

# Lei Hsiung

☎ (603) 507-3550 · ✉ Lei.Hsiung.GR@dartmouth.edu · 🏠 Hsiung.cc · 🌐 twweeb · 🎓 scholar

## RESEARCH INTERESTS

My current research focuses on **vision and language model behavior**. I aim to understand and advance LLMs and VLMs by examining how **post-training techniques** impact vision/language understanding capabilities and **reasoning behaviors**. Before that, I worked on improving the **robustness, efficiency, and trustworthiness** of machine learning models.

## EDUCATION

- **Dartmouth College** Sept. 2023 - Present  
*PhD in Computer Science*  
Hanover, NH
  - Advisors: Prof. Yaoqing Yang and Dr. Pin-Yu Chen
- **National Tsing Hua University** Mar. 2022  
*Master of Science in Computer Science*  
Hsinchu, Taiwan
  - Advisor: Prof. Tsung-Yi Ho
- **National Tsing Hua University** June 2020  
*Bachelor of Science, Computer Science, Mathematics*  
Hsinchu, Taiwan

## RESEARCH EXPERIENCE

- **Trusted AI Group, IBM Thomas J. Watson Research Center** Oct. 2022 - Dec. 2022  
*Research Intern*  
Yorktown Heights, NY  
Mentors: Dr. Pin-Yu Chen and Dr. Nandhini Chandramoorthy  
*Project 1: NeuralFuse On-chip Energy-efficient Inference*
  - Proposed a protection module (NeuralFuse) for on-chip AI accelerators, enabling them to withstand bit errors caused by low voltage while maintaining stable performance. [CP.6]*Project 2: Neural Network Calibration and Visualization*
  - Developed a neural network calibration package to help ensure consistency between the confidence of model prediction and the actual correctness likelihood. [CP.3]

## PUBLICATIONS & PREPRINTS

\* EQUAL CONTRIBUTION; S=IN SUBMISSION, CP=CONFERENCE PROCEEDINGS

- [S.1] **Lei Hsiung**, Tianyu Pang, Yung-Chen Tang, Linyue Song, Tsung-Yi Ho, Pin-Yu Chen, and Yaoqing Yang. **Your Task May Vary: A Systematic Understanding of Alignment and Safety Degradation when Fine-tuning LLMs**. 2025.  
💡 LLM Alignment 💡 AI Safety
- [CP.7] Hsi-Ai Tsao, **Lei Hsiung**, Pin-Yu Chen, and Tsung-Yi Ho. **When Does Visual Prompting Outperform Linear Probing for Vision-Language Models? A Likelihood Perspective**. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2025.  
💡 Vision-Language Model 💡 Visual Prompts 💡 Parameter-Efficient Fine-Tuning
- [CP.6] Hao-Lun Sun, **Lei Hsiung**, Nandhini Chandramoorthy, Pin-Yu Chen, and Tsung-Yi Ho. **NeuralFuse: Learning to Recover the Accuracy of Access-Limited Neural Network Inference in Low-Voltage Regimes**. *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.  
💡 Bit-Error Robustness 💡 Energy-Efficient Inference
- [CP.5] **Lei Hsiung**\*, Hsi-Ai Tsao\*, Pin-Yu Chen, Sijia Liu, and Tsung-Yi Ho. **AutoVP: An Automated Visual Prompting Framework and Benchmark**. In *Proceedings of the Twelfth International Conference on Learning Representations (ICLR)*, 2024.  
💡 AutoML 💡 Model Reprogramming 💡 Parameter-Efficient Fine-Tuning
- [CP.4] **Lei Hsiung**, Yun-Yun Tsai, Pin-Yu Chen, and Tsung-Yi Ho. **Towards Compositional Adversarial Robustness: Generalizing Adversarial Training to Composite Semantic Perturbations**. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.  
💡 Adversarial Robustness 💡 Combinatorial Optimization
- [CP.3] **Lei Hsiung**, Yung-Chen Tang, Pin-Yu Chen, and Tsung-Yi Ho. **NCTV: Neural Clamping Toolkit and Visualization for Neural Network Calibration**. In *Proceedings of the 37th AAAI Conference on Artificial Intelligence (AAAI)*, 2023.  
💡 Neural Network Calibration 💡 Model Reprogramming
- [CP.2] **Lei Hsiung**, Yun-Yun Tsai, Pin-Yu Chen, and Tsung-Yi Ho. **CARBEN: Composite Adversarial Robustness Benchmark**. In *Proceedings of the 31st International Joint Conferences on Artificial Intelligence (IJCAI)*, 2022.  
💡 Adversarial Robustness Benchmark
- [CP.1] **Lei Hsiung**, Yung-Ju Chang, Wei-Ko Li, Tsung-Yi Ho, and Shan-Hung Wu. **A Lab-Based Investigation of Reaction Time and Reading Performance using Different In-Vehicle Reading Interfaces during Self-Driving**. In *Proceedings of the 14th Int'l ACM Conf. on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI)*, 2022.  
💡 Human-Computer Interaction 💡 Eye-Tracking Analysis 💡 User Study

## HONORS, AWARDS, AND GRANTS

---

- **Guarini Travel Grant (2024)**. Guarini School of Graduate and Advanced Studies, Dartmouth College
- **Scholar Award (2024)**. The Twelfth International Conference on Learning Representations, Vienna, Austria
- **Volunteer Award (2023)**. The 37th Annual Conference on Neural Information Processing Systems, New Orleans, LA
- **Dartmouth Fellowship (2023)**. Guarini School of Graduate and Advanced Studies, Dartmouth College
- **Honorary Member (2022)**. The Phi Tau Phi Scholastic Honor Society of R.O.C. (top 3% graduands)
- **Mayor's Award (2015)**. Kaohsiung City, Taiwan (1st place in high-school graduation)

## TALKS

---

- **Building Trustworthy Systems: Compositional Adversarial Robustness and Low-Voltage Inference**. *TrustML Workshop at The University of British Columbia*, June 23, 2023 (Vancouver, BC, Canada)
- **Bit Errors of SRAM-Based Weight Storage: Trade-offs Between Energy and Accuracy**. *AI Accelerators Short Course, Taiwan AI Academy*, Mar. 29, 2022 (invited by Prof. [H. T. Kung](#)) (Hsinchu, Taiwan)

## TEACHING EXPERIENCES

---

Deep Learning Generalization and Robustness, Head TA (2024, 2025); Integrated Circuit Design, Head TA (2021); Calculus I, Undergraduate Teaching Assistant (2019); Calculus II, Undergraduate Teaching Assistant (2019/2020)

## PROFESSIONAL SERVICES

---

- **Reviewers:** (Conf.) ICLR 2024-2025, NeurIPS 2022-2024, ICML 2022-2024, AAAI 2025, CVPR 2025; (Journal) TMLR
- **PC Members:** AdvML Frontiers Workshop (ICML' 22/23), AdvMLDM Workshop (KDD' 22)
- **Volunteers:** NeurIPS 2023

## SKILLS

---

- **Programming Languages:** C, C++, Python, JavaScript, SQL, Verilog
- **Machine Learning:** PyTorch, DeepSpeed, Slurm, Scikit-learn, Pandas, NumPy, SciPy, Matplotlib, Seaborn
- **Languages:** English (fluent), Chinese (native)
- **Tools I love:** Vim, Git, tmux