

Vinu Sankar Sadasivan

Final year PhD candidate
Department of Computer Science
The University of Maryland, College Park
[Research Interests](#) — AI/ML Security & Privacy, GenAI

[vinusankars.github.io](#)
vinu@umd.edu
[Google Scholar](#)

EDUCATION

The University of Maryland, College Park Ph.D. & M.S. in CS advised by Prof. Soheil Feizi	<i>Aug '21 – May '25 (Expected)</i> GPA - 4.00/4.00
Indian Institute of Technology, Gandhinagar B. Tech. in CSE [🏆 Director's Silver Medalist]	<i>Jul '16 – Jul '20</i> GPA - 9.21/10.00

INVITED TALKS

MLOps Podcast – Red-teaming for AI	<i>Dec '24</i>
UK AI Safety Institute – How to Jailbreak AI Efficiently?	<i>Nov '24</i>
US Securities and Exchange Commission – Can AI-Generated Content be Reliably Detected?	<i>May '24</i>
Amazon AWS Responsible AI – Fast Adversarial Attacks on Language Models	<i>Apr '24</i>
Google Research – Hardness of AI Text Detection	<i>Nov '23</i>

RESEARCH EXPERIENCES

Google Research → DeepMind, Mountain View <i>PhD Student Researcher (full-time until Jan '25)</i>	<i>Sep '24 – May '25</i> Manager: Dr. Lun Wang
Fundamental AI Research, Meta, Paris <i>Research Scientist Intern</i>	<i>May – Aug '24</i> Managers: Dr. Matthijs Douze , Dr. Jakob Verbeek
University of Maryland <i>Research Assistant in CS</i>	<i>Aug '21 – Present</i> Advisor: Prof. Soheil Feizi
IIT Gandhinagar <i>Junior Research Fellow in CSE</i>	<i>Aug '20 – Jul '21</i> Advisor: Prof. Anirban Dasgupta
California Institute of Technology <i>Undergraduate Research Fellow in Astronomy Department</i>	<i>May – Jul '19</i> Advisor: Dr. Ashish Mahabal
Microsoft Research India <i>Research Intern in Machine Learning and Optimization Group</i>	<i>Jan – Apr '19</i> Managers: Dr. Harsha Simhadri & Dr. Prateek Jain
Indian Institute of Science <i>Research Intern at Spectrum Lab for Signal Processing</i>	<i>May – Jul '17, Dec '17, Feb '18, May – Jul '18</i> Advisor: Prof. Chandra Seelamantula

RESEARCH PAPERS

* equal contribution

LLM-Check: Investigating Detection of Hallucinations in Large Language Models

G Sriramanan, S Bharti, **VS Sadasivan**, S Saha, P Kattakinda, S Feizi
Accepted at Conference on Neural Information Processing Systems (NeurIPS) 2024. [\[PDF\]](#)

DREW: Towards Robust Data Provenance by Leveraging Error-Controlled Watermarking

M Saberi, **VS Sadasivan**, A Zarei, H MahdaviFar, S Feizi
Preprint on arXiv. June, 2024. [\[PDF\]](#)

Fast Adversarial Attacks on Language Models In One GPU Minute

VS Sadasivan, S Saha*, G Sriramanan*, P Kattakinda, A Chegini, S Feizi
Accepted at International Conference on Machine Learning (ICML) 2024. [\[PDF\]](#)
[Media Coverage](#) [📰 The Register](#)

Can AI-Generated Text be Reliably Detected?

VS Sadasivan, A Kumar, S Balasubramanian, W Wang, S Feizi
Accepted at Transactions on Machine Learning Research (TMLR) 2025. [\[PDF\]](#)
[Media Coverage](#) [📰 Nature](#), [Washington Post](#), [Wired](#), [New Scientist](#), [The Register](#), [TechSpot](#)

Robustness of AI-Image Detectors: Fundamental Limits and Practical Attacks

M Saberi, **VS Sadasivan**, K Rezaei, A Kumar, A Chegini, W Wang, S Feizi

Accepted at International Conference on Learning Representations (ICLR) 2024. [PDF]

[Media Coverage](#)  [Wired](#), [The Verge](#), [MIT Technology Review](#), [Bloomberg](#), [The Register](#)

Exploring Geometry of Blind Spots in Vision Models

S Balasubramanian*, G Sriramanan*, **VS Sadasivan**, S Feizi

Accepted [[spotlight](#) ☆] at Conference on Neural Information Processing Systems (NeurIPS) 2023. [PDF]

Provable Robustness for Streaming Models with a Sliding Window

A Kumar, **VS Sadasivan**, S Feizi

Preprint on arXiv. March, 2023. [PDF]

CUDA: Convolution-based Unlearnable Datasets

VS Sadasivan, M Soltanolkotabi, S Feizi

Accepted at Computer Vision and Pattern Recognition Conference (CVPR) 2023. [PDF]

Statistical Measures For Defining Curriculum Scoring Function

VS Sadasivan, A Dasgupta

Accepted [[spotlight](#) ☆] at SubSetML Workshop at International Conference on Machine Learning (ICML) 2021. [PDF]

Shallow RNN: Accurate Time-series Classification On Resource Constrained Device

D Dennis, DAE Acar, V Mandikal, **VS Sadasivan**, V Saligrama, HV Simhadri, P Jain


Accepted at Conference on Neural Information Processing Systems (NeurIPS) 2019. [PDF]


High Accuracy Patch-Level Classification Of Wireless Capsule Endoscopy Images Using A Convolutional Neural Network

VS Sadasivan, CS Seelamantula

Accepted at IEEE International Symposium on Biomedical Imaging (ISBI) 2019. [PDF]

AWARDS AND HONORS

[Kulkarni Fellowship Awardee](#)  at University of Maryland in 2023.

[Notable reviewer](#)  top ~1% reviewer in ICLR 2023.

[Director's Silver Medalist](#)  CSE, IIT Gandhinagar in 2020.

[Special mention for poster](#) Undergraduate Research Conclave, IIT Gandhinagar in 2019.

[Summer Undergraduate Research Fellowship](#)  Caltech in 2019 (awarded ~ 6,350 USD).

[Kerala State Topper](#), [Regional Mathematics Olympiad](#) in 2014.

[KVPY awardee](#) by Government of India in 2016. Ranked 85 out of ~ 100,000 in the country.

[NTSE scholar](#) awarded by Government of India in 2012.

SERVICES & TEACHING

Reviewer for prominent machine learning conferences such as ICML 2021, NeurIPS 2022, ICLR 2023 ([Notable reviewer](#)), NeurIPS 2023, ICML Neural Compression Workshop 2023, ICML 2024, TACL.

Teaching assistant for CMSC720: Foundations of Deep Learning (Spring 2024), CMSC 422: Introduction to Machine Learning (Fall 2021), and CMSC 320: Introduction to Data Science (Spring 2022) at UMD.

Peer-assisted learning mentor at IIT Gandhinagar, helping freshmen who found it difficult to cope with their academic workload.

RESEARCH REPORTS

OSSuM: A Gradient-Free Approach For Pruning Neural Networks At Initialization

VS Sadasivan, J Malaviya, and A Dasgupta [PDF]

Improved Generalized Adaptive Exponential Functional Link Network Approximates

VS Sadasivan, SS Bhattacharjee, V Patel, and NV George [PDF]

FPGA-Based Area, Power, and Latency Optimized Approximate Multipliers For Neural Networks

VS Sadasivan, CK Jha, and J Mekie [PDF]