

O. FELIX AMORUWA

famoruwa@berkeley.edu | 909-731-9011 | [felimamoruwa.info](mailto:famoruwa@berkeley.edu)

AI product leader with 12+ years driving model strategy, evaluation frameworks, and developer ecosystems at scale — from hand-coding BPTT in C++ (2004) to building RL post-training workbenches that benchmark GRPO/DPO across TRL, VeRL, OpenRLHF, and NeMo RL today. Built production model evaluation platforms (aeval) with systematic quality frameworks across factuality, reasoning, safety, and code generation. Deep expertise in agentic flows, multi-model LLM orchestration, and inference optimization; scaled platform infrastructure to 675M+ engagements at Intuit. NeurIPS published researcher; UC Berkeley Engineering + CMU MBA.

AI/ML RESEARCH & PROJECTS

aeval — AI Model Evaluation Platform (2025–2026)

- Built local-first model evaluation platform with 5 core eval types (factuality, reasoning, instruction-following, safety, code generation), adversarial safety testing with refusal detection, and data contamination detection — directly implementing the systematic evaluation frameworks and quality standards required for production model catalogs.
- Statistical rigor: bootstrap confidence intervals, Welch's t-test, Cohen's d effect size, saturation detection; CI/CD integration with regression detection and automated safety gates. Stack: FastAPI orchestrator, TimescaleDB, Redis job queue, Next.js dashboard, Ollama.

RL Workbench — Post-Training RL Platform (2026)

- Built 3-phase RL post-training workbench covering the full RLHF/DPO pipeline: Reward Lab for designing and A/B testing reward functions across 4 datasets (GSM8K, MATH, HumanEval, UltraFeedback), Playground for real TRL-powered GRPO/DPO training with live SSE metric streaming on Apple Silicon (MPS) or CUDA, and Arena for head-to-head framework benchmarking (TRL, VeRL, OpenRLHF, NeMo RL) with GPU passthrough in Docker containers.
- Implemented 12 RL algorithms (PPO, GRPO, DAPO, REINFORCE++, RLOO, DPO, SimPO, IPO, KTO, ORPO, SPPO) with standardized throughput/memory/convergence benchmarking across frameworks — the same benchmark-and-evaluate discipline required to prove production-grade inference performance on Cerebras hardware.

BRAIN — Protein Structure Prediction ML Platform (UC Berkeley 2004; rewritten 2026)

- Built production ML platform in PyTorch with 5 neural architectures (feedforward, GRU, Transformer, ESM-2, multi-task), MLflow experiment tracking, Optuna HPO, and FastAPI serving — spanning 413 to 8B parameters (19M-fold scale increase); original 2004 system hand-coded in C++ with custom BPTT.
- NeurIPS 2014 accepted paper on artificial neural networks for protein secondary structure prediction; computational analysis at Lawrence Berkeley National Laboratory under Dr. Steven Holbrook.

AutoEval — Automated Visual Evaluation for Robot Model Training (2025)

- Built automated visual evaluation system using multimodal AI (Claude/GPT-4V) to score model outputs (grasp poses, segmentation maps, bounding boxes) against natural-language evaluation rubrics — reducing evaluation cycles from 72 hours to ~4 minutes; zero-integration architecture captures from any visualization tool via screen capture.

PROFESSIONAL EXPERIENCE

STREAMIO AI | Oakland, CA

Founder & CEO — AI Platform (macOS, Linux, iOS & Web) 09/24 – Present

- Implemented OpenClaw multi-agent orchestration framework with gateway protocol, subagent delegation, profile management, and session switching — delivering production agentic flows across real estate, insurance, health/dental, and financial markets use cases; directly applicable to Cerebras' agentic computation inference workloads.
- Integrated Claude (MCP SDK) and multimodal AI for real-time screen analysis, contextual conversations, and automated report generation — hands-on experience with chat completions API, system prompt design, and LLM application code across code generation and deep research use cases.
- Shipped cross-platform desktop builds with native macOS ScreenCaptureKit integration, embedded terminal (20+ slash commands), and MCP server exposing screen capture tools to AI coding assistants; built end-to-end auth and payments pipeline (Kinde OAuth 2.0, Stripe) with ElevenLabs TTS/STT integration.
- Led 0-to-1 product strategy, AI development, and go-to-market execution including customer discovery interviews and iterative product refinement — entrepreneurial ownership across the full product lifecycle.

FINTELLECT AI | Oakland, CA

Founder & CEO — AI Financial Platform (iOS/Android/Web) 09/24 – Present

- Architected multi-provider LLM orchestration pipeline (Claude, GPT-4, Gemini) with fallback routing, structured output validation, and token budget optimization — hands-on model portfolio management across frontier and open-source models.
- Built domain-specific conversational AI agents with RAG retrieval (ChromaDB vector store) and AI-powered macroeconomic analysis tools; executed go-to-market through App Store launch with customer discovery and influencer partnerships.

INTUIT | Mountain View, CA

Staff Product Manager — Developer Frameworks & Platform Infrastructure 06/21 – 09/24

- Achieved 275% YoY growth in platform engagements, scaling to 675M+ in FY23 across QuickBooks, TurboTax, Mint, Mailchimp, and Credit Karma; scaled throughput from 6K to 50K TPS via rSocket migration supporting ~1.5M concurrent connections with sub-25ms TP99 — enterprise-scale inference infrastructure experience.
- Extended Java and Python SDK Starter Kits with scaffolding templates, build configurations, testing frameworks, and CI/CD integration — empowering developers to go from zero to production-ready microservice in minutes; delivered ICE Self-Service platform reducing developer onboarding from 2–3 weeks to <24 hours.
- Conducted enterprise-wide Service Language Assessment across 9 languages, analyzing usage data and developer feedback to inform strategic technical decisions presented to CTO; built Asterias, a declarative asset lifecycle management platform with GraphQL API.
- Implemented ICE Presence in async chat generating \$480K/month in additional invoicing; deployed Background-to-Foreground Messaging on iOS/Android with <100ms latency. DeveloperWeek 2022 speaker.

SPLUNK INC. | San Francisco, CA*Senior Product Manager — Search Orchestration* 01/19 – 01/21

- Led query performance optimization initiative achieving up to 10x performance improvements for beta customer — the same performance-vs-quality tradeoff discipline required to maximize Cerebras' inference speed advantage.
- Owned Search Service (Go microservices), Search Catalog (PostgreSQL), and Splunk Processing Language (SPL/SPL2) roadmaps; delivered Scheduler Service end-to-end in ~4 months, demoed at .conf19. Splunk .conf18 Speaker.

KAISER PERMANENTE | Pleasanton, CA*SOA Technical Product Manager* 08/12 – 12/18

- Led development and enterprise rollout of Splunk Logging-as-a-Service (1.7 TB daily volume, 200+ internal enterprise customers); built Redis caching capability across the enterprise for scalability, fault tolerance, and data redundancy.

IBM | Pittsburgh, PA / Foster City, CA / RTP, NC*Software Engineer — Business Intelligence Products* 07/05 – 03/10

- Resolved multiple \$1–5M high-severity customer escalations for enterprise BI software; led cross-functional root cause analysis and customer-facing resolution, improving case resolution time by 20% — 5+ years SWE foundation underpinning technical PM credibility.

BANK OF AMERICA MERRILL LYNCH | Charlotte, NC*Tech MBA Summer Associate — Global Corporate & Investment Banking T&O* 06/11 – 08/11

- Developed enhanced estimation model using Monte Carlo simulation to optimize DMAIC phase distribution across a \$494M portfolio in Global Technology & Operations.

EDUCATION

CARNEGIE MELLON UNIVERSITY, SILICON VALLEY

Master of Science — Software Management Mountain View, CA (08/16)

CARNEGIE MELLON, TEPPER SCHOOL OF BUSINESS

Master of Business Administration — Concentrations: Finance, Quantitative Analysis, Economics Pittsburgh, PA (03/12)

UC BERKELEY, COLLEGE OF ENGINEERING & HAAS SCHOOL OF BUSINESS

Bachelor of Science — Computational Engineering Science Berkeley, CA (12/05)

Certificate — Corporate Finance for Financial Engineering

TEACHING EXPERIENCE

DE ANZA COLLEGE | Cupertino, CA*Adjunct Faculty — Computer Information Systems Department* 04/18 – Present

- CIS 35A/B, 36A/B — Java Programming; CIS 64C/64G — PL/SQL; CIS 44F — Data Analytics; CIS 64E — Large Scale Cloud Computing (AWS, GCP); CIS 102 — Ethical Hacking; CIS 104 — Digital Forensics; CIS 95F — Managing Cloud Projects.

ADDITIONAL INFORMATION

- FAA Commercial Drone License; FAA Private Pilot License (intended CFI)
- O'Reilly & OER Publications: Data Analytics and Managing Cloud Projects (forthcoming Summer 2026)
- First Tee Assistant Golf Coach — San Francisco Chapter (2012–Present), Pittsburgh Chapter (2011–2012)
- Interests: Classical & Jazz Saxophone (Alto/Tenor/Baritone), Triathlon, Golf, Intramural Basketball & Soccer