

Hung Yun Tseng

htseng23@wisc.edu | linkedin.com/in/hung-yun-tseng | github.com/AndreaTseng

EDUCATION

University of Wisconsin - Madison	Sep 2025 - present
Electrical and Computer Engineering, Ph.D.	
University of Wisconsin - Madison	Sep 2021 - May 2025
Bachelor of Science in Computer Science - 3.55/4.0 GPA	

EXPERIENCE

Undergraduate Research Assistant , Prof. Grigoris Chrysos	Feb 2024 – Present
<ul style="list-style-type: none">Design LJ-Bench, a legally-grounded benchmark for evaluating LLM safety across 76 crime types that identified new vulnerabilities in major language models.Develop ontology framework and design analysis evaluating 16 models across 16 models, demonstrating higher attack success rates than existing benchmarks.Propose privacy algorithms that address "test-time privacy" vulnerabilities where adversaries exploit confident model predictions on sensitive data, creating Pareto-optimal finetuning methods that achieve $>3\times$ improved uncertainty on protected instances while maintaining $<0.2\%$ accuracy loss on standard benchmarks.	
Full Stack Mobile Application Developer IPM, UW-Madison	Apr 2023 – Sept 2023
<ul style="list-style-type: none">Developed PVY Predictor using WeatherBit API and Maps SDK to help farmers forecast the risk of potatoes virus, resulting in a 20% cost reduction in paraffinic oil usage for Wisconsin farmers.Launched updated apps on Google Play, achieving a 15% increase in downloads by transitioning from SQLite to Room database and implementing an alert system for high PVY risk.Improved old apps UI performance by utilizing lint to spot inefficient layout hierarchy and adopting lazy view inflation, ensuring smoother animation during user scroll.	

PROJECTS

Geometry of Truth (Pytorch)	Dec 2024
<ul style="list-style-type: none">Applied SVM classifiers and PCA throughout 10 layers in TinyLlama-1.1B and LLaMA 3-70B, demonstrating that larger models exhibit stronger linear separation of truth-related representations.	
OrwellGPT (Pytorch)	Sep 2024
<ul style="list-style-type: none">Used reinforcement learning on a tiny GPT-2 model to mimic George Orwell's writing style, achieving a perplexity of 18.1.	
Image Classification (Numpy)	May 2024
<ul style="list-style-type: none">Built a CNN model from scratch with NumPy, implementing custom convolutional layers, batch normalization, and backpropagation, achieving 90% accuracy on an ImageNet subset.	
BaderRoo APP (Kotlin)	Oct 2023
<ul style="list-style-type: none">Engineered an Android app to help UW-Madison students find compatible roommates, collaborating in a 4-member team.Implemented a scalable backend using Firebase for real-time data synchronization and storage. Integrated UI/UX features including swipe-based matching and real-time messaging.	

PUBLICATIONS

<i>LJ-Bench: Ontology-based Benchmark for Crime, COLM 2025 Under Review</i>	Mar 2024
<i>Inducing Uncertainty for Test-Time Privacy</i>	Oct 2024

RELATED WORK

Reviewer for NeurIPS 2024 Workshop for Women in Machine Learning	Oct 2024
Reviewer for AAAI 2025 Workshop on Connecting Low-Rank Representations in AI	Nov 2024
Reviewer for ICLR 2025 Workshop on Uncertainty and Hallucination in Foundation Models	Feb 2025
Paper Presentation	Feb 2024 - present
<ul style="list-style-type: none">Present in-depth review of foundation papers such as <i>An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale</i>, <i>Fast Inference from Transformers via Speculative Decoding</i>, and <i>Denoising Diffusion Probabilistic Models</i>.	
Wisconsin AI Safety Initiative, Technical Team	Sep 2023 - May 2024
<ul style="list-style-type: none">Facilitate weekly AI safety meetings, guiding discussions on topics like scalable oversight and mechanistic interpretability.	
Peer Mentor - CS 540, CS 300	Jan 2025 - May 2025
<ul style="list-style-type: none">Serve as peer mentor for CS540: Intro to Artificial Intelligence and CS300: Java Programming, clarifying complex algorithms and machine learning concepts while providing hands-on guidance for programming assignments.	

HONORS

Awards: Dean's List (GPA > 3.85): Spring 2022, Fall 2023, Fall 2024