Ashish Hooda

✉ *ashish1995hooda@gmail.com*
🌐 *pages.cs.wisc.edu/ hooda*
AshishHoodaIITD

Research Interests

Security & Privacy, Post-training Language Models

Education

2019 – 2025 **Ph.D.**, Computer Sciences *(Minor in Electrical Engineering)*, UW-Madison

2014 – 2018 **B.Tech. (Hons.)**, Electrical Engineering *(Minor in Computer Science)*, IIT Delhi

Experience

Summer 2023 **Research Internship**, *Google, with Mihai Christodorescu and Miltiadis Allamanis*
Fall 2024 Worked on evaluating program semantics understanding of Large Language Models for Code. Built the first framework for counterfactual evaluation of code completion models.

- Paper accepted to ICML 2024.

Summer 2022 **Applied Scientist Internship**, *Amazon AWS*
Developed an efficient Graph Neural Network training framework that scales to billion node scale graphs. Utilized residual quantization for efficiency without sacrificing precision.

2018 – 2019 **Software Engineer**, *Microsoft India*
Developed a task distribution model for agent assignments in the Omnichannel CRM team.

Invited Talks, Research in News

Mar 2025 **More Fun(-Tuning) in the New World**, *Ars Technica*, [link]

Mar 2025 **LLM hackers are using its own tools against it**, *Android Authority*, [link]

Feb 2025 **PRP: Attacking LLM Guard-Rails**, *UCSD*

Oct 2024 **Counterfactual Analysis for Code Predicates**, *Jet Brains Research*

July 2024 **Preemptive Monitoring in E2E Encrypted Services**, *Internet Society*, [link]

June 2024 **Is Detection A Viable Defense For against Attacks?**, *Visa Research*

Nov 2023 **Do Code LLMs understand program semantics?**, *Deepmind ML4Code Team*

Oct 2023 **Do Stateful Defenses Work Against Black-Box Attacks?**, *Google AI Red Team*

Aug 2023 **Deepfake Detection Against Adaptive Attackers**, *Google AI Red Team*

Selected Awards and Services

- Reviewer for ICML, ICLR, NeurIPS, Usenix
- Student Travel Award, NDSS 2024
- Doctoral Consortium Award, WACV 2024
- Mentor at Individualized Cybersecurity Research Mentoring (iMentor), CCS 2023
- Co-Mentor at Wisconsin Science and Computing Emerging Research Stars (WISCERS), UW-Madison 2022
- Runner up in CS Research Symposium, UW-Madison 2022
- Regionals at ACM International Collegiate Programming Contest (ICPC), 2017
- Runner-up at Microsoft CODE-FUN-DO Hackathon, 2015
- Ranked 4 in Board Examination among 2 million students
- Ranked 17 in Joint Entrance Exam (JEE) among 1 million students

Publications

(* indicates joint first authorship)

IEEE S&P 2025 **Fun-tuning: Characterizing the Vulnerability of Proprietary LLMs to Optimization-based Prompt Injection Attacks via the Fine-Tuning Interface**
Andrey Labunets, Nishit Pandya, Ashish Hooda, Xiaohan Fu, Earlence Fernandes
46th IEEE Symposium on Security and Privacy, Acceptance Rate: 14.8% [Paper]

ICLR 2025 **Functional Homotopy: Smoothing Discrete Optimization Via Continous Parameters for LLM Jailbreak Attacks**
Zi Wang*, Divyam Anshuman*, Ashish Hooda, Yudong Chen, Somesh Jha
International Conference on Learning Representations, Acceptance Rate: 31.24% [Paper]

ACL 2024 **PRP: Propagating Universal Perturbations to Attack LLM Guard-Rails**
Ashish Hooda*, Neal Mangaokar*, Jihye Choi, Shreyas Chandrashekaran, Kassem Fawaz, Somesh Jha, Atul Prakash
Association for Computational Linguistics, Acceptance Rate: 21.3% [Paper][Code]

SATA 2024 **PolicyLR: A LLM compiler for Logic Representation of Privacy Policies**
Ashish Hooda, Rishabh Khandelwal, Prasad Chalasani, Kassem Fawaz, Somesh Jha
Safe & Trustworthy Agents Workshop, NeurIPS [Paper]

ICML 2024 **Do Large Code Models Understand Programming Concepts? Counterfactual Analysis for Code Predicates**
Ashish Hooda, Mihai Christodorescu, Miltiadis Allamanis, Aaron Wilson, Kassem Fawaz, Somesh Jha
International Conference on Machine Learning, Acceptance Rate: 27.5% [Paper]

WACV 2024 **D4: Detection of Adversarial Diffusion Deepfakes Using Disjoint Ensembles**
Ashish Hooda*, Neal Mangaokar*, Ryan Feng, Kassem Fawaz, Somesh Jha, Atul Prakash
IEEE/CVF Winter Conference on Applications of Computer Vision, Acceptance Rate: 41.41% [Paper] [Code]

NDSS 2024 **Experimental Analyses of Physical Surveillance Risks in Client-Side Content Scanning**
Ashish Hooda, Andrey Labunets, Tadayoshi Kohno, Earlence Fernandes
Network and Distributed System Security Symposium, Acceptance Rate: 19.9% [Paper]

AdvML 2023 **Theoretically Principled Trade-off for Stateful Defenses against Query-Based Black-Box Attacks**
Ashish Hooda*, Neal Mangaokar*, Ryan Feng, Kassem Fawaz, Somesh Jha, Atul Prakash
2nd AdvML Frontiers Workshop, ICML [Paper]

CCS 2023 **Stateful Defenses for Machine Learning Models Are Not Yet Secure Against Black-box Attacks**
Ashish Hooda*, Ryan Feng*, Neal Mangaokar*, Kassem Fawaz, Somesh Jha, Atul Prakash
ACM Conference on Computer and Communications Security, Acceptance Rate: 19.15% [Paper][Code]

IMWUT 2022 **SkillFence: Systems Approach to Mitigate Voice-Based Confusion Attacks**
Ashish Hooda, Matt. Wallace, Kushal Jhunjhunwalla, Earlence Fernandes, Kassem Fawaz
ACM Interactive, Mobile, Wearable and Ubiquitous Technologies, Acceptance Rate≈ 20% [Paper]

CVPR 2021 **Invisible Perturbations: Physical Adv Examples Exploiting the Rolling Shutter Effect**
Ashish Hooda*, Athena Sayles*, Mohit Gupta, Rahul Chatterjee, Earlence Fernandes
Conference on Computer Vision and Pattern Recognition, Acceptance Rate: 23.7% [Paper][Code]

Preprint **Synthetic Counterfactual Faces**
Guruprasad V Ramesh, Ashish Hooda, Harrison Rosenberg, Shimaa Ahmed, Kassem Fawaz
arXiv:2407.13922 [Paper]