Barbara Hidalgo-Sotelo, PhD, CDMP

AI Engineering Consultant | Machine Learning Engineer | Data Science & GenAI

📞 512.804.9995   @ barbs@balex.com   🔗 https://barbhs.com   📍 Austin, TX (On-site or Remote)   🚩 US Citizen

SUMMARY

AI engineering consultant and data scientist with 6+ years delivering end-to-end machine learning and data products across professional services, public sector, and healthcare. Build Python-based ML and generative AI applications (LangChain/LlamaIndex; RAG patterns) and partner closely with engineers and stakeholders to translate business needs into reliable, well-documented solutions. Strong foundation in data engineering (SQL/ETL, data quality), cloud platforms (AWS/Azure/GCP), and pragmatic MLOps practices (Git, automated testing, CI/CD exposure).

CORE STRENGTHS

AI solution design & client delivery
Align stakeholders → Gather requirements → Design architecture → Execute implementation → Smoothly handoff

GenAI Application Engineering
LLM apps, RAG/GraphRAG patterns, Retrieval, Prompt iteration, Evaluation mindset

Machine Learning Development
Supervised/unsupervised; Model selection, Validation, Error analysis

Data Engineering & Preprocessing
SQL, ETL, Validation, Feature engineering, Metadata management

Cloud & Deployment foundations
AWS, Azure, GCP; Secure credential configs; Balancing performance/scalability considerations

Engineering Collaboration
Git workflows, Documentation for technical + non-technical audiences, Agile delivery

PROFESSIONAL EXPERIENCE

Independent AI Engineering (Project-Based)
Self-Directed Portfolio Work | May 2025–Present | Remote

- Built and iterated on Python-based GenAI applications using RAG/GraphRAG patterns, focusing on retrieval quality, traceability (citations/provenance), and reproducible workflows
- Developed local-first systems with a clear deployment path (environment configuration, modular services, provider abstraction) to support rapid iteration and future productionization
- Strengthened engineering fundamentals: Git/GitHub workflows, documentation, structured evaluation mindset, and repeatable pipelines

Technical Consultant (Program Delivery / Change Management)
Inflective, Inc. | January 2024 – May 2025 | Remote

- Partnered with client stakeholders to assess program status, risks, and quality; communicated progress and blockers to keep delivery moving
- Resolved cross-functional delivery issues across business and technical teams, reducing friction and supporting on-time execution
- Supported change-management planning and execution using internal methods and software

Data Scientist (Federal Client — Financial Markets & Data Governance)
Metric5 Consulting | April 2021 – January 2023 | Remote

- Built **Python** utilities to streamline **metadata export** and reporting workflows; produced documentation to support adoption & knowledge transfer
- Developed POC reporting utility integrating ServiceNow, enterprise data catalog & analytics outputs to improve handling of data access requests
- Collaborated with technical teams in a standards-driven environment; supported version control and automated testing pipelines

Technical Consultant & Project Manager (Risk management / Healthcare analytics)
Inflective, Inc. | Aug 2017 – May 2019 | Remote

- Designed a **Python / scikit-learn** classification model to automate medical bill categorization with performance comparable to human coders
- Led a digital transformation effort improving adjudication & analytics workflows, achieving **~30% faster** processing time, meaningful cost saving
- Co-led Lean Six Sigma workshops to identify process inefficiencies and operational improvements, reducing error rates by ~25%
- Delivered in an **Agile** environment (Jira), presenting analysis and recommendations to executive stakeholders

Technical Consultant (Requirements, QA/QC, Enablement)
Persis Consulting | 2014–2017 | Remote

- Documented software requirements & supported Agile delivery for healthcare & private-sector clients, aligning technical implementation to business needs
- Managed QA/QC testing and created end-user training materials to improve reliability and adoption of EMR software

Research Project Manager (Biometrics & Marketing Effectiveness)
MediaScience | Feb 2017 – July 2017 | Austin, TX

- Ran hypothesis and A/B testing using multimodal biometric data (eye tracking, GSR, facial expression signals) and delivered client-ready insights

PROFESSIONAL EXPERIENCE

Research Engineer/Scientist

The University of Texas at Austin. | **January 2013 – June 2016** | **Austin, TX**

- Engineered secure data collection and automated analysis workflows (Python/MATLAB) for longitudinal studies; supervised research staff and undergraduates contributing to the data pipelines

SELECTED PROJECTS

GraphRAG with Podcasts — knowledge graph + vector retrieval for podcast Q&A (Neo4j Aura, embeddings)

- Designed a multi-stage ingestion pipeline for podcast episodes and transcripts with checkpointing/idempotent re-runs for iterative testing
- Implemented GraphRAG-style retrieval patterns to support question answering with timestamp/citation-style grounding for traceability
- Modeled entities and relationships in Neo4j Aura to connect episodes, concepts, and sources for richer retrieval than vector-only search

Poolula Platform — modular RAG system with evaluation harness (FastAPI, SQLModel, ChromaDB)

- Built a local-first, modular RAG architecture combining a structured data layer and document retrieval (ChromaDB) with tool-style routing
- Implemented audit-friendly logging patterns & an evaluation workflow that scores responses against defined expectations/ground truth prompts
- Designed the system to be provider-agnostic (LLM backend abstractions) to support rapid client-oriented experimentation

ChronoScope — resume timeline extraction & visualization (Streamlit, LLMs, Plotly)

- Built a Streamlit application that extracts timeline events from resumes and generates interactive Plotly timelines and TimelineJS-ready exports
- Implemented local-first processing and export utilities to support privacy-sensitive workflows and repeatable iterations
- Added validation/quality checks to improve consistency of extracted events across runs

TECHNICAL SKILLS

PROGRAMMING: Python (NumPy, pandas, scikit-learn; exposure to PyTorch/TensorFlow), SQL (Postgres/MySQL/BigQuery), R, MATLAB, Bash

GenAI / LLM: LangChain, LlamaIndex, OpenAI GPT family; RAG patterns (including GraphRAG-style workflows); embeddings & retrieval; prompt iteration (evaluation-minded)

Vector / Search: ChromaDB; vector indexing/retrieval concepts; graph + vector hybrid retrieval patterns (Neo4j Aura + embeddings)

Knowledge Graphs: Neo4j Aura; graph data modeling; relationship-driven retrieval and provenance/citation patterns

APIs & Apps: FastAPI, Flask; Streamlit; basic full-stack patterns (React/JS exposure); logging and configuration patterns

Data Engineering: ETL pipelines, data validation/quality checks, feature engineering, metadata management; familiarity with Spark concepts

Cloud: AWS (EC2, RDS, Lambda; Bedrock exposure), Azure (AI Studio), GCP (BigQuery); Snowflake (currently ramping—strong SQL/ELT transfer)

MLOps / Collaboration: Git & GitHub workflows; GitHub Actions/Azure DevOps exposure; documentation and stakeholder communication; Jira/Agile

Analytics & Visualization: Power BI, Tableau, Plotly/Matplotlib

EDUCATION

PhD, Cognitive Science
Massachusetts Institute of Technology

BS in Electrical Engineering & BS in Biology
The University of Texas at Austin

CERTIFICATIONS

Microsoft Certified: Azure AI Engineer Associate (2024)

Certified Data Management Professional Associate (2024)

AWS Certified Cloud Practitioner (2023)

DCAM Accredited V2 (2022)