

Keith Steward, Ph.D.

Generative AI Systems Architect | Enterprise Agent Platforms

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

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

PROFILE

Hands-on Generative AI architect who designs and delivers secure enterprise systems from use-case selection through production operation. Combines LLMs, RAG, agents, vector retrieval, Python, APIs, microservices, cloud-native infrastructure, evaluation, guardrails, governance, and cost management with technical mentorship. Built the AssistX core platform and helped deliver 70+ AI automations across more than 20 workflows, accelerating selected work by up to 50x. Brings practical depth across architecture, implementation, reliability, and continuous AI quality improvement.

CORE EXPERTISE

Generative AI • LLMs • RAG • AI Agents • Agentic Workflows • Python • LangChain • LangGraph • OpenAI API • Amazon Bedrock • PGVector • REST APIs • AI Evaluation • AI Guardrails • AWS

SELECTED TECHNICAL IMPACT

- 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, workflow orchestration, and enterprise integration.
- Implemented evaluation, LLM-as-judge quality measurement, guardrails, model selection, token-cost controls, security, observability, and 99.9% uptime.
- Evaluated GPT, Claude, Amazon Bedrock, Google Vertex, Cerebras, and open-weight models while mentoring 11 engineers and guiding production releases.

PROFESSIONAL EXPERIENCE

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

Architect and build AI product, platform, and cloud capabilities while guiding engineering execution and operations.

- 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, workflow orchestration, and enterprise integration.
- Implemented evaluation, LLM-as-judge quality measurement, guardrails, model selection, token-cost controls, security, observability, and 99.9% uptime.
- Evaluated GPT, Claude, Amazon Bedrock, Google Vertex, Cerebras, and open-weight models while mentoring 11 engineers and guiding production releases.
- 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

Directed architecture and engineering for an AI-powered contact-center Workforce Management SaaS product.

- Recruited and led an 11-person organization including a manager, with responsibility for hiring, coaching, performance reviews, bonus allocation, and a \$600,000 budget.
- Directed Agile delivery across security, data, platform, BI, and quality engineering, translating product priorities into forecasting, optimization, integrations, and releases.
- Delivered an AI-powered workforce-management SaaS product for enterprise and government customers using AWS, Python, PostgreSQL, forecasting, and optimization.

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

Owned enterprise data modernization for a \$1B+ organization integrating 11 acquired businesses and heterogeneous ERP systems.

- Prioritized 50+ analytics use cases with Finance, Marketing, Supply Chain, and technology leaders and secured support for an AWS data strategy.
- Upskilled four analysts and grew the team to seven AWS engineers with hiring, mentoring, planning, budget, risk, and delivery accountability.
- Established nightly data ingestion, cataloging, consolidation, normalization, and global querying across 11 geographically distributed ERP environments.

- Delivered the first AWS data lake within seven months and reduced 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.

- Converted policy and product requirements into secure cloud architecture, APIs, automated infrastructure, testing, CI/CD, and production releases.
- Delivered the first cloud-native privacy-rights platform within eight months, automating approximately 90% of a manual workflow.
- Partnered with Enterprise Architecture and SRE on IAM, serverless patterns, managed services, infrastructure automation, resilience, and cloud operating standards.
- Guided model-governance practices for evaluation, transparency, bias, drift, monitoring, auditability, lifecycle controls, security, and human oversight.

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, and fault-tolerant operations.
- At AWS, advised 200+ enterprises on AI/ML, data, security, and cloud architecture and developed 180+ technical specialists.

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 • prompt engineering • AI agents • agentic workflows • tool calling • function calling • memory • workflow orchestration • LangChain • LangGraph • MCP • OpenAI API • Amazon Bedrock • Google Vertex • open-weight models • Python • FastAPI • REST APIs • microservices • event-driven architecture • embeddings • vector search • PostgreSQL • PGVector • AWS • SageMaker • AI evaluation • LLM-as-judge • AI guardrails • responsible AI • AI governance • MLOps • observability • security • IAM • Docker • Kubernetes • CI/CD • AWS CDK