

# 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](mailto:vinu@umd.edu)  
[Google Scholar](#)

---

## EDUCATION

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

---

## INVITED TALKS

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

---

## RESEARCH EXPERIENCES

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

---

## 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  
Preprint on arXiv. March, 2023. [\[PDF\]](#)  
[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]