

Keith Steward, Ph.D.

Enterprise GenAI Systems Architect | Agentic Platforms

Boston, MA | 508-318-8035 | keith.steward@gmail.com

[LinkedIn](#) | [IntelligentDataWorks.ai](#) | [IDW LinkedIn](#)

PROFILE

Hands-on architect who designs and delivers secure enterprise Generative AI systems from use-case selection through production operation. Combines LLMs, RAG, agents, multi-agent orchestration, vector retrieval, Python, APIs, microservices, cloud-native infrastructure, evaluation, guardrails, governance, and cost management. Built the AssistX core and first 80% of its initial application, enabling 70+ automations across more than 20 workflows with 99.9% availability. Brings practical depth across architecture, implementation, technical mentorship, reliability, and continuous AI quality improvement.

CORE TECHNOLOGIES

Generative AI • LLMs • RAG • AI Agents • Multi-Agent Systems • Python • LangChain • LangGraph • MCP • OpenAI API • Amazon Bedrock • Google Vertex AI • PGVector • AI Evaluation • AI Guardrails

SELECTED CAREER HIGHLIGHTS

- Architected and built AssistX using LLMs, RAG, agents, tool calling, Python, FastAPI, PostgreSQL/PGVector, microservices, APIs, and AWS.
- Designed reusable retrieval and agent patterns supporting document ingestion, embeddings, semantic search, context management, orchestration, and enterprise integration.
- Implemented evaluation, LLM-as-judge quality measurement, guardrails, model routing, token-cost controls, security, observability, and 99.9% availability.
- Evaluated GPT, Claude, Amazon Bedrock, Google Vertex AI, Cerebras, and open-weight models while mentoring 11 engineers and guiding production releases.
- Advised 200+ enterprises on AI/ML, cloud, data, security, and architecture while translating business requirements into practical solution designs.

PROFESSIONAL EXPERIENCE

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

Build and operate a production AI platform while owning strategy, architecture, delivery, governance, security, and cloud operations.

- Architected and built the AssistX core platform and first 80% of its initial application hands-on using Python, FastAPI, LLMs, RAG, agents, PostgreSQL/PGVector, APIs, Docker, AWS, and CDK.
- Recruited and technically led 11 engineers while owning a \$900,000 budget, product priorities, architecture, releases, security, and operations.
- Delivered 70+ automations across more than 20 workflows, accelerating selected work by up to 50x and sustaining 99.9% availability.
- Established evaluation, guardrails, human oversight, observability, model routing, cost controls, incident response, and release practices.
- Initiated strategic acquisition discussions regarding IDW's AI platform as part of a planned transition to a senior AI/engineering leadership role.

Senior Director of Engineering | SuccessKPI | Remote | Jun 2022 to Aug 2023

Built and led engineering for an AI-powered workforce-management component of a cloud-native contact-center SaaS platform.

- Recruited and led an 11-person organization, including a manager, spanning software, data, platform, security, BI, and quality engineering.
- Owned Agile execution, architecture, operational priorities, performance reviews, bonus allocation, team capacity, and a \$600,000 budget.
- Delivered forecasting, scheduling, optimization, analytics, APIs, and integrations for enterprise and government customers.

Senior Director, Data Analytics & Operations | TriMark USA | Mansfield, MA | Aug 2021 to May 2022

Established the enterprise data strategy, engineering team, architecture, roadmap, and operating capability across a fragmented multi-division environment.

- Aligned executives around more than 50 analytics use cases and secured funding for an AWS data-lake strategy and delivery roadmap.

- Upskilled an analytics team with no prior AWS experience and expanded it to seven cloud data engineers.
- Delivered the first enterprise data lake across 11 divisions in seven months, reducing cross-ERP reporting from roughly two months to one day.

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

Led architecture and hands-on implementation across five enterprise Data Privacy and AI Governance initiatives.

- Led architecture and engineering across five enterprise Data Privacy and AI Governance initiatives spanning privacy rights, enterprise metadata, policy automation, and governance APIs.
- Advised the AI Governance and Ethics organization on model transparency, bias, drift, ownership, risk classification, monitoring, human oversight, and auditability.
- Established AWS security, IAM, API, event-driven, and CI/CD patterns while coordinating Product, Enterprise Architecture, SRE, governance, and delivery teams.

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

Recruited out of AWS by the NorthBay Solutions CEO to establish a new AI/ML practice and reusable production data-and-ML capability.

- Built an AWS Data Lake and ML accelerator hands-on with Python, Django, React, APIs, and AWS, then directed a 24-person remote team toward delivery.
- Delivered an AWS Neptune knowledge graph integrating complex hierarchical enterprise data and expanded reusable cloud and AI architectures for customers.
- Advised enterprise customers on AI strategy, data platforms, knowledge graphs, solution architecture, and production implementation.

Senior Specialist Solutions Architect, AI/ML & Big Data | Amazon Web Services | Boston, MA | May 2015 to Dec 2018

Advised enterprises and AWS teams on production machine learning, distributed analytics, data platforms, and cloud adoption.

- Advised 200+ enterprises on AI/ML strategy, production architecture, SageMaker, data lakes, distributed systems, security, and scalable adoption.
- Created and led global Big Data and AI/ML communities, developing 180+ specialists through standards, mentoring, workshops, and reusable assets.
- Built a SageMaker and TensorFlow forecasting hackathon for Morgan Stanley data scientists and created 170+ technical enablement assets.
- Delivered 50+ technical presentations and workshops and earned AWS ML Elite Gold and Platinum recognition.

EARLIER RELEVANT EXPERIENCE

- Built production cloud and software platforms spanning AI, data, regulated applications, customer-facing SaaS, fault-tolerant operations, security, measurable service reliability, integration architecture, release automation, and incident-aware operations.
- Built an AWS data-and-ML accelerator hands-on and delivered enterprise AI/ML architecture guidance.

EDUCATION & RECOGNITION

Ph.D., Molecular Biology / Microbiology & Immunology, University of Western Ontario. AWS ML Elite Gold and Platinum; AWS re:Invent Rockstar; named inventor on multiple patents.

TECHNOLOGIES

Generative AI • LLMs • RAG • AI agents • multi-agent systems • prompt engineering • tool calling • function calling • workflow orchestration • LangChain • LangGraph • MCP • OpenAI API • Anthropic Claude • Amazon Bedrock • Google Vertex AI • open-weight models • Python • FastAPI • REST APIs • microservices • event-driven architecture • embeddings • vector search • PostgreSQL • PGVector • AWS • SageMaker • TensorFlow • AI evaluation • LLM-as-judge • AI guardrails • Responsible AI • AI governance • MLOps • LLMOps • observability • security • IAM • Docker • Kubernetes • CI/CD • AWS CDK