Mingma Lama

Software Engineer · Full-Stack & AI Systems

Baltimore, MD · 814-500-3723 · lamamingma9811@gmail.com · linkedin.com/in/mingma23 · github.com/lama9811 · mingmalama.xyz

EDUCATION

Morgan State University — B.S. Computer Science — Expected May 2027

- GPA 3.74/4.00 · Tau Sigma National Honor Society · Dean's List (Community College of Baltimore County)
- Coursework: Data Structures & Algorithms, Object-Oriented Programming, Discrete Structures, Software Engineering

EXPERIENCE

Applied AI Lab, Morgan State University — Baltimore, MD · Jan 2026 – Present

Lead Engineer, AI Engineering Intern *— ORA Navigator, Office of Research Administration* — *May 2026 – Present*

- Shipped an AI assistant (ora.inavigator.ai) to **ORA staff, PIs and researchers**, grounded in 751 documents
- Built a 3-layer grounding pipeline that keeps answers to the knowledge base, adds missing citations and flags weak ones
- Designed an AI grant reviewer that checks proposals against NSF rules, with code verifying every quoted finding
- Eliminated AI score drift across uploads (**10/10 identical runs**) using code-based rule checks and 3-pass majority voting
- Built an NSF grant budget builder that calculates costs, flags **23 funding-policy issues** and exports to Excel and PDF
- Deployed 3 services (React PWA, FastAPI, ADK agent) on Cloud Run with Cloud SQL, Redis and CI/CD

Lead Engineer, AI Engineering Intern *— CS Navigator, Department of Computer Science* — *Jan 2026 – May 2026*

- Shipped an AI advisor (cs.inavigator.ai) to **CS students** that answers course and degree questions from their degree audit
- Made responses **3x faster** with a 3-layer cache (in-memory, Redis, semantic)
- Built a semester planner that helps students pick classes using **live seat data from Banner**, refreshed every 5 minutes
- Hardened a sandboxed Python/JavaScript runner for an AI coding tutor (AST blocklists, CPU/memory limits)
- Added five-layer conversation memory with **FERPA** privacy controls, voice input and Google Search mode

Research Assistant *— NAO CBT Companion Robot* — *Aug 2026 – Present*

- Built a real-time voice pipeline (WebSockets, Silero VAD, Deepgram, ElevenLabs) and a multi-agent Claude backend
- Routed each turn to the right Claude model: Haiku for fast replies, Sonnet for CBT depth, Opus for crisis triage
- Added a self-harm safety layer that runs before any agent replies; fixed a cross-user history leak with session scoping
- Built speaker-aware turn-taking with sound localization and face tracking, keeping a separate history per person
- Connected the robot to the CS Navigator API for campus questions and automated Raspberry Pi deploys

Department of Computer Science, Morgan State University — Baltimore, MD · Sep 2025 – Apr 2026

Undergraduate Researcher *— Mental Distress Prediction · Mentor: Prof. Radhouane Chouchane*

Python, Pandas, scikit-learn · github.com/lama9811/Mental-Distress-Prediction

- Modeled how depression and financial stress predict mental distress (CDC PLACES 2025, n=51), linear vs. random forest
- Engineered a composite **financial-threat** feature from four CDC hardship indicators (housing, utilities, food)
- Presented a poster, selected by abstract review, at the 31st Annual Research Symposium (Apr 2026)

PROJECTS

Kairos — Knowledge OS · Runner-Up ($3,000) + MLH Best Use of Gemini API — MorganHacks 2026

Next.js, FastAPI, PostgreSQL/pgvector, Three.js · team of 4 · github.com/lama9811/kairos-ai

- Built a knowledge graph that turns local code and notes into a 3D map of 1,000+ concepts via AST parsing and Gemini
- Implemented hybrid semantic + keyword search (pgvector, Postgres full-text, rank fusion) with nightly self-improvement
- Deployed **20 AI agents** with distinct reasoning styles via Gemini function calling, grounded in the knowledge graph

EmerFlow — Hospital Surge Coordinator — HopHacks 2026, Johns Hopkins

FastAPI, Next.js, Gemini on Vertex AI, Server-Sent Events · team of 4 · github.com/lama9811/Emerflow

- Built a real-time command board where 10 AI agents negotiate patient placement with all bed math enforced in code
- Added a 5-level escalation ladder, **human approval** for big actions, and a circuit breaker that runs without AI
- Gave every agent its own **5-hour memory**, with each patient memory stored as a node in a 3D Three.js graph

SKILLS

Languages: Python, TypeScript, JavaScript, SQL, Java
Backend & Frontend: FastAPI, REST APIs, WebSockets, SSE, React, Next.js, Tailwind CSS, Three.js
Cloud & DevOps: Google Cloud (Cloud Run, Cloud Build, Cloud SQL, Vertex AI), GitHub Actions, CI/CD, Linux/systemd, Git
Databases & Tools: PostgreSQL, MySQL, Redis, pgvector, pytest
AI Systems: RAG, multi-agent systems, embeddings, Google ADK, Gemini API, Claude API, scikit-learn, Pandas