

Derek Fox

GenAI Engineer

derek-fox@mail.com | [linkedin.com/in/fox-derek](https://www.linkedin.com/in/fox-derek) | github.com/foxintheloop

Denver, CO | (720) 336-0774

GenAI Engineer with 4+ years of experience shipping production AI systems across enterprise consulting, finance, and telecom. Specializes in designing multi-agent architectures, building RAG pipelines, and developing LLM evaluation tooling at scale. Combines a Master's in Data Science (AI specialization) with hands-on delivery for Fortune 500 clients in regulated, high-accuracy environments.

CORE SKILLS

AI & LLM: Generative AI, Large Language Models, Retrieval-Augmented Generation (RAG), Agentic Systems, LLM Evaluation Frameworks, Prompt Engineering, Vector Databases, Embeddings, MCP Servers, Langchain, Hugging Face Transformers

ML & Data Science: Machine Learning, Deep Learning, NLP, Predictive Analytics, Classification, Ensemble Methods, Time-Series Forecasting, A/B Testing, Statistical Analysis

Engineering: Python, SQL, TypeScript, R, PySpark, TensorFlow, Node.js, Docker, Git, REST APIs, JSON

Cloud & Infrastructure: AWS (Bedrock, Redshift, Lambda), Google Cloud (Vertex AI, Gemini Agentspace), Microsoft Azure (Synapse, Fabric), Databricks

WORK EXPERIENCE

Generative AI Engineer Senior Consultant

Jun 2025 – Present

Deloitte

Denver, CO

- Built full-stack LLM-as-a-Judge evaluation framework that accelerated testing cycles by 90% for CFO-facing finance agentic application. Scaled tool from single functional team to company-wide adoption across 3 teams, improving test coverage and model quality by automating identification of functional bugs, data quality issues, and performance regressions.
- Architecting agents within 30-agent fleet on Google Gemini Agentspace focused on revenue leakage detection, cash collections optimization, and journal entry intelligence. Driving technical design and implementation for production-scale finance automation.

ML Engineer | Data Scientist Senior Consultant

Jan 2022 – Jun 2025

RevGen Partners

Denver, CO

- Built Customer Support Generative AI data product using AWS Bedrock and Langchain with RAG — utilized customer conversations, product catalogs, sentiment analysis, and conversation history to support customer service operations and recommend new products.
- Engineered machine learning driven lead routing engine with Microsoft Fabric and Azure Synapse. Incorporated Ranking, Ensemble, Classification, and Forecasting models.
- Built revenue assurance classification model that identified 5M at-risk accounts most likely to carry an unpaid balance, enabling targeted collection strategies.
- Engineered pricing optimization solution with genetic algorithm and demand forecasting models to determine optimal product and service pricing.
- Uncovered \$150M in annual telecom revenue share to divest through data pipeline development integrating 12 data sources.

Data Scientist | Analyst

Jan 2021 – Jan 2022

Freelance / Hack for LA / EnCiv

Remote

- Built time-series forecasting models, data pipelines (AWS Redshift), and statistical analysis tools across civic tech and nonprofit engagements. Contributed to LA 311-Data project and EnCiv's Unpoll application.

Business Systems Analyst

Dec 2019 – Oct 2020

Charles Schwab

Lone Tree, CO

- Designed reporting and automation systems using Teradata, Oracle, Apache Hive, SQL, and Python to optimize survey business processes and program evaluation.

PROJECTS

ReasonForge — github.com/foxintheloop/ReasonForge

Neurosymbolic mathematics toolkit for LLMs. Built 111 tools across 7 domain-specific MCP servers that route calculations through SymPy, NumPy, and SciPy for provably correct results — eliminating math hallucinations

entirely. Modular architecture allows LLM agents to compose multi-step reasoning chains with exact computation at each step.

UpTier — github.com/foxintheloop/UpTier

MCP-powered task manager with Claude AI integration. Full desktop application with Microsoft To Do-style UI, intelligent prioritization strategies (Eisenhower, quick wins, high impact), and goal tracking. Built end-to-end from MCP server design through TypeScript frontend to production deployment.

TransitGuard — github.com/foxintheloop/TransitGuard

Predictive safety intelligence for Chicago CTA, built as both a web application and mobile app. Trained XGBoost and DBSCAN clustering models on 50K+ crime incidents to identify high-risk transit stations before incidents occur. Includes a RAG-powered chatbot for natural language querying of safety data and predictions.

EDUCATION

Northwestern University

2023 – 2025

Master of Science, Data Science — AI Specialization

UC Berkeley

Bachelor of Arts, Pre-Law

CERTIFICATIONS

Microsoft Certified Data Science Associate • Databricks LLM • IBM Data Science