Neil Kale

nkale@cs.cmu.edu | (508) 936-3455 | neilkale.github.io | www.linkedin.com/in/neil-kale/

Education

Carnegie Mellon University (CMU) **Pittsburgh, PA**
Master of Science in Machine Learning December 2025
GPA 4.24/4.0
Advisors Prof. Virginia Smith, Prof. Aditi Raghunathan
Committee Service MSML Admissions Committee, MSML Social Committee, SCS Master's Advisory Committee

Worcester Polytechnic Institute (WPI) **Worcester, MA**
Bachelor of Science in Computer Science, **2nd Major** in Mathematics, **Minor** in Robotics Engineering May 2024
GPA 4.0/4.0
Awards and Activities Graduation with High Distinction**,** National Academic Achievement (Upsilon Pi Epsilon), Meritorious Senior (Pi Mu Epsilon), Men's Varsity Rowing

Selected Publications

Multimodal AI

[1] Wu, C. H.*, **Kale, N.***, & Raghunathan, A. Mitigating modal imbalance in multimodal reasoning. In *2025 Conference on Language Modeling (CoLM).*
[2] **Kale, N.**, Zhang, C. B. C., Zhu, K., Aich, A., Rodriguez, P., Team, S. R., ... & Wang, Z. (2025). Reliable Weak-to-Strong Monitoring of LLM Agents. arXiv preprint arXiv:2508.19461. Under review at ICLR 2026.

Privacy and Auditing
[3] **Kale, N.**, ..., & Smith, V. (2025). Position: child safety necessitates new approaches to AI safety.
[4] Thaker, P., Hu, S., **Kale, N.**, Maurya, Y., Wu, Z. S., & Smith, V. (2025, April). Position: LLM unlearning benchmarks are weak measures of progress. In *2025 IEEE Conference on Secure and Trustworthy Machine Learning (SaTML)* (pp. 520-533). IEEE.
[5] Thaker, P.*, **Kale, N.***, Wu, Z. S., & Smith, V. (2025). Membership Inference Attacks for Unseen Classes. *arXiv preprint arXiv:2506.06488.* Under review at ICLR 2026.
[6] Hu, S., **Kale, N.**, Thaker, P., Fu, Y., Wu, S., & Smith, V. (2025). BLUR: A Benchmark for LLM Unlearning Robust to Forget-Retain Overlap. *arXiv preprint arXiv:2506.15699.*

Healthcare
[7] Gjuka, D., Adib, E., ..., **Kale, N.**, ..., Kwiatkowski, D., & Stone, E. (2023). Enzyme-mediated depletion of methylthioadenosine restores T cell function in MTAP-deficient tumors and reverses immunotherapy resistance. Cancer cell, 41(10), 1774-1787

Industry Experience

Scale AI (Hosted by **Christina Knight**) October 2025 - January 2026
Security and Policy Research Lab Fellow
Engaging with global policymakers to solve real-world problems in agent robustness, AI control, and frontier risk evaluation.

Scale AI (Hosted by **Sail Wang** & **Julian Michael**) May 2025 - August 2025
SEAL Research Scientist Intern
Developed novel chain-of-thought monitoring techniques to oversee superhuman agents with weaker LLM monitors.

GoDaddy Inc (Hosted by **Harsh Nilesh Pathak)**
Machine Learning Scientist Intern May 2024 - August 2024
Built GoDaddy's first agentic pipeline for CRM and prototyped an early agentic chatbot.

Harvard Medical School (Hosted by **David Kwiatkowski**) May 2021 – August 2023
Research Assistant
Analyzed 1,200 bladder cancer genomes/transcriptomes as the only computer scientist on a 20-person research effort over 2 years.

Teaching

Teaching Assistant for 10701 Introduction to Machine Learning (CMU), CS541 Deep Learning (WPI)

Relevant Coursework

Graduate 10715 Introduction to Machine Learning (A+), 36700 Probability and Statistics (A), 15789 Foundations of Modern Machine Learning (A+), 10708 Probabilistic Graphical Models (A+), 10725 Convex Optimization (A), 10703 Deep Reinforcement Learning (ongoing), 10718 Machine Learning in Practice (ongoing)

Undergraduate CS534 Artificial Intelligence (A), CS541 Deep Learning (A), MA4635 Statistical Learning (A)