

Keith Steward, Ph.D.

Hands-On Head of AI Engineering | Principal AI Engineer

Boston, MA | 508-318-8035 | keith.steward@gmail.com | [LinkedIn](#) | [IntelligentDataWorks.ai](#) | [Portfolio](#)

PROFILE

Hands-on AI engineering leader who architects, builds, and operates production LLM applications, retrieval systems, agentic workflows, APIs, and cloud platforms. Built a GenAI platform supporting 70+ automations across 20+ workflows using Python, LangGraph, RAG, tool execution, PostgreSQL/PGVector, evaluation, guardrails, and observability. Combines principal-level engineering depth with architecture ownership, technical strategy, mentoring, and delivery leadership. Brings financial-services experience and the judgment to improve model quality, system reliability, scalability, and business impact from concept through production.

RELEVANT TECHNOLOGIES

Python • LLM applications • RAG • AI agents • LangGraph • LangChain • Vector databases
FastAPI • REST APIs • PostgreSQL • PGVector • AWS • Distributed systems • Evaluation • Observability

SELECTED TECHNICAL IMPACT

- Architected and built *AssistX*, a production GenAI platform supporting 70+ automations across 20+ workflows and accelerating selected work up to 50x.
- Implemented retrieval over labor-law sources from all 50 states using crawling, embeddings, PostgreSQL/PGVector, semantic search, grounding, and deterministic controls.
- Built production agentic workflows spanning model reasoning, tool execution, validation, evaluation, guardrails, observability, and human review; deployed AWS agents with Bedrock AgentCore and CDK.
- Advised 200+ enterprises on production AI, data, and cloud systems at AWS; created 170+ reusable technical assets and trained 180+ architects.

PROFESSIONAL EXPERIENCE

Technical Co-Founder | Intelligent DataWorks inc | Boston, MA | Aug 2023–Present

Hands-on architect and engineer for an applied-AI SaaS platform spanning agent orchestration, retrieval, enterprise data, full-stack delivery, and cloud operations.

- Built 100% of the reusable AssistX core platform and the first 80% of its HR application using Python, FastAPI, PostgreSQL/PGVector, LLM APIs, RAG, agents, LangChain, LangGraph, REST services, and AWS.
- Designed multi-step agent workflows with model reasoning, deterministic logic, tool execution, validation, evaluation, guardrails, observability, and human review for production reliability.
- Built and deployed a customer-service agent with AWS Bedrock AgentCore and Strands, using AWS CDK for infrastructure and runtime deployment.
- Operated the production stack end to end—systems administration, security, testing, AWS deployment, monitoring, troubleshooting, and release management—while guiding an 11-engineer team.
- Delivered the first commercial release within eight months; 70+ automations now span 20+ workflows, with selected work accelerated up to 50x and projected annual savings above \$50K per customer.

Sr. Engineering Director, Workforce Management Service | SuccessKPI | Jun 2022–Aug 2023

Player-coach for an AI-powered workforce-management SaaS service integrated with enterprise contact-center platforms.

- Owned architecture and hands-on implementation across Python, ML forecasting, operations-research scheduling, APIs, PostgreSQL, AWS, data pipelines, automated testing, and production delivery.
- Recruited and led seven distributed engineers through design reviews, feature decomposition, Agile execution, code quality, documentation, release planning, and operations; shipped the initial platform in six months.

Sr. Director, Data Analytics & Operations | TriMark USA | Mansfield, MA | Aug 2021–May 2022

Hands-on owner of a cloud data platform integrating 11 heterogeneous enterprise ERP environments.

- Designed and deployed the first AWS data lake in seven months using PySpark, Glue, S3, JDBC, AWS CDK, CI/CD, catalogs, and automated data pipelines.
- Reduced cross-ERP reporting from roughly two months to one day—a 60x improvement—while expanding an upskilled four-person team to seven engineers.

Technical Lead, Cloud Data Privacy & AI Governance | MassMutual | Boston, MA | Nov 2020–Aug 2021

Designed production engineering, security, and governance patterns for regulated enterprise AI and data systems.

- Established patterns for CI/CD, model ownership, transparency, bias and drift monitoring, auditability, human oversight, security, and responsible AI across five initiatives.

- Built a serverless privacy platform with Python, REST APIs, Lambda, Kafka, Aurora PostgreSQL, and AWS services; delivered it in eight months and automated roughly 90% of a manual workflow.

VP of AI/ML | NorthBay Solutions | Boston, MA | Jan 2019–Nov 2020 (Role ended in company-wide COVID layoff).

Recruited from AWS by the CEO to establish an AI/ML practice while remaining hands-on in platform architecture and implementation.

- Built the first Data Lake + ML Accelerator full-stack with Python, Django, React, HTML/CSS, APIs, and AWS, then guided a 24-person remote engineering team toward repeatable delivery.
- Delivered production AI, analytics, and knowledge-graph systems for enterprise customers including financial-services firms such as Invesco.

Sr. Specialist Solutions Architect, AI/ML, Big Data & Analytics | Amazon Web Services | Boston, MA | May 2015–Dec 2018

Served as AWS's first Senior Specialist Solutions Architect for AI/ML, combining hands-on architecture with enterprise delivery and technical coaching.

- Advised 200+ enterprise customers—including financial-services organizations—on AI/ML, distributed data systems, APIs, cloud architecture, security, reliability, and production adoption.
- Developed an early agent architecture in 2018; created demos and hackathons with SageMaker, TensorFlow, EMR, Spark, and other AWS AI/data services.
- Founded technical communities of 80 and 100 specialists, trained 180+ architects, created 170+ enablement assets, contributed to certification exams, and delivered 50+ presentations.

EDUCATION & RECOGNITION

- Ph.D., Molecular Biology / Microbiology & Immunology, University of Western Ontario; Canadian Medical Research Council Post-Doctoral Research Fellowship.
- AWS ML Elite Gold and Platinum; AWS re:Invent Rockstar; AWS Certification SME; U.S. Patents 8,793,073 and 8,489,334.

TECHNOLOGIES

Python • LLM applications • RAG • AI agents • Agentic AI • LangGraph • LangChain • Prompt engineering
 Tool calling • AI evaluation • Guardrails • Observability • Responsible AI • FastAPI • REST APIs
 Microservices • Distributed systems • PostgreSQL • PGVector • SQL • Embeddings • Semantic search
 Vector databases • AWS • Amazon Bedrock • Bedrock AgentCore • Strands • SageMaker • AWS CDK
 Docker • Kubernetes • Lambda • Step Functions • API Gateway • EventBridge • Kafka • PySpark
 Data pipelines • CI/CD • Automated testing • Git/GitHub • Linux • React • SaaS