Michael Huang

Oakland, CA | (415) 505-8933 | michael.micah.huang@gmail.com | www.linkedin.com/in/michael-micah-huang

Summary

CS engineer focused on applied AI/ML — from prompt optimization and multi-agent orchestration to containerized ML deployments. Built production-grade AI systems as a contract engineer (MCP server, Gemini integration, 157-test codebase), deployed multi-agent pipelines to k3s, and won 2nd place at two hackathons building real-time LLM-powered applications.

Education

California Polytechnic State University - San Luis Obispo San Luis Obispo, CA
Bachelor of Science in Computer Science, Minor in Philosophy; Graduated: August 2025

- **Relevant Coursework:** Data Structures & Algorithms, Computer Architecture, Operating Systems, Systems Programming, Database Systems, Deep Learning

Professional Experience

Dynamic Experts December 2025 – January 2026
Contract AI Engineer *Remote*

- Architected a production-grade MCP (Model Context Protocol) server to expose deep-research capabilities, integrating Gemini 2.5 Pro with markdown sanitization and automated citation handling
- Delivered a fully documented, production-ready codebase with **157 unit tests** (Vitest), ensuring seamless handoff to the client engineering team
- Optimized async polling architecture with **90-minute timeout handling** to support long-running inference tasks (**60+ minutes**), solving critical timeout issues in standard HTTP request flows

Solidigm (Formerly Intel NAND) June 2022 – August 2023
Automation and Product Engineering Intern *Folsom, CA*

- Migrated legacy Python 2.7 automation frameworks to Python 3.x, modernizing the codebase for **600+ production-line tests** and significantly reducing false-positive failures
- Engineered end-to-end automation scripts that increased test coverage, saving the engineering team approximately **8 hours of manual verification per SKU** during testing cycles
- Performed root cause analysis on non-product failures, implementing patch fixes that improved overall testing stability and reduced infrastructure downtime

Hackathon

Hackathon Wins | *Gemini, Azure Speech, ElevenLabs, TypeScript* Jan 2026

- Won **2nd place** at **two** hackathons, building a real-time election misinformation detector (Before the Ballot) and a live speech fact-checker (Claude Code Hack Day @ AWS, out of **25 teams**)
- Upgraded static LLM validation to **live search grounding** with Gemini, enabling real-time source retrieval and contradiction flagging
- Engineered a JSON Schema validation layer using Gemini to ensure deterministic, error-free outputs from non-deterministic LLMs
- Led orchestration of **four microservices** and shipped fully integrated, crash-resistant MVPs under tight sprint deadlines

Projects

Axion Prompt Learning Module | *Python, Pydantic, AsyncIO* January 2026 – Present

- Built a prompt optimization module that iteratively improves system prompts via an evaluate-analyze-optimize loop using structured English-language feedback
- Implemented beam search with candidate generation, auto-revert on regressions, and hard negative sampling for failure analysis

SAR Multi-Agent Deployment Pipeline | *Docker, k3s, Redis Streams, Python* January 2026 – Present

- Containerized a 7-agent AI search-and-rescue system, building optimized Docker images and deploying to a k3s cluster
- Configured in-cluster Redis and networking for local orchestration agents, preserving the existing Redis Streams protocol

Technical Skills

Languages: Python, C/C++, Java, JavaScript (ES6+), TypeScript, SQL
AI/ML: PyTorch, Gemini API, Azure Speech SDK, Sklearn, Jupyter Notebook, LLMs, Generative AI
Web Development: Node.js, Express, React.js, MongoDB
Tools & Infrastructure: Git, Linux, POSIX, Docker, k3s/Kubernetes, Redis, MCP, Firecrawl API, Vitest, Confluence, Jira