

# BARBARA HIDALGO-SOTELO, PhD

AI Engineer & Consultant | AI/ML • Data Governance • Business–Technology Bridge

📞 +1-512-804-9995 📧 barbs@balex.com 🌐 barbhs.com 🌐 dagny099.github.io 🌈 Austin, TX, USA 🌟 US Citizen

## SUMMARY

AI Engineer with hands-on experience building and deploying LLM-powered applications and production ML systems. Strong Python programming foundation combined with expertise in generative AI frameworks, vector databases, and graph-based retrieval. Recent focus on RAG architectures, multi-modal AI integration, and scalable inference pipelines. PhD in Cognitive Science (MIT) with 6+ years applying ML/AI to real-world problems across healthcare, federal, and enterprise domains. Proven ability to rapidly learn emerging AI technologies and translate research into production systems.

## TECHNICAL EXPERIENCE

### Metric5 Consulting – Data Scientist

2021 - 2023 Washington, DC | Remote

*Federal client: Financial markets & data governance initiatives*

- Designed Python utilities enabling superusers to extract/validate metadata, supporting compliance with OPEN Government Data Act.
- Built dashboards in Power BI to visualize workflows and user behavior, helping teams identify and resolve bottlenecks.
- Partnered with business stakeholders and developers to align federal data strategy mandates with technical delivery.

### Inflective, Inc. – Technical Consultant & Project Manager

2017 - 2019 NJ, PA, FL | Remote

*Client: Third-party administrator of healthcare claims*

- Directed a digital transformation project optimizing a pipeline for 4M+ annual medical bills, improving throughput by 30% while maintaining accuracy.
- Conducted workshops with executives and developers to map business/data flows and align technology with business goals.
- Applied Lean Six Sigma to streamline compliance processes, reduce errors.

### MediaScience – Research Project Manager

2017 Austin, TX

- Led analysis of biometric and eye-tracking data to assess advertising impact, applying anomaly detection and statistical modeling.
- Delivered customized statistical reports and dashboards to non-technical stakeholders, influencing client strategy.
- Coordinated multi-disciplinary teams and managed client-facing projects for major media organizations.

### University of Texas at Austin – Research Engineer / Scientist

2013 - 2016 Austin, TX

*Longitudinal study on bilingual development and cognition*

- Built a secure data pipeline to collect and analyze speech and behavioral data from over 1,000 participants
- Implemented a GUI-based tool using Python and MATLAB to extract and validate speech onset times from raw audio in behavioral research settings
- Mentored research staff & ensured protocol compliance across teams

## EDUCATION

### PhD, Cognitive Science

#### Massachusetts Institute of Technology

- *Dissertation: Applied computer vision and statistical modeling to analyze human visual attention across 900+ scenarios. Research cited 417+ times.*

### BS, Electrical Engineering | BS, Biology

#### University of Texas at Austin

## CERTIFICATIONS

### Microsoft: Azure AI Engineer Associate (2024)

### Certified Data Management Associate (2024)

### AWS Certified Cloud Practitioner (2023)

### DCAM v2 Accredited (2022)

## TECHNICAL SKILLS

### AI Engineering & LLM Development

- **GenAI Frameworks:** LangChain, LlamaIndex, Hugging Face Transformers, prompt engineering, RAG architectures
- **LLM Platforms:** OpenAI (GPT-4, embeddings), Anthropic Claude, Azure OpenAI, AWS Bedrock, Google Gemini
- **Retrieval & Search:** Vector databases (ChromaDB, Pinecone, Weaviate), semantic search, embedding models
- **Graph Databases:** Neo4j, Cypher query language, knowledge graph construction

### Programming & Software Engineering

- **Core Languages:** Python (expert), SQL, JavaScript, Bash scripting
- **ML Libraries:** PyTorch, TensorFlow, scikit-learn, NumPy, XGBoost
- **Development:** Git, Docker, API design, microservices architecture, CI/CD pipelines

### Cloud & MLOps

- **Cloud Platforms:** AWS (Lambda, SageMaker, Bedrock, EC2, S3, RDS), Azure (AI Studio, ML Studio), GCP (BigQuery, Vision API)
- **MLOps:** Model deployment, versioning, monitoring, evaluation pipelines, A/B testing frameworks
- **Data Engineering:** ETL pipelines, data quality assessment, automated data collection

### Application Development

- **Frameworks:** Streamlit, Flask, FastAPI
- **Databases:** PostgreSQL, MySQL, MongoDB, Neo4j
- **Visualization:** Plotly, Matplotlib, interactive dashboards

## PROJECTS & AI APPLICATIONS

### 💬 ConvoScope – Production LLM Application Platform

*Python, OpenAI API, Anthropic, LangChain, AWS / 2024*

- Built production-ready GenAI application integrating multiple LLM providers (OpenAI, Anthropic) with real-time streaming and conversation management
- Implemented context management patterns for multi-turn conversations using LangChain orchestration
- Deployed on AWS EC2 with SSL security, custom domain & automated monitoring

### 🐝 Beehive Analytics – Graph & Computer Vision Platform

*Python, Neo4j, Google Cloud Vision API / 2025–Present*

- Built knowledge graph (Neo4j) integrating 4+ years of images with temporal, environment, and observation metadata
- Applied computer vision APIs and custom image processing pipelines to extract structured knowledge from unstructured data

### 🏃 Workout Analytics – Time-Series ML Classification System

*Python, scikit-learn, AWS RDS, Streamlit / 2023–Present*

- Built end-to-end ML pipeline: automated CSV export → AWS Lambda processing → model training → inference
- Created interactive dashboard (workouts.barbhs.com) with explainable AI features and prediction confidence scores
- Deployed on AWS backend and secure credential mgmt