efficient Learning Techniques for Amazon Scale, Amazon Machine Learning Conference  
Stochastic Activation Pruning for Robust Adversarial Defense  
2017 Machine Deception Workshop, Conference on Neural Information Processing Systems (NeurIPS) [[web](#)]

## Teaching

- Module Teaching  
2025 Conformal Prediction © Doctoral Training Centre, University of Oxford  
Teaching Assistant  
2022–23 Foundations of Statistical Inference (George Deligiannidis) © Department of Statistics, University of Oxford  
2016 Matrices and Matrix Calculations (John Gilbert) © Department of Mathematics, University of Texas at Austin

## Academic Services

- Organizing Committee Member  
2024 ELLIS Robust Large Language Models Workshop [[web](#)]  
2024 ELLIS Robust Machine Learning Workshop [[web](#)]  
2023 ELLIS Robust Machine Learning Workshop [[web](#)]  
Reviewer  
NeurIPS, ICML, ICLR, AISTATS, TMLR

## Honors and Awards

- 2021– Clarendon Fund Scholarship, University of Oxford *(only 1 awardee from the Department of Statistics)*  
2018 College Scholar, College of Natural Sciences, University of Texas at Austin *(only top 20% awardees)*  
2017–18 Out-of-State Tuition Waiver, College of Natural Sciences, University of Texas at Austin *(only 5-7 awardees)*  
2017–18 Thomas and Elizabeth Merner Scholarship, College of Natural Sciences, University of Texas at Austin  
2017–18 Angus G. and Erna Pearson Endowed Scholarship, Department of Computer Science, University of Texas at Austin  
2017 College Scholar, College of Natural Sciences, University of Texas at Austin *(only top 20% awardees)*  
2016–17 Motorola Endowed Scholarship, Department of Computer Science, University of Texas at Austin  
2015–16 Angus G. and Erna Pearson Endowed Scholarship, Department of Computer Science, University of Texas at Austin  
2015 TIDES FRI Summer Research Fellowship, College of Natural Sciences, University of Texas at Austin  
2014–15 Freshman Scholarship, College of Natural Sciences, University of Texas at Austin  
2014–15 Schein Memorial Scholarship, Department of Computer Science, University of Texas at Austin