

Ethan Harianto

eharianto@stanford.edu | ethanharianto.github.io | (347) 475-7671

Education

Stanford University

MS in Computer Science (AI)

Expected June 2026

BS in Computer Science (Systems)

Expected June 2026

Relevant Coursework: Deep Learning, Spoken Language Processing, Artificial Intelligence Principles, Design & Analysis of Algorithms, Operating Systems Principles

Experience

Junior Software Developer, Pantheon Lab

June 2025 – Aug 2025

- Architected a voice-activated Retrieval-Augmented Generation (RAG) pipeline using LlamaIndex, integrating open-source LLMs into a real-time WebRTC framework.
- Engineered a distributed inference system in Go, optimizing high-concurrency request handling and quota management for GPU clusters on Vast.ai.
- Reduced input-to-response latency for conversational agents by processing audio streams in-memory, optimizing the interface between the ASR/TTS models and the LLM.

Research Assistant, Stanford PinCS Lab

Mar 2025 – June 2025

- Engineered a real-time memory-aid application leveraging AssemblyAI for speech-to-text, designing accessible UI patterns for patients with cognitive impairments.
- Implemented rapid-prototyping cycles to integrate user feedback from clinical pilot studies, refining the interaction model for high-stakes healthcare environments.

Founding Engineer, Slide Social

Dec 2022 – Sep 2023

- Designed a custom image compression algorithm to optimize mobile-to-cloud data transmission, reducing backend storage costs by 90% while maintaining perceptual quality.
- Architected a scalable Firebase backend to handle concurrent user interactions, optimizing NoSQL database queries for sub-100ms latency.

Projects

LyricNet: Multimodal Music Emotion Recognition | Python, PyTorch, Transformers

Oct 2025 - Dec 2025

- Developed a multimodal deep learning framework fusing DistilBERT lyric encodings with normalized audio features to predict emotional states.
- Achieved 95.58% validation accuracy, outperforming audio-only baselines (70.83%) and lyric-only models (90.58%) by stabilizing training through feature fusion.
- Mitigated overfitting on a long-tail distribution of 500,000+ tracks by implementing a class-balanced weighted loss function.

Operating Systems Kernel (Pintos) | C, x86 Assembly

Apr 2025 - June 2025

- Engineered a multi-threaded OS kernel in C, implementing priority scheduling, virtual memory paging, and a hierarchical file system on x86 architecture

Reinforcement Learning Agent (Coup) | Python, NumPy

June 2024

- Developed an autonomous game agent implementing a Q-learning algorithm, achieving a >90% win rate against random-action opponents through iterative self-play training.

Rock Climbing ML Grading System | Python, PyTorch, YOLOv5

May 2024 - June 2024

- Engineered a self-supervised learning pipeline using PyTorch and YOLOv5 to automate route grading, achieving 80% accuracy against expert evaluations.

Skills

Languages: Python, Go, C++, C, SQL, Swift, JavaScript

Machine Learning: PyTorch, Transformers (Hugging Face), LlamaIndex, Scikit-learn, Pandas, NumPy, YOLO

Infrastructure & Tools: Docker, Git, AWS, Azure, Vast.ai, WebRTC, Linux/Unix