

JAY OZER

San Francisco, CA | 415-515-4700 | jayozzer@gmail.com

[linkedin.com/in/jayozzer](https://www.linkedin.com/in/jayozzer) | github.com/jayozzer | medium.com/@jayozzer | dataacrobat.com

SUMMARY

AI Engineering Manager with 10+ years building production-grade AI/ML systems in regulated environments. Built and shipped cross-functional engineering teams delivering enterprise-scale agentic AI solutions that achieved 99% processing time reduction while maintaining strict compliance and security standards. Deep hands-on expertise in Python, RAG architectures, multi-agent systems, and MLOps—combined with strategic leadership that scales teams from 0 to 6 engineers. Architected data engineering and business intelligence systems at Federal Reserve Bank processing terabytes of economic data for monetary policy decisions, then led applied AI pods at Doma delivering breakthrough innovations that broke the 30-year monopoly of the title industry—catalyzing Doma's strategic transformation from title-and-tech company to pure-tech platform serving wholesalers. Passionate about staying close to code while building high-performing teams that ship reliable, responsible AI at scale.

TECHNICAL EXPERTISE

- Languages & Frameworks: Python (primary), SQL, JavaScript
- AI/ML Stack: RAG, LLM Fine-Tuning (OpenAI, Fireworks.ai, TogetherAI), Multi-Agent Systems (AutoGen, CrewAI, LangGraph), PyTorch, Prompt Engineering, Agent Frameworks (AgentBuilder, Claude Agents SDK), semantic search, embeddings
- Vector Databases: Pinecone, Qdrant, ChromaDB, ClickHouse
- MLOps/LLMOps: CI/CD pipelines, Docker, model registries, feature stores, A/B testing, canary deployments, offline/online evaluation, observability, cost management
- Data Engineering: Snowflake, Postgres, Airflow, Fivetran, dbt, ETL/ELT pipelines, real-time data processing
- Cloud & Infrastructure: AWS (Lambda, S3, CloudWatch), Azure (AI Foundry, Power Automate), GPU workflows
- Security & Compliance: Guardrails, GDPR, IAM, data minimization, content safety
- Leadership: Agile/Scrum, Technical Mentorship, Cross-functional Team Building, Hiring & Performance Management

PROFESSIONAL EXPERIENCE

DOMA TECHNOLOGY, San Francisco, CA

Senior Manager, Applied AI (Mar 2023 – Present)

Led applied AI engineering pod delivering production agentic AI systems for title insurance automation, partnering with Fannie Mae and Blend to transform real estate workflows at enterprise scale. Built breakthrough products that achieved 99% processing time reduction, broke the 30-year monopoly of the title industry, and catalyzed Doma's strategic transformation from title-and-tech company to pure-tech platform serving wholesalers.

Agentic AI & Multi-Agent Systems

- Architected and deployed DeedIQ, an autonomous 10+ agent LLM system using CrewAI that eliminated manual title underwriting and reduced deed processing time by 99% (30min-4hrs → seconds per deed), breaking the industry's 30-year dependency on manual processes
- Implemented dual-team extraction with cross-validation, county-specific pattern recognition, and self-learning capabilities achieving 97%+ accuracy in testing and 100% accuracy in production with zero errors to date
- Established production evaluation frameworks with human-in-the-loop validation ensuring 100% accuracy through systematic review cycles, guardrails, and model performance tracking against acceptance thresholds
- Designed ClerkIQ automated document retrieval system combining browser automation with intelligent navigation across 200+ county clerk websites, reducing per-document costs by 99% (\$40 → \$0.50) with projected annual savings of \$2.3M + 11 FTE at full deployment
- Built IntentIQ production email triage system processing 14K+ emails daily through sophisticated two-stage AI pipeline for intelligent workflow categorization across 7 distinct title insurance categories

MLOps & Infrastructure

- Built and maintained CI/CD pipelines for LLM deployment at enterprise scale with automated model versioning, canary deployments, and rollback capabilities
- Implemented automated LLM evaluation frameworks using RAGAS methodology for continuous quality monitoring in production
- Designed observability infrastructure with real-time monitoring, cost tracking, and performance optimization for production AI systems
- Led database modernization through LLM-powered automation tools, migrating 25K+ lines of SQL from SQL Server to Snowflake—reducing 6-month project timeline to 1 month through intelligent code transformation (published simplified tool on OpenAI GPT Store)

Strategic Partnerships & Analytics

- Architected analytics infrastructure and reporting systems for strategic partnerships with Fannie Mae and Blend, creating data visualizations published in Wall Street Journal
- Built comprehensive data acquisition pipelines, predictive models, and performance dashboards driving successful product launches through data-driven insights

Manager, Data Engineering (Jul 2021 – Mar 2023)

Built and led high-performing data team of 6 engineers, establishing engineering excellence and modern data practices that scaled company-wide analytics.

Team Leadership & Technical Excellence

- Hired, mentored, and developed team of 6 data engineers and analysts, implementing structured technical assessments, coding standards, and career development frameworks
- Established code review processes, GitHub workflows, and engineering best practices—resolving 1000+ data quality issues through systematic peer review

Enterprise Data Integration & Performance

- Led enterprise-wide data integration initiative unifying data from three merged companies into centralized Snowflake analytics platform serving BI, analytics, and data science teams
- Modernized legacy reporting infrastructure achieving 10x dashboard performance improvement and 90% reduction in query times through query optimization and data modeling best practices

Staff Data Analyst (Jan 2020 – Jul 2021)

Transformed financial services data infrastructure by establishing unified source of truth and driving company-wide adoption of modern analytics tools for post-merger integration.

Data Quality & Governance

- Spearheaded company-wide adoption of Looker, consolidating disparate data sources into unified analytics platform enabling standardized reporting
- Architected enterprise-wide row-level security framework for materialized non-public data, collaborating with Risk and Compliance teams for IPO readiness—leading to promotion to Data Engineering Manager

Thought Leadership & Innovation

- Published machine learning research on Analytics Vidhya: predictive model for San Francisco restaurant inspection health scores achieving 90%+ accuracy to optimize Department of Health resource allocation post-COVID-19

FEDERAL RESERVE BANK OF SAN FRANCISCO, San Francisco, CA

Senior Data Analyst (Aug 2013 – Jan 2020)

Led data engineering and business intelligence infrastructure development processing terabytes of economic data for data-driven monetary policy decisions in highly regulated central banking environment.

Enterprise Data Platform & Infrastructure

- Architected FedQuery Hub using HDFS and Tableau—a centralized data platform with curated views enabling self-service analytics for monetary policy research
- Processed terabytes of economic data with high availability, strict security controls, and comprehensive audit trails meeting Federal Reserve compliance standards

- Implemented data governance framework establishing Fed-wide quality metrics that increased data reliability across multiple departments

EDUCATION & CERTIFICATIONS

Education

- MS, Data Science – Regis University, Denver, CO
- MBA, Sustainable Business – San Francisco State University, San Francisco, CA
- MS, Industrial Engineering – University of Central Florida, Orlando, FL
- BEng, Chemical Engineering – Newcastle University, UK

AI/ML Certifications & Teaching

- AI Engineering Bootcamp – AI Makerspace, San Francisco, CA
- Teaching Assistant – AI Engineering Bootcamp Cohort 6, AI Makerspace (mentoring engineers in production LLM systems, RAG, and multi-agent architectures)
- Advanced LLM Application Building & Fine-Tuning LLMs – Maven, San Francisco, CA

KEY ACHIEVEMENTS

Production AI Systems Delivered

- 14+ production LLM systems shipped across regulated industries (Title insurance, healthcare)
- Enterprise-scale data engineering and BI systems at Federal Reserve Bank processing terabytes of economic data for monetary policy decisions in highly regulated central banking environment
- 99% processing time reduction through intelligent automation (DeedIQ deed parsing, ClerkIQ document retrieval)
- Broke 30-year monopoly of title industry, catalyzing strategic transformation at Doma from title-and-tech to pure-tech platform
- \$2.3M+ annual projected savings from automated document retrieval at scale

Team Building & Leadership

- Scaled engineering teams from 0→6 members with structured hiring, mentorship, and performance management
- Established engineering rituals including design reviews, code reviews, and sprint planning across cross-functional pods
- Led teams through successful IPO preparation including data governance and security framework implementation
- Returned as Teaching Assistant for AI Engineering Bootcamp Cohort 6, mentoring engineers in production LLM systems and multi-agent architectures

PATENTS & NOTABLE PROJECTS

- Document Processing Patent (Pending: 19/080655) – DeedIQ intelligent deed analysis system filed March 14, 2025
- Automated County Document Retrieval Patent (Pending) – ClerkIQ intelligent document retrieval system to be filed Q1 2026
- PoppyNote: HIPAA-ready automatic transcription system with RAG-powered patient history chatbot using Pinecone for contextual retrieval. Generates structured SOAP notes for dental professionals on AWS infrastructure (S3, Lambda, CloudWatch), enabling clinicians to query prior visit notes before appointments
- Ask Poppy: Fine-tuned GPT-4o-mini chatbot with RAG for pediatric dentistry parent inquiries. Trained on 750+ verified Q&A pairs, integrated with GuardGroq (LLaMA Guard-powered guardrails). Reduced on-call weekend stress by handling routine inquiries, enabling focus on true emergencies
- KnowledgeFlow: Popular Streamlit app for Voiceflow knowledge base optimization using contextual retrieval (inspired by Anthropic's approach)—presented at Voiceflow developer meetup, widely adopted by community for improving RAG accuracy
- browser-use: Fixed critical duplicate download event bug improving download handling stability (merged PR to core library)