Pin-Han (Hank) Huang

Ph.D. Student, Computer Science, University of California, Los Angeles
pinhank@ucla.edu | pinhank0121.github.io

Research Interests

I am interested in understanding the undesired behavior of ML models, with the goal of working towards an actionable understanding for building efficient and reliable AI systems. Recently, I have been working on data selection for efficient LLM post-training and harness optimization.

Education

University of California, Los Angeles — Los Angeles, CA
Ph.D. in Computer Science · Advisor: Prof. Baharan Mirzasoleiman — Sep. 2026 – Present

National Taiwan University — Taipei, Taiwan
B.S. in Computer Science · Advisors: Prof. Shang-Tse Chen & Prof. Hsuan-Tien Lin — Sep. 2021 – Aug. 2025

Research Experience

Computational Learning Lab, National Taiwan University — Taipei, Taiwan
Undergraduate Researcher, Research Assistant · Advisor: Prof. Hsuan-Tien Lin — Sep. 2024 – Aug. 2026

- **Representation Learning.** Building robust discriminative models leveraging generative representations. A project on better leveraging diffusion models to improve adversarial training is accepted to NeurIPS 2026.

Artificial Intelligence Security Lab, National Taiwan University — Taipei, Taiwan
Undergraduate Researcher, Research Assistant · Advisor: Prof. Shang-Tse Chen — Jul. 2023 – Aug. 2026

- **Certified Robustness.** Worked on methods to improve the certified robustness of vision models against adversarial attacks. A project on constructing better Lipschitz networks was presented at ICML 2025 (Spotlight).

CompLING Lab, University of Waterloo — Waterloo, Canada
Visiting Research Intern (Mitacs GRI) · Advisor: Prof. Freda Shi — Jun. 2025 – Aug. 2025

- **Multilingual VLMs & Object Hallucination.** Studied how multilingual capability affects object hallucination in MLLMs via mechanistic interpretability.

Publications

[1] **Expanding the Role of Diffusion Models for Robust Classifier Training**
Pin-Han Huang, Shang-Tse Chen†, Hsuan-Tien Lin†
NeurIPS 2026

[2] **Enhancing Certified Robustness via Block Reflector Orthogonal Layers and Logit Annealing Loss**
Bo-Han Lai*, Pin-Han Huang*, Bo-Han Kung*, Shang-Tse Chen
ICML 2025 (Spotlight)

*Equal contribution †Co-advisors

Awards and Scholarships

- **Mitacs Globalink Research Internship.** Mitacs, Canada — Jun. 2025
- **1st Prize, Undergraduate Research Seminar.** Dept. of CSIE, National Taiwan University — Jun. 2024
- **NSTC Undergraduate Research Grant.** National Science and Technology Council, Taiwan — Jun. 2024