

ANTHONY MAIO

AI Platform / LLMOps Engineer - Agent Reliability, Evaluation, and Secure Systems

Danbury, CT (Remote US) | anthony.maio@gmail.com | linkedin.com/in/anthony-maio | github.com/anthony-maio | making-minds.ai

SNAPSHOT

Staff/Lead engineer building reliable, secure, observable distributed systems and applying the same discipline to LLM and agent systems (evaluation, monitoring, cost/latency controls).

20 years across IAM/security, fintech, and high-volume platforms. Known for measurable outcomes: 99.99% uptime, ~20% latency improvement, ~\$654K/year savings (2FA + cloud), 5K tx/sec at <10ms latency, and 5TB/day compliance pipelines.

Led 8-18 engineers (often without formal authority) through modernization, reliability programs, and 0-to-1 delivery under ambiguity.

CORE SKILLS

Applied AI / LLMOps: Eval harnesses, regression testing, monitoring/observability, failure-mode analysis, RAG foundations, safe rollout and measurement

Distributed Systems: Event-driven architecture, Kafka, CQRS/event sourcing, PostgreSQL, Redis, reliability engineering

Security / IAM: AuthN/AuthZ, MFA/2FA, OAuth/OIDC patterns, API security, audit/compliance pipelines

Cloud / Platform: AWS, Kubernetes, Terraform, CI/CD, incident response, cost optimization

Languages: Python, C#/.NET

PROFESSIONAL EXPERIENCE

MAKING-MINDS.AI

Independent AI Systems Engineer (LLMOpS, Eval, Reliability)

Remote

Sep 2024 - Present

- Shipped evaluation and monitoring tooling for LLM/agent systems focused on failure detection under real operational constraints (latency/cost/throughput)
- Built automated verification workflows (test-first + gated trust) to reduce silent failures in generated code and tool-using agents
- Released open-source toolkits with reproducible datasets/benchmarks and documented operational tradeoffs and limitations for deployment teams

DRAFTKINGS

Staff Software Engineer, Identity and Access Management

Remote

Feb 2023 - Sep 2024

- Tech lead and mentor for 8-engineer remote IAM team; improved throughput ~5x (8 to 40 story points/sprint over ~9 months) via coaching, ownership norms, pairing, and review workflow improvements
- Led 2FA SMS provider migration, maintaining 99.99% login uptime while improving auth latency (~20%) and reducing spend by ~\$450K/year; aligned engineering, compliance, and finance
- Modernized legacy auth services toward cloud-native patterns (service decomposition, reliability hardening, operational tooling) with zero downtime
- Reduced AWS spend by ~\$204K/year through right-sizing, targeted serverless adoption, and autoscaling (HPA) tuning
- Built/operated compliance data flows (Kafka to Redshift), processing ~5TB/day to support regulatory reporting and audits

Lead Software Engineer, R&D (Trading)

Aug 2022 - Feb 2023

- Led early-stage technical strategy for two remote teams; delivered 0-to-1 MVP for greenfield trading platform

- under changing regulatory constraints
- Designed event-sourced CQRS core sustaining ~5K transactions/sec at <10ms latency (PostgreSQL + Kafka)
- Built self-service experimentation foundations reducing feature launch time from days to <1 hour

BROADRIDGE*Lead Software Engineer*

Remote

Jan 2015 - Aug 2022

- Architected Kafka-based workflow orchestration across ~22 interdependent business processes; improved resilience and operational visibility
- Led staged modernization from monolithic .NET/WCF to AWS microservices as foundation for ~\$2B AUM platform
- Onboarded 6 enterprise clients via staged MVP delivery, enabling \$50M+ in revenue

TWOFOUR SYSTEMS*Product Engineer*

White Plains, NY

Jul 2007 - Dec 2014

- Built bespoke trading/risk platforms for Tier-1 financial institutions; delivered real-time risk/margin capabilities
- Improved delivery reliability via CI/CD and deployment practices (containerized builds, automated pipelines, blue/green patterns)

EDUCATION

Binghamton University - B.E., Computer Engineering**SELECTED PUBLIC WORK**

Open-source tooling and technical writing: making-minds.ai and github.com/anthony-maio