

# ALAN SUN

AWSUN@CMU.EDU • ALANSUN17904.GITHUB.IO

## EDUCATION

<b>Carnegie Mellon University</b> Master of Science in Computer Science	December 2025 GPA 4.0/4.0
<b>Dartmouth College</b> Bachelor of Arts, Computer Science, Mathematics, <i>magna cum laude</i> , High Honors	June 2024 GPA 3.9/4.0

## HONORS AND AWARDS

• U.S. National Science Foundation Graduate Research Fellowship	2025–2030
• G-Research Grant for Early Career Researchers	May 2025
• Max-Planck Institute for Software Systems Visiting Scholar Fellowship	June 2024
• John G. Kemeny Computing Prize for Innovation	May 2024
• Neukom Prize for Outstanding Undergraduate Research—First Prize	May 2024
• James O. Freedman Presidential Scholar	2022–2023
• Francis L. Town Prize for Achievement in Computer Science	December 2022
• Johns Hopkins Applied Physics Lab Intern Achievement Award for Technical Excellence	August 2022

## PUBLICATIONS

1. *Provably Tracking Equivalent Mechanistic Interpretations Across Neural Networks*  
**A. Sun**, M. Toneya.  
ICLR (2026) [[pdf](#)]
2. *Circuit Stability Characterizes Language Model Generalization*  
**A. Sun**  
ACL (2025) [[pdf](#)]
3. *Achieving Domain-Independent Certified Robustness via Knowledge Continuity*  
**A. Sun**, C. Ma, K. Ge, S. Vosoughi  
NeurIPS (2024) [[pdf](#)]
4. *Deciphering Stereotypes in Pre-Trained Language Models*  
W. Ma, H. Scheible, B. Wang, G. V., P. C., **A. Sun**, A. K., L. Wang, D. Yang, S. Vosoughi  
EMNLP (2023) [[pdf](#)]
5. *ThanosNet: A Novel Trash Classification Method Using Metadata*  
**A. Sun** and H. Xiao  
IEEE Big Data (2020) [[pdf](#)]

## TEACHING

• Optimization for Machine Learning (TA)	CMU
• Probability (Grader)	Dartmouth
• Real Analysis (Grader)	Dartmouth
• Algorithms (TA)	Dartmouth
• Multivariate Calculus (Peer Tutor)	Dartmouth
• Digital Electronics (TA)	Dartmouth

## SERVICE

• ICLR, Reviewer	2025, 2026
• NeurIPS, Reviewer	2024, 2025
• ACL Student Research Workshop, Reviewer	2025
• {MATH-AI, ATTRIB, MechInterp} Workshop @ NeurIPS, Reviewer	2024, 2025