

Hrithik Ravi

hrithikravi007@gmail.com | <https://hrithikravi.github.io> | Work Eligibility: US Citizen

EDUCATION

University of Michigan

M.S. in Elec. and Comp. Engineering (focus in Machine Learning, GPA: 4.0/4.0)

Ann Arbor, MI

B.S.E. in Computer Science and Engineering (Minor in Mathematics)

(Aug. 2023 – Dec. 2024)

(Aug. 2019 – April 2023)

PUBLICATIONS

The Implicit Bias of Gradient Descent on Separable Multiclass Data

Hrithik Ravi, Clayton Scott, Daniel Soudry, Yutong Wang

NeurIPS 2024

RESEARCH EXPERIENCE

Michigan AI Laboratory – Research Assistant

Ann Arbor, MI

Analysis of Chain-of-Thought Reasoning in LLMs – Advised by Prof. Ambuj Tewari

(June 2025 – present)

- ❖ Designed setup to probe reasoning limitations of (non-thinking) frontier LLMs on synthetic, verifiable NLP reasoning tasks.
- ❖ Conducted experiments demonstrating poor OOD generalization abilities of base models.
- ❖ Devised “in-context distillation” approach to distill reasoning skills from thinking models into non-thinking models in-context.
- ❖ Analyzing surprising brittle generalization of SFT-based post-training compared to our in-context strategy.
- ❖ Preparing paper for ICML 2026.

Google DeepMind – Open-Source Developer

Remote

Benchmark for Multimodal LLMs – Advised by Paige Bailey

(May 2025)

- ❖ Aimed to design a novel multimodal benchmark for Gemini and other (M)LLMs.
- ❖ Reviewed state-of-the-art literature about needle-in-a-haystack (NIAH) evals. Identified key limitation of NIAH that needles were highly distinct from rest of the context.
- ❖ Wrote design proposal and implementation plan for a new synthetic multimodal eval to fix this limitation, essentially generalizing Multi-Round Coreference Resolution (MRCR) beyond textual modalities.
- ❖ Proposal accepted to Google DeepMind (Google Summer of Code, 5% acceptance rate).
- ❖ Withdrawn from program to focus on research more closely relevant to my interests.

Michigan AI Laboratory – Research Assistant

Ann Arbor, MI

Mechanistic Interpretability to Debias VLMs – Advised by Prof. Rada Mihalcea

(Dec. 2024 – May 2025)

- ❖ Researched cost-effective and interpretable inference-time steering approaches to debias OpenAI CLIP.
- ❖ Implemented linear probing and Contrastive Activation Addition, in both text as well as image embedding space.
- ❖ Observed negative results. Investigated and identified that observed bias was primarily an artifact of strict scoring metrics.

University of Michigan EECS Dept. – Student Researcher

Ann Arbor, MI

Mechanistic Interpretability to Elicit Chain-of-Thought from LLMs – Advised by Prof. Wei Hu

(Feb. 2024 – May 2024)

- ❖ Surveyed mechanistic interpretability literature to understand representations that emerge from unsupervised pretraining.
- ❖ Observed that linearity seemed to be restricted to high-level behavioral changes (such as stylistic changes, truthfulness, toxicity, etc.).
- ❖ Aimed to investigate whether precise, procedural reasoning skills also had linear representations. Planned to test Inference-Time Intervention and Contrastive Activation Addition on synthetic syllogism dataset (PrOntoQA).
- ❖ Was unable to continue working on the project due to health issues.

University of Michigan EECS Dept. – Student Researcher

Ann Arbor, MI

Generalization Theory for Deep Learning – Advised by Prof. Clayton Scott

(Aug. 2023 – May 2024)

- ❖ Surveyed theoretical literature on the “implicit bias” hypothesis, which attributes generalization of overparameterized neural nets to (S)GD’s implicit bias towards simple solutions.
- ❖ Identified a gap in literature – multiclass results were sparse and limited to cross-entropy loss.
- ❖ Generalized a seminal binary result to the multiclass setting using a unified theoretical framework.
- ❖ Empirically validated our theorems, thus demonstrating first ever multiclass implicit bias effect for not just cross-entropy.
- ❖ Published findings at NeurIPS 2024.

UofM Transportation Research Institute – Temp. ML Research Engineer

Ann Arbor, MI

3D Object Detection for Self-Driving Cars – Advised by Dr. Arpan Kusari

(May 2023, June – Aug 2023: part-time)

- ❖ Implemented a computer vision 3D IOU algorithm in OpenCV to evaluate performance of Detectron2 object detection model.
- ❖ Designed the algorithm to penalize false negatives harshly resulting in a model optimized for maximal passenger safety.

- ❖ Implemented a sensor fusion algorithm in MATLAB to provide a 360° view and enable safe path-planning in self-driving cars.
- ❖ Used Hungarian algorithm to fuse LIDAR and camera tracks, programmatically adding new tracks and discarding stale ones.

WORK EXPERIENCE

University of Michigan – Computer Science Graduate Student Instructor (TA) <i>EECS 281: Data Structures and Algorithms</i>	<i>Ann Arbor, MI</i> (Aug. 2023 – Dec. 2024)
❖ Received full tuition waiver to teach undergraduates foundations of data structures and algorithms.	
❖ Held lectures, office hours to walk through example problems and to help students implement and debug C++ project solutions.	
Amazon Web Services – Software Engineering Intern <i>Security and Networking; Time-series Analysis for Predictive Managed Scaling</i>	<i>Herndon, VA</i> (June 2023 – Sept. 2023)
❖ Built an iptables firewall with a token-bucket algorithm to fortify DNS servers against request floods, ensuring QoS for 3MN clients per region.	
❖ Leveraged AWS Python API to implement an ML time-series pipeline in AWS Forecast, S3 to predict city-wide electricity consumption.	
Amazon Web Services – Software Engineering Intern <i>Full-Stack Mobile and Web Development for Supply Chain Analytics</i>	<i>Austin, TX</i> (May 2022 – Aug. 2022)
❖ Mobile: Built notification UI in React Native, Typescript; backend in Expo, Firebase, Apple Notification Service, and NodeJS.	
❖ Web: Developed interface to configure analytics, implementing frontend in React, backend in Kotlin and AWS DynamoDB.	
Icuro Inc. – Machine Learning Intern <i>Natural Language Processing (NLP) – Market Sentiment Analysis</i>	<i>Santa Clara, CA</i> (July 2021 – Sept. 2021)
❖ Fine-tuned GPT-3 to analyze sentiment towards Amazon consumer product reviews in key aspects (“comfort”, “quality”, etc.).	

AI, ML, AND CS PROJECTS

❖ LLMs – Built LLM pipeline (data, large GPT model, self-supervised pretraining, LoRA finetuning) in PyTorch.	<i>March 2025</i>
❖ AI – Enhanced Sakana AI Scientist with RAG-based experiment reproduction from user-uploaded research papers.	<i>Feb. 2025</i>
❖ ML – Trained CNN in PyTorch to extract goals from soccer footage, creating engaging highlight reels for viewers.	<i>Aug. 2024</i>
❖ Embedded Systems – Built wireless gesture-controlled Glove, demoed “Hot or Cold” hidden object detection game.	<i>Aug. 2022</i>
❖ OS – Implemented full-fledged OS kernel in C++.	<i>Jan. – May 2022</i>

INVITED TALKS

“A Theoretical Overview of Test-time Scaling”, STATS 700 Guest Lecture (invited by Prof. Ambuj Tewari). *Sept. 2025*

“Is AI Changing Who We Are?”, Michigan Friday Night AI, Invited Panelist (along with Prof. Rada Mihalcea and Prof. Jonathan R. Brennan). *Sept. 2025*

“Chain-of-Thought Reasoning: Trends and Directions”, UMich Language and Information Technologies Lab. *May 2024*

“Multiclass Implicit Regularization”, Michigan Student Symposium for Interdisciplinary Statistical Sciences (MSSISS). 3rd place. *March 2024*