

Suhas Kotha

suhask@andrew.cmu.edu | [homepage](#)

EDUCATION

Stanford University, PhD Computer Science 2024 - current

Carnegie Mellon University, MS+BS Computer Science 2020 - 2024

Advised by Aditi Raghunathan, MS fully funded by Apple

WORK EXPERIENCE

Jane Street, Quantitative Trading Intern 5/22 – 8/22

Improved algorithmic trading strategies for bonds and commodities

NASA (SpaceML), Machine Learning Intern 6/20 – 1/21

Proposed self-supervised scene classification for satellite imagery, deployed at NASA

TEACHING

- Instructor for Introductory Type Theory Spring '22, Fall '22
- TA Functional Programming Spring '21, Spring '22
- TA Algorithm Design/Analysis Fall '22
- TA Math Foundations for CS Fall '21

SERVICE

- Reviewer for NeurIPS (2023-), ICML (2023-), ICLR (2023-), TMLR (2023-)
- Co-founded CMU Undergraduate Theory Community

AWARDS

- Stanford School of Engineering Fellowship 2024 - 2025
- Putnam Top 500, USA Math Olympiad, USA Computing Olympiad Platinum

SELECTED COURSEWORK

(PhD) Foundations of Modern ML
(PhD) Analytic Convex Geometry
(PhD) Algorithms for Big Data
(PhD) Uncertainty Quantification
(PhD) Intermediate Statistics

(PhD) Probabilistic Combinatorics
(PhD) Automated Reasoning & SAT
(PhD) Convex Optimization
(PhD) Generative Computer Vision
Parallel Computer Architecture

PUBLICATIONS

Jailbreaking is Best Solved by Definition

Taeyoun Kim*, **Suhas Kotha***, Aditi Raghunathan
Preprint 2024

Repetition Improves Language Model Embeddings

Jacob Springer, **Suhas Kotha**, Daniel Fried, Graham Neubig, Aditi Raghunathan
Preprint 2024

A Safe Harbor for AI Evaluation and Red Teaming

Shayne Longpre, Sayash Kapoor, Kevin Klyman, Ashwin Ramaswami, Rishi Bommasani, Borhane Blili-Hamelin, Yangsibo Huang, Aviya Skowron, Zheng-Xin Yong, **Suhas Kotha**, Yi Zeng, Weiyan Shi, Xianjun Yang, Reid Southen, Alexander Robey, Patrick Chao, Diyi Yang, Ruoxi Jia, Daniel Kang, Sandy Pentland, Arvind Narayanan, Percy Liang, Peter Henderson
ICML 2024, Position

Understanding Catastrophic Forgetting in Language Models via Implicit Inference

Suhas Kotha, Jacob Springer, Aditi Raghunathan
ICLR 2024

Provably Bounding Neural Network Preimages

Suhas Kotha*, Christopher Brix*, Zico Kolter, Krishnamurthy Dvijotham[†], Huan Zhang[†]
NeurIPS 2023, Spotlight

TALKS

- Understanding Catastrophic Forgetting in Language Models via Implicit Inference
 - Brave Machine Learning Research
- Provably Bounding Neural Network Preimages
 - UIUC Formal Verification Group
 - Carnegie Mellon Undergraduate Theory Community