Alan Sun

17904 Wheatridge Dr.
Germantown, Maryland, USA 20874
+1 301-316-8199 • alansun@andrew.cmu.edu

EDUCATION

Carnegie Mellon University Aug. 2024 – Dec. 2025
Master of Science in Computer Science

Dartmouth College June 2024
Bachelor of Arts, Computer Science, Mathematics, *magna cum laude*, high honors
Honors Thesis: Achieving Domain-Independent Certified Robustness via *Knowledge Continuity*
GPA: 3.90/4.00, **Major GPA:** 3.99/4.00

Relevant Coursework: Machine Learning and Stat. Data Analysis, Information Theory, Probability (Honors), Real Analysis (Honors), Metric Spaces and Measure Theory[*], Probability and Statistical Inference, Computer Vision, Data-Driven Uncertainty Quantification, Algorithms, Differential Equations, Computer Architecture, Complex Analysis, Randomized Algorithms, Software Design and Implementation, Multivariable Calculus, Digital Electronics, Abstract Algebra (Honors), Intro. to Computability

RESEARCH EXPERIENCE

BrAIN (Max-Planck Institute for Software Systems) June 2024 – Present
Adivsor: Mariya Toneva

Mind, Machines, Society Group (Dartmouth College) Sept. 2022 – June 2024
Advisors: Soroush Vosoughi, Chiyu Ma, Weicheng Ma

- **Domain-Independent Certified Robustness via *Knowledge Continuity*** *(Sept. 2023 – June 2024); First Author*
  Formulated a new metric, *knowledge continuity*, which when minimized provable minimizes adversarial robustness while not limiting the expressiveness of the hypothesis class. Constructively demonstrating that robustness and accuracy are not at odds. *Under review at ICML 2024.*
- **Automated Soft Modular Robot Design** *(March 2023 – June 2023); Co-Author*
  Created language models trained using policy-gradient methods to automatically assemble soft-lattice robots based on user-specified task and robot deployment environment. This work was done in collaboration with the Dartmouth Robotics Lab.

Fu Lab (Dartmouth College) Aug. 2023 – Jan. 2024
Advisor: Feng Fu

- **Mechanistic Interpretation with Large Language Models** *(Aug. 2023 – Present)*
  Developing novel methods to mechanistically interpret neural networks trained to play iterated prisoner's dilemma. The goal is to generate (end-to-end) natural language explanations both at the neuronal and network level using large language models.
- **Information Bottleneck Theory to Explain Adversarial Attacks** *(March 2022 – June 2022)*
  Used information bottleneck theory and the information plane to characterize and explain adversarial robustness of neural networks based on their activation functions.

Applied Physics Lab (Johns Hopkins University), Threat Analytics Group June 2022 – Sept. 2022
Advisors: Sarah Prata, Alex Memory

- Built and designed graph neural networks to detect and predict trends of toxic posts and comments throughout Reddit communities.
- Worked with agency responsible for generating the data on collecting techniques to avoid implicit biases.
- Performed exploratory data analysis on online forum data and proposed novel, graph-based metrics for quantifying post-comment relationships.

[*] Guided study with Professor Erik van Erp

HONORS AND AWARDS

- **John G. Kemeny Computing Prize for Innovation (2024).** Intended to encourage novel uses of computing by undergraduate Dartmouth students. Rewards students who produce original, creative, well-designed, and well-implemented computer programs.
- **Neukom Prize for Outstanding Undergraduate Research–First Prize (2024).** Recognizes outstanding graduate/undergraduate research in computational sciences at Dartmouth.
- **Francis L. Town Prize for Achievement in Computer Science (2023).** Presented annually to one exceptional student in computer science at Dartmouth.
- **James O. Freedman Presidential Scholar (2023).** Provides funding for undergraduate students to work as research assistants with Dartmouth faculty.
- **Goldwater Scholarship Program Nominee (2023).** One of five students nominated to represent Dartmouth in the national Barry Goldwater Scholar selection.
- **Dartmouth College Second Honors Group (2023; 2022).** Awarded annually to top 15% of all undergraduate students.
- **JHU/APL Achievement Award for Technical Excellence (2022).** Given to interns who make meaningful technical contributions to their projects, produce work of exception quality.
- **Bronze Medal in Options Trading at UChicago Trading Competition (2022).** Created a real-time algorithm which makes markets for options sensitive to catastrophic events.
- **Silver Medal for Kaggle Toxic Comment Classification Challenge (2020).** Achieved a top 4% finish, 72nd out of 1621 teams. Created language models to identify multilingual toxic comments using only English training set.

PUBLICATIONS

W. Ma, H. Scheible, B. Wang, G. Veeramachaneni, P. Chowdhary, A. Sun, A. Koulogeorge, L. Wang, S. Vosoughi. **Deciphering Stereotypes in Pre-Trained Language Models**. *Accepted to* 2023 Conference on Empirical Methods in Natural Language Processing.

A. Sun. **A Decentralized Insurance Exchange**," U.S. Patent 63/153,349 *Pending*. February 24, 2021.

A. Sun and H. Xiao. **ThanosNet: A Novel Trash Classification Method Using Metadata**, in 2020 IEEE International Conference on Big Data (Big Data), 2020, pp. 1394–1401. doi: 10.1109/BigData50022.2020.9378287.

TEACHING EXPERIENCE

Probability, Grader (2024).

Real Analysis, Grader (2024).

Algorithms, Undergraduate Teaching Assistant (2023).

Responsibilities: Attended class as a teaching aide. Hosted office hours once a week to answer student questions and help students on problem sets. Graded homework and in-class worksheets. (53 students).

Multivariate Calculus, Peer Tutor (2022).

Responsibilities: Hosted one-on-one sessions, where I clarified concepts and proofs from class.

Digital Electronics, Undergraduate Teaching Assistant (2022).

Responsibilities: Graded quizzes, homework, and in-class worksheets. Hosted office hours once a week to answer student questions on problem sets. Mentored students on their culminating term project. (35 students).

STUDENT MENTEES

Ava Carlson (BA, Dartmouth College)
Kenneth Ge (BS, Carnegie Mellon University)
Chikwanda Chisha (BA, Dartmouth College, E.E Just Summer Research Intern)

TECHNICAL SKILLS

Programming Languages: Python, C, Java, Bash.
Frameworks/Technologies: PyTorch, Scikit-Learn, HuggingFace, Docker, Docker Compose, NumPy, Pandas, Git, Django.

REFERENCES

Soroush Vosoughi, Assistant Prof. of Computer Science, Dartmouth College, soroush.vosoughi@dartmouth.edu
Erik van Erp, Associate Prof. of Mathematics, Dartmouth College, erikvanerp@dartmouth.edu
Yoonsang Lee, Assistant Prof. of Mathematics, Dartmouth College, yoonsang.lee@dartmouth.edu
Chiyu Ma, PhD Student in Computer Science, Dartmouth College, chiyu.ma@dartmouth.edu