

# O. FELIX AMORUWA

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

Technical PM at the research-product boundary — 12+ years building developer-facing platforms and AI agent frameworks, from shipping SDK tooling at 675M+ engagements (Intuit) to building multi-agent orchestration systems, RL post-training workbenches, and AI evaluation harnesses from scratch. Designed and implemented subagent delegation frameworks, MCP-integrated agent pipelines, and statistical evaluation systems measuring task completion, error recovery, and hallucination frequency. NeurIPS published ML researcher; hands-on practitioner across PPO, GRPO, DPO, and 9 additional RL algorithms.

## AI/ML RESEARCH & PROJECTS

---

### RL Workbench — Post-Training RL & Agent Evaluation Platform (2026)

- Built 3-phase RL post-training workbench: Reward Lab for designing and A/B testing reward functions (RLVR, learned, hybrid) across 4 datasets (GSM8K, MATH, HumanEval, UltraFeedback); Playground for real TRL-powered GRPO/DPO training with live SSE metric streaming on MPS/CUDA; 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, REINFORCE++, RLOO, DPO, SimPO, IPO, KTO, ORPO, SPPO) with algorithm-specific metric profiles, cross-tab workflow lineage tracking, and standardized throughput/memory/convergence benchmarking — the same empirical evaluation discipline required to measure agent quality at scale.

### aeval — AI Model Evaluation Platform (2025–2026)

- Built local-first agent evaluation harness with 5 core eval types (factuality, reasoning, instruction-following, safety, code generation), adversarial safety testing with refusal detection, and data contamination detection — directly measuring task completion rate, hallucination frequency, and error recovery across model versions.
- 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.

### AutoEval — Automated Visual Evaluation for Agent Model Training (2025)

- Built automated agent evaluation harness that scores model outputs against natural-language rubrics — reducing evaluation cycles from 72 hours to ~4 minutes; multimodal AI (Claude/GPT-4V) performs spatial reasoning on prediction frames, generating structured PASS/FAIL reports with confidence scores and edge-case flagging.
- Zero-integration architecture captures agent outputs from any visualization tool via screen capture; directly applicable to analyzing agent traces at scale and identifying failure modes, loops, and unproductive execution paths.

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

- NeurIPS 2014 published. Production ML platform in PyTorch with 5 neural architectures, MLflow experiment tracking, Optuna HPO, and FastAPI serving — 823 automated tests, Docker orchestration (6 containers); spans 413 to 8B parameters. Original 2004 system: hand-coded BPTT in C++.

## PROFESSIONAL EXPERIENCE

---

### STREAMIO AI | Oakland, CA

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

- Designed and implemented OpenClaw multi-agent orchestration framework with gateway protocol, subagent delegation, profile management, and session switching — enabling coordinated AI agent workflows that decompose tasks across real estate, insurance, health/dental, and financial markets domains; directly analogous to agent planning and execution frameworks that balance autonomy with predictability.
- Integrated Claude MCP SDK to expose screen capture tools to AI coding assistants, defining agent extensibility primitives — tool access, codebase context, and external service calls — that mirror the MCP and plugin architecture central to the Agent Harness.
- Shipped MCP server and embedded terminal (xterm.js/node-pty, 20+ slash commands) enabling developers to observe and steer agent behavior in real time; built cross-platform desktop builds (macOS DMG with code signing/notarization, Linux .deb) with native ScreenCaptureKit integration via Swift.
- Built production Electron + React + TypeScript desktop application with 100+ components, Redux Toolkit state management, real-time HLS livestreaming pipeline (up to 9 concurrent sources, FFmpeg transcoding, WebSocket layer), and end-to-end auth/payments (Kinde OAuth 2.0, Stripe, Electron SafeStorage).

### FINTELLECT AI | Oakland, CA

Founder & CEO — LLM Agent Platform (iOS/Android/Web) 09/24 – Present

- Architected multi-provider LLM orchestration layer (Claude, GPT-4, Gemini) with fallback routing, structured output validation, and token budget optimization — implementing failure recovery and error handling patterns essential to reliable agent execution.
- Built domain-specific Fintellect Agents with RAG retrieval (ChromaDB), each scoped to distinct financial focal points — designing agent decomposition, context scoping, and guardrails that keep agents on productive paths without micromanagement.
- Executed 0-to-1 product strategy through App Store launch, incorporating customer discovery and iterative refinement based on empirical usage patterns.

### INTUIT | Mountain View, CA

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

- Achieved 275% YoY growth in ICE 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.
- Extended Java and Python SDK Starter Kits with scaffolding templates, build configurations (Gradle/Maven), testing frameworks, and CI/CD integration — empowering developers to go from zero to production-ready microservice in minutes; delivered ICE Self-Service DevPortal reducing onboarding from 2–3 weeks to <24 hours.
- Worked closely with telemetry and usage data (SQL, BigQuery) to prioritize developer pain points across ~20 mobile apps and 30+ product SKUs — the same measurement discipline needed to analyze agent traces at scale and turn patterns into concrete improvements.
- Conducted enterprise-wide Service Language Assessment across 9 languages, analyzing usage data and developer feedback to inform strategic decisions presented to CTO; built Asterias, a declarative asset lifecycle management platform with GraphQL API.

**SPLUNK INC.** | San Francisco, CA

*Senior Product Manager — Search Orchestration* 01/19 – 01/21

- Owned Search Service (Go microservices), Search Catalog (PostgreSQL metadata service), and Splunk Processing Language (SPL/SPL2) — built roadmaps, PRDs, and acceptance criteria for Splunk Cloud Services; delivered Scheduler Service end-to-end in ~4 months, demoed at .conf19.
- Led query performance benchmarking initiative, building mirrored Enterprise topology for benchmark testing and achieving up to 10x performance improvements — applying the same empirical, measurement-first approach to defining what 'good' looks like for complex system behavior.

**KAISER PERMANENTE** | Pleasanton, CA

*SOA Technical Product Manager* 08/12 – 12/18

- Led enterprise rollout of Splunk Logging-as-a-Service (1.7 TB daily volume, 200+ internal customers), Redis/XC10 caching for fault tolerance, and capacity planning across multiple datacenters — building platform reliability and observability discipline at enterprise scale.

**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% — developing the failure-mode diagnosis discipline central to agent trace analysis.

## 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

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

## 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