Alan Sun

awsun@cmu.edu • alansun17904.github.io

EDUCATION

Carnegie Mellon University — Expected Dec. 2025
Master of Science in **Computer Science**

Dartmouth College — June 2024
Bachelor of Arts, Computer Science, Mathematics

AWARDS

- U.S. National Science Foundation Graduate Research Fellowship — 2025–2030
- G-Research Grant for Early Career Researchers — May 2025
- Visting Scholar Fellowship @ Max-Planck Institute for Software Systems — 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 — Dec. 2022

PUBLICATIONS

Provably Tracking Equivalent Mechanistic Interpretations Across Neural Networks
A. Sun, M. Toneva
In Review @ ICLR (2026)

Circuit Stability Characterizes Language Model Generalization
A. Sun
In Proceedings of ACL 2025 (Main)

Achieving Domain-Independent Certified Robustness via Knowledge Continuity
A. Sun, C. Ma, K. Ge, S. Vosoughi
In Proceedings of NeurIPS 2024

Deciphering Stereotypes in Pre-Trained Language Models
W. Ma, H. Scheible, B. Wang, G. V., P. C., **A. Sun**, A. K., L. Wang, D. Wang, S.Vosoughi
In Proceedings of EMNLP 2023 (Main)

ThanosNet: A Novel Trash Classification Method Using Metadata
A. Sun and H. Xiao
In Proceedings of IEEE Big Data 2020

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 (2025, 2026) — Reviewer
- NeurIPS (2024, 2025); Workshops: MATH-AI (2024), ATTRIB (2024), MechInterp (2025) — Reviewer
- ACL Student Research Workshop (2025) — Reviewer