

Keith Steward, Ph.D.

Agentic AI Engineer | Production AI Systems Builder

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

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

PROFILE

Hands-on AI engineer who turns ambiguous product workflows into production LLM, agentic, and retrieval systems. Architected and helped build a commercial AI platform supporting 70+ automations across 20+ workflows, with selected work accelerated up to 50x and the first release shipped in eight months. Combines Python, Django, React, PostgreSQL/PGVector, RAG, LangChain/LangGraph, OpenAI and Anthropic models, AWS, MLOps, and cloud deployment. Brings player-coach depth to rapid prototyping, production scaling, and close Product, Design, and engineering collaboration.

SELECTED QUALIFICATIONS

- **Production Agentic AI:** LLM applications, AI agents, RAG, embeddings, vector search, tool orchestration, evaluation, observability, reliability, and cost awareness
- **Full-Stack Engineering:** Python, FastAPI, Django, React, PostgreSQL/PGVector, REST APIs, data pipelines, cloud services, and user-facing workflow integration
- **Product Delivery:** discovery, rapid prototyping, architecture, implementation, deployment, feedback-driven iteration, Agile Scrum, CI/CD, and production operations
- **Cloud & MLOps:** AWS, SageMaker, Bedrock AgentCore, CDK, Docker, Kubernetes, monitoring, security, governance, testing, and scalable platform architecture

TECHNICAL TOOLKIT

Python • Django • React • PostgreSQL/PGVector • RAG • Vector Search • OpenAI • Anthropic Claude • LangChain • LangGraph • AI Agents • FastAPI • AWS • Docker • Kubernetes

PROFESSIONAL EXPERIENCE

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

Architect and build production agentic-AI software while contributing directly to engineering and cloud operations.

- Architected and helped build AssistX, a production AI platform combining Python, FastAPI, PostgreSQL/PGVector, OpenAI and Anthropic models, RAG, agents, APIs, LangChain, LangGraph, workflow orchestration, and AWS services.
- Shipped the first commercial release within eight months; the platform supports 70+ AI automations across more than 20 workflows and accelerates selected work by up to 50x.
- Recruited and technically led an 11-engineer Agile team while coding and contributing directly to architecture, backend services, cloud infrastructure, orchestration, security, and platform operations.
- Implemented a customer-service AI agent with AWS Bedrock AgentCore and Strands, deployed to the managed agent runtime using AWS CDK.
- Built a full-stack product spanning FastAPI services, PostgreSQL/PGVector, Streamlit, HTML/CSS, voice input, semantic retrieval, and multi-step automation.
- Developed retrieval over labor-law source material from all 50 states, using embeddings and PostgreSQL/PGVector to answer open-ended questions through grounded semantic search.
- Evaluated OpenAI, Anthropic Claude, Amazon Bedrock, Cerebras, and open-weight models while balancing quality, latency, cost, security, and workflow requirements.
- Initiated strategic acquisition discussions regarding IDW's AI platform as part of a planned transition to a senior AI/engineering role.

Sr. Director of Engineering | SuccessKPI | Remote | Jun 2022 to Aug 2023

Built an AI-enabled Workforce Management SaaS capability for enterprise contact-center operations.

- Translated detailed Product requirements into AWS architecture, algorithms, software design, and engineering work breakdown for forecasting, optimization, scheduling, and analytics features.
- Led a seven-member distributed team through two-week Agile sprints, unit testing, automated deployment, CI/CD, and collaboration with Product and QA.
- Delivered the first production platform within six months and onboarded two large enterprise customers; later releases served four customers including U.S. government agencies.
- Implemented early LLM applications and AI-driven forecasting and optimization workflows integrated with Amazon Connect, Genesys, and Talkdesk.

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

Built a production AWS data platform and automated pipelines across 11 heterogeneous enterprise systems.

- Partnered with Marketing, Supply Chain, Finance, and technology stakeholders to gather and prioritize more than 50 analytics use cases and translate them into a delivery roadmap.
- Delivered TriMark's first AWS data lake within seven months, automating nightly ingestion, cataloging, consolidation, and normalization while reducing selected reporting work from two months to one day.

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

Led hands-on cloud engineering across enterprise data-privacy and AI-governance initiatives.

- Partnered with Product Managers to define requirements and AWS architectures, then guided developers through Scrum delivery using test-driven development and CI/CD.
- Advised on MLOps, model testing, monitoring, transparency, bias, drift, observability, security, and auditability using Amazon SageMaker.
- Delivered a cloud-native privacy-rights platform within eight months, automating approximately 90% of a formerly manual workflow.

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

Built production AI, ML, and data-platform capabilities while remaining hands-on in implementation.

- Built the initial AWS Data Lake + ML Accelerator hands-on using Python, Django, React, APIs, and AWS, then guided a distributed team toward customer demonstrations and production delivery.
- Partnered with the CEO, Product, and customers to translate needs into architecture, ML capabilities, APIs, prototypes, and implementation plans.
- Delivered applied-ML solutions including cancellation prediction and bestseller forecasting for enterprise customers.
- Worked across customer discovery, rapid prototyping, architecture reviews, and implementation planning in a fast-moving AWS consulting environment.

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

Developed hands-on AI/ML solutions and advised enterprises on production AWS architectures.

- Advised 200+ customers on AI/ML, data platforms, and cloud architecture while building technical examples with SageMaker, TensorFlow, Spark, APIs, and managed AWS services.
- Developed an early agent architecture presented at AWS re:Invent and created a SageMaker/TensorFlow ML hackathon for Morgan Stanley.
- Built technical demonstrations and reference architectures across SageMaker, TensorFlow, EMR, Spark, Glue, S3, Lambda, Step Functions, and related AWS services.
- Trained 180+ global Solution Architects and earned AWS ML Elite Gold and Platinum recognition.

EARLIER RELEVANT EXPERIENCE

- Built and operated a fault-tolerant AWS/Cassandra platform for an FDA-cleared connected dental product at Dimensional Photonics.
- Built BioWizard, a 400-rule AI system for genomic drug-target evaluation, and helped create the globally adopted Ingenuity Pathway Analysis platform.

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

Python • FastAPI • Django • React • HTML • CSS • REST APIs • PostgreSQL • PGVector • SQL • LLMs • OpenAI API • Anthropic Claude • Generative AI • AI agents • RAG • embeddings • vector search • LangChain • LangGraph • MCP • prompt engineering • AI evaluation • MLOps • observability • data pipelines • workflow orchestration • AWS • Amazon SageMaker • Bedrock AgentCore • AWS CDK • Docker • Kubernetes • CI/CD • TensorFlow • Spark • security • Linux • Git