

Harutoshi Okumura

(404) 316-9572 | HOkumura032@gmail.com | linkedin.com/in/harutoshi-okumura | github.com/HarutoshiOkumura

EDUCATION

Emory University, Laney Graduate School

Master of Science, Double Major: Computer Science

Atlanta, GA

Expected May. 2027

Relevant Courses: Linear Optimization, Machine Learning, AI Research Practicum, Systems Programming

EXPERIENCE

Robust.io (Z-Fellows — \$2.5M Pre-seed)

Founding Software Engineer

NYC, NY

Jun. 2025 – Present

- Designed an **end-to-end**, Agent Controller (**TypeScript + Fastify**) that orchestrates creation, tracking, and termination of cloud **Letta** agents and local sandbox workers, establishing a unified, reusable lifecycle API
- Architected a **multi-agent orchestration** platform with **Redis** caching for completed workstreams and **SQL** snapshotting for execution-state tracking, enabling **parallel agent workflows** with robust global-undo support
- Engineered a **Fastify** backend ingesting, indexing, and searching files, enabling sub-100 ms AI retrieval

Scale AI

Remote

Technical Advisor Intern

May 2025 – Aug. 2025

- Researched generative model reasoning in **C++/Python**, boosting competitive programming benchmarks by 20%
- Enhanced LLM SWE agent's ability to navigate large codebases leveraging **Git**, **Docker**, and **Python**
- Reduced model reasoning error rate by 25% through cross-functional error analysis with AI research teams and targeted fine-tuning of LLM architectures

Pagent TV (\$10M Val. Mobile App)

Bellevue, WA

Founding Machine Learning Engineer

Mar. 2025 – Present

- Built a recommendation system using **TensorFlow Serving**, hosted **DeepFM** model with **Apache Flink Kafka joins** and GPU-accelerated inference, cutting p99 latency by 90%
- Engineered a two-tiered storage solution for unused Kafka streams, leveraging **AWS Redis** for immediate caching & **S3 Glacier Deep Archive** for long-term storage, optimizing cost-performance by **95.7%**
- Exposed a **Socket.io API** for live match updates and real-time **WebSocket communications**, supporting over **1000 concurrent players** with **97.8% uptime**
- Engineered a **MongoDB**-backed scheduling engine that executes recurring workflows with **Node.js**
- Built & published video ranker for a TikTok-style video competition mobile app using **React-Native & Swift**
- Containerized** the application stack with **Docker** to provision reproducible multi-container environments, cutting CI/CD deployment failures by 40%

CuraStone Corp., KnoWhiz

Bellevue, WA

Full-Stack Software Engineer Intern

Jun. 2024 – Dec. 2024

- Researched MCQ generation using LLMs with **Q&A** fine-tuning, driving accuracy from **76%** to **92%**
- Engineered an **asynchronous math pipeline** using **Quart** & integrating **Azure Blob Storage**, **MongoDB**, & **OpenAI** to automate generation & storage of math learning material, increasing STEM engagement by **20%**
- Diagnosed LLM content generation latency issues & optimized delivery by dynamically rendering courses, improving render time by **30%** & active customer base to **300**
- Developed an optimized study feature using a **job scheduling algorithm (MFQS)** to prioritize new flashcards & revisit earlier topics systematically

PROJECTS [\(MORE PROJECTS HERE\)](#)

Research Paper Author: Automatic Contradiction Detection in Logical Texts with LLMs (1st Author)

- Built an NLP pipeline leveraging **BERTopic** to vectorize & cluster debate speeches in **QDrant**, reduce dimensions via **UMAP**, & assign labels using **c-TFIDF**; validated with a fine-tuned **NLI model** from HuggingFace, improving detection rate by **32%** and reducing false positives by **21%**
- Designed a novel framework integrating Symbolic AI (formal logic) & LLMs to classify debate statements & identify logical inconsistencies using a proprietary taxonomy based on Aristotle's law

TECHNICAL SKILLS

Languages + OS: Java, Python, C++, TypeScript, JavaScript, HTML/CSS, Go, Kotlin, Linux (Asahi Fedora Remix)

Database & Services: MySQL, PostgreSQL, Google Cloud, GraphQL, MongoDB, AWS, Roboflow, CUDA

Libraries & Frameworks : Git, React, OpenAI, Next.js, Node.js, PyTorch, TensorFlow, Docker, YOLO,

Leadership Activities: Alpha Kappa Psi (VP Technology), Emory Symphony Orchestra (1st Violin), Asian Student Org (VP Finance), LEAP, Atlas Consulting Group (Director Finance & Admin), Computer Change++ (Founder)