Gaurang Sriramanan

(+1)240-733-6585
gaurangs@umd.edu

Research Interests

Adversarial Attacks & Defenses, AI Reliability, Security and Robustness in Deep Learning

Education

2021–Present **PhD in Computer Science**, *University of Maryland, College Park*.
Research Advisor: Prof. Soheil Feizi
Anticipated Graduation Date: 2026
CGPA - 4.0/4.0

2020–2021 **Master of Science (Research) in Mathematics**, *Indian Institute of Science, Bangalore*.
CGPA - 9.4/10

2020 **Research Assistant**, *Video Analytics Lab (VAL), Indian Institute of Science, Bangalore*.

2016–2020 **Bachelor of Science (Research) in Mathematics**, *Indian Institute of Science, Bangalore*.
CGPA - 9.3/10

Selected Publications

2024 Vinu Sankar, Shoumik S*, **Gaurang Sriramanan***, Priyatham K, Atoosa C, Soheil Feizi, Fast Adversarial Attacks on Language Models In One GPU Minute, arXiv Preprint

- Developed an efficient class of Beam Search based adversarial attacks (BEAST) on Language Models for jailbreaking, eliciting hallucinations, and privacy attacks in one GPU minute
- With interpretable parameters, BEAST enables attackers to balance attack speed, success rate, and the readability of adversarial prompts

2023 Sriram B*, **Gaurang Sriramanan***, Vinu Sankar, Soheil Feizi, Exploring Geometry of Blind Spots in Vision models, **NeurIPS 2023**, **(Spotlight Paper, top 3%)**

- Developed Level Set Traversal algorithm that iteratively explores regions of high confidence with respect to the input space using orthogonal components of the local gradients
- By exploring the level sets of common vision models, we discover blindspot inputs misclassified with high confidence, and uncover a star-like connected substructure for superlevel sets

2022 **Gaurang Sriramanan**, Maharshi G, Soheil Feizi, Toward Efficient Robust Training against Union of Lp Threat Models, **NeurIPS 2022** and Oral Paper at AML Workshop at ICML 2022 (top 4%)

- Developed the first L1 robust model trained solely with single-step attacks using a curriculum
- Extends to robust training against union of Lp threat models, generalizes to unseen adversaries

2022 Sravanti A*, Samyak J*, **Gaurang Sriramanan**, R Venkatesh Babu, Towards Achieving Adversarial Robustness Beyond Perceptual Limits, **ECCV 2022**

- Developed Oracle-Aligned Adversarial Training to achieve state-of-the-art robustness within larger perturbation constraint sets, modelling real-world adversaries using LPIPS distance

2021 **Gaurang Sriramanan***, Sravanti A*, Arya B, R Venkatesh Babu, Towards Efficient and Effective Adversarial Training, **NeurIPS 2021**

- Developed Nuclear Norm Adversarial Training (NuAT), achieving state-of-the-art robustness amongst efficient defenses and Hybrid-NuAT to mitigate catastrophic failure from robust-overfitting

* Equal Contribution authors

2020 **Gaurang Sriramanan***, Sravanti A*, Arya B, R Venkatesh Babu, Guided Adversarial Attack for Evaluating and Enhancing Adversarial Defenses, **NeurIPS 2020 (Spotlight Paper, top 4%)**

- Developed the Guided Adversarial Margin Attack (GAMA), a state-of-the-art L-infinity attack that reliably determines the true robustness of deep networks
- Developed Guided Adversarial Training (GAT), that achieves state-of-the-art robustness among single-step adversarial defense methods

Project and Thesis Contributions

2020–2021 **Master's Thesis Project**, *Video Analytics Lab (VAL), IISc.*
Research Advisor: Prof. Venkatesh Babu
Thesis topic: Efficient and Effective training of Adversarially Robust Networks

2019–2020 **Undergraduate Thesis Project**, *Video Analytics Lab (VAL), IISc.*
Research Advisor: Prof. Venkatesh Babu
Thesis topic: Evaluating and Enhancing Adversarial Defenses

2018 **Summer Internship**, *Machine and Language Learning Lab (MALL), IISc.*
Research Advisor: Prof. Partha Pratim Talukdar
Deep Learning Theory, Semi-Supervised Learning, Orthogonal embeddings and GCNs for Graph Transduction

2018 **Summer Internship**, *Mathematics Department, IISc.*
Research Advisor: Prof. Kaushal Verma
Topics in Banach and Hilbert Space Theory including Spectral theory, Diagonalisation of Self-Adjoint Compact Operators, Hardy Spaces, Classical Fourier series etc.

Activities

2021–2023 **Reviewer** for ICCV 2021, CVPR 2022, ECCV 2022, ACML 2022, NeurIPS 2022/23, ICLR 2023, JMLR 2022/23, **Outstanding Reviewer Award** recipient at CVPR 2022, ECCV 2022, ICLR 2023

2020–2023 Served on the **Program Committee** of various Deep Learning workshops on Adversarial Robustness, Security and Socially Responsible ML: ECCV 2020 2022, ICLR 2021 2022, CVPR 2021 2022, ICML 2021 2022, ICCV 2021, 2023

2022 **Invited Speaker** at the TrustML Young Scientist Seminars hosted by the RIKEN Center for Advanced Intelligence Project (AIP) Japan

2018 **Technical Coordinator**, Pravega Laser-Tag: Responsible for overseeing the entire process, from ideation to the implementation of sensor and system electronics, server-client program code, and their integration into wearable suits for the Laser-Tag Event at Pravega, IISc's annual science festival

2017 **Co-founder, Singularity, the Undergraduate Engineering Club of IISc**: Dealt with the designing of hobby electronics projects and organised tutorial sessions for fellow batch-mates

Academic Achievements

2022 Recipient of the **Neural Information Processing Systems (NeurIPS) Scholar Award** 2022

2021-2023 Recipient of the **Dean's Fellowship** and **Chair's Fellowship**, awarded by University of Maryland

2021 Accepted to attend the highly competitive **Oxford Machine Learning Summer School** 2021

2020 Selected as a **Finalist** in The **Qualcomm Innovation Fellowship (QIF)** 2020

2019 Among the 100 students selected worldwide to attend the prestigious **Cornell Maryland Max-Planck Pre-Doctoral Research School (CMMRS)** 2019, at MPI-SWS, Saarbruecken, Germany

2016–2021 Recipient of the **Kishore Vaigyanik Protsahan Yojana (KVPY)** Scholarship, awarded by The Government of India

2015–2016 Secured **97.4%** aggregate, Centum in Computer Science in CBSE Class XII board examinations

2014-2016 Recipient of the **National Talent Search Examination (NTSE)** Scholarship, awarded by The Government of India

Relevant Coursework

Machine Learning	Statistical Pattern Recognition, Foundations of Deep Learning, Deep Learning: Theory and Practice, Advanced Numerical Optimization, Data Analytics, Information Theory, Scientific Computing
Mathematics	Calculus on Manifolds, Linear Algebra, Probability and Statistics, Real and Complex Analysis, Number Theory, Abstract Algebra, Topology, Measure Theory, Ordinary and Partial Differential Equations, Functional Analysis, Algebraic Topology

Programming Frameworks

Python, PyTorch, TensorFlow, Keras, Numpy, Pandas, Docker, OpenCV, C, C++