Samuel Sasaki

samuel.sasaki@vanderbilt.edu | sammsaski.github.io | github.com/sammsaski | linkedin.com/in/sam-sasaki

Research Interests

My research is at the intersection of formal methods and machine learning, with a focus on improving the safety and reliability of modern AI systems. I am motivated by the question of what verification can contribute to AI safety in an era of frontier models, where worst-case formal guarantees increasingly give way to questions about the underlying quality of learned behavior. I am also drawn to neurosymbolic approaches, where symbolic structure can provide interpretability and safety properties that purely learned systems struggle to support, and to the broader question of how training procedures shape model internals.

Education

Vanderbilt University, Nashville, TN — June 2023 – Present
Ph.D. in Computer Science (Advisor: Taylor T. Johnson)

Pomona College, Claremont, CA — Sept. 2018 – Dec. 2022
B.A. in Computer Science and Mathematics

Publications

* Denotes equal contribution.

Conference Papers

[C1] **Samuel Sasaki**, Ben Wooding, Taylor T. Johnson, "Learning Reach-Set Geometry for Tighter Probabilistic Neural Network Verification," *The Fortieth Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2026.

[C2] Anne M. Tumlin*, **Samuel Sasaki***, Ben Wooding, Diego Manzanas Lopez, Muhammad Usama Zubair, Navid Hashemi, Hongchao Zhang, Waseem Abbas, Ipek Oguz, Meiyi Ma, Taylor T. Johnson, "NNV3: Expanding Neural Network Verification to New Architectures and Domains," *The 24th International Symposium on Automated Technology for Verification and Analysis (ATVA)*, December 2026.

[C3] **Samuel Sasaki**, Ben Wooding, Hanchen David Wang, Anne M. Tumlin, Meiyi Ma, Taylor T. Johnson, "n2v: Neural Network Verification in Python," *The 9th International Symposium on AI Verification (SAIV)*, July 2026.

[C4] Quang Pham, Ben Wooding, Luke Nam, **Samuel Sasaki**, Taylor T. Johnson, "MetaMoE: Formal Verification of Compositional Robustness and Scalability of Mixture-of-Experts Architecture," *The 9th International Symposium on AI Verification (SAIV)*, July 2026.

[C5] Navid Hashemi, **Samuel Sasaki**, Ipek Oguz, Meiyi Ma, Taylor T. Johnson, "Scaling Data-Driven Probabilistic Robustness Analysis for Semantic Segmentation Neural Networks," *The Thirty-Ninth Annual Conference on Neural Information Processing Systems (NeurIPS)*, December 2025.

[C6] Preston K. Robinette, Thuy Dung Nguyen, **Samuel Sasaki**, Taylor T. Johnson, "Trigger-Based Fragile Model Watermarking for Image Transformation Networks," *30th European Symposium on Research in Computer Security (ESORICS)*, September 2025.

[C7] Diego Manzanas Lopez, **Samuel Sasaki**, Taylor T. Johnson, "NNV: A Star Set Reachability Approach," *The 8th International Symposium on AI Verification (SAIV)*, July 2025.

[C8] **Samuel Sasaki**, Diego Manzanas Lopez, Taylor T. Johnson, "Neurosymbolic Finite and Pushdown Automata: Improved Multimodal Reasoning versus Vision Language Models (VLMs)," *2nd International Conference on Neuro-symbolic Systems (NeuS)*, May 2025.

[C9] **Samuel Sasaki**, Diego Manzanas Lopez, Preston K. Robinette, Taylor T. Johnson, "Robustness Verification of Video Classification Neural Networks," *13th International Conference on Formal Methods in Software Engineering (FormaliSE)*, April 2025.

Workshop Papers

[W1] **Samuel Sasaki**, Ben Wooding, Taylor T. Johnson, Matthias Althoff, Luis Benet, Samuel Coogan, Marcelo Forets, Akash Harapanahalli, Lukas Koller, Tobias Ladner, Christian Schilling, Huan Zhang, and Xiangru Zhong, "ARCH-COMP26 Category Report: Artificial

Intelligence and Neural Network Control Systems (AINNCS) for Continuous and Hybrid System Plants," *13th International Workshop on Applied Verification of Continuous and Hybrid Systems (ARCH-COMP)*, 2026.

Preprints

[P1] Navid Hashemi, **Samuel Sasaki**, Diego Manzanas Lopez, Ipek Oguz, Meiyi Ma, Taylor T. Johnson, "Probabilistic Robustness Analysis in High Dimensional Space: Application to Semantic Segmentation Network," arXiv, `arxiv.org/abs/2509.11838v1`, November 2025.

Tools

`n2v` is an open-source Python adaptation and extension of the `NNV` MATLAB toolbox for formal verification of deep neural networks. Some new added features include broader GPU support, a more diverse collection of LP solvers, new falsification methods, and support for PyTorch models. `https://github.com/sammsaski/n2v/`

Talks & Presentations

Co-presenter, "The Verification of Neural Networks Competition (VNN-COMP): A Lab for Benchmark Proposers, Verification Tool Participants, and the Broader AI Community," Lab, *40th AAAI Conference on Artificial Intelligence (AAAI)*, January 2026.

Samuel Sasaki, Diego Manzanas Lopez, Taylor T. Johnson, "Neurosymbolic Automata: Improved Multimodal Reasoning versus Vision Language Models," *9th Workshop on Learning in Verification (LiVe)*, May 2025.

Academic Honors & Grants

Grants

VSS NSF Travel Award ($808)	2025
VMW Student Travel Scholarship — VMW and CAV 2024 ($1,740)	2024
Vanderbilt Graduate Travel Grant ($1,000)	2024
NCAA Postgraduate Scholarship ($10,000)	2023
Pomona College Summer Undergraduate Research Grant ($4,800)	2022

Awards & Honors

Pomona-Pitzer Male Student-Athlete of the Year	2022
Pomona College Scholar	2022

Research Experience

Vanderbilt University, Nashville, TN — June 2023 – Present
Research and Teaching Assistant

- Research assistant working on developing methods for efficiently verifying robustness properties of neural networks, specifically those used for video classification.

Pomona College, Claremont, CA — June 2022 – Dec. 2022
Undergraduate Research Assistant

- Creating and training a computer vision model to inform robot movement and designing the architecture for model deployment.
- Contributed to a project on identifying locations on the Pomona College campus using a convolutional neural network by performing data collection, defining the network architecture, and executing model training.

Teaching Experience

Vanderbilt University — 2023–2024
Teaching Assistant, introductory engineering and embedded systems courses

Vanderbilt University — April 2026
Guest Lecturer, CS6315: Automated Verification — "Introduction to Neural Network Verification"

Work Experience

Elementary — Jan. – June 2022
Software Engineer

- Improved reliability by updating the Python testing suite to handle product configurations in a simulated environment.

– Constructed an API to capture and write data originating from edge devices to a performance monitoring stream built on Redis and created a dashboard to display performance metrics of the hardware.
– Investigated performance of machine learning product based on run-time and latency metrics across the set of hardware specifications and motivated decision-making for future sprints with detailed reports of findings.

Elementary May – Aug. 2021
Software Engineer Intern

– Increased flexibility in testing by adding an Ansible workflow that could run multiple testing configurations (CPUs, devices) at a time, which provided a more robust and streamlined CI/CD pipeline.
– Initiated the development of an internal code coverage comparison tool written in Bash and Python and architected an automated workflow for it on GitHub that brought awareness of 5+ repositories with code coverage less than 90%.

Service

Conference organization — co-leading the AINNCS track, ARCH-COMP 2026 2026

Subreviewer — ICLR (2026), NeurIPS (2025), EMSOFT (2024, 2025, 2026), ICCPS (2024, 2025, 2026), UAI (2024), and KR (2024) 2024–2026

Artifact evaluation committee — QEST+FORMATS 2025 2025

Outreach — reviewer, *Young Scientist Journal* (YSJ), a peer-reviewed journal of high school research 2026

Last updated: October 19, 2026