

Mridul Sarkar

Applied AI Engineer · Forward Deployed Engineer

Chicago, IL · (630) 640-6813 · mridulsarkarbiz@gmail.com

github.com/mertall · linkedin.com/in/mridul-s

SUMMARY

Full-stack AI engineer, **5+ years shipping applied ML and LLM systems from discovery to production** across federal, enterprise, and SMB. Owns the full arc — agentic AI, RAG, and large-scale data platforms — across the application layer (agents, orchestration, evals), the model layer (fine-tuning, embeddings, retrieval), and the systems layer (cloud, microservices, on-device inference).

EXPERIENCE

Founder & Principal AI Engineer — Multimodal Works LLC Jul 2025 – Present

AI consulting for SMBs — local AI, cloud AI, and custom full-stack apps, discovery to production-hardened handoff · Greater Chicago Area

- Built **Live Job Costing** (React · Node/Python · GraphQL/FastAPI), tying every cost to a job the moment it enters **QuickBooks** so margin is visible mid-build; a \$5M firm saves **~\$30K/yr** in bookkeeping and CPA fees.
- Built an **Operational Intelligence Agent** over client CRM + financial data with a first-class **memory graph**, so it survives long-running, multi-step workflows instead of resetting like a stateless chatbot.
- Design and custom-build **\$3K on-site AI workstations (128 GB unified memory)** running open-source pipelines fully on client hardware (no data egress), plus a cross-platform installer that right-sizes the model tier to each machine.

Senior Software Engineer · Associate Manager (CL8) — Accenture Federal Services Sep 2021 – Jul 2025

Custom Software Engineering / MLE Group, Remote · promoted from Software Engineering Specialist (CL9)

- Co-architected and scaled an **enterprise data platform for 100k+ users — 99.9% SLOs, p95 latency -40%**.
- Led **GenAI prototyping (RAG + vector databases)**, improving **Recall@10 ~20%**; co-authored an **LLM deployment white paper for the Intelligence Community**.
- Drove **LLM adoption** as the bridge across legal, security, and engineering — turning a “no policy” stance into a deployed path to production.

- Engineered an **ACID-compliant federated search platform** (JDBC→MongoDB ingest, NLP enrichment, Elasticsearch) and distributed microservices serving **2M+ users**.

AI Engineer (Contract) — Ravata Solutions · CV for embryo-health analysis (reproductive-health startup) Jul 2021 – Sep 2021

Associate Instructor, Quantum Computing — MIT Lincoln Laboratory · applied quantum-computing summer course Jul 2021 – Aug 2021

Edge AI & Software Engineering — Space and Satellite Systems · edge AI on LEO-satellite imagery; PID control Sep 2018 – Aug 2021

Transformer Representation Learning — VMware · BERT-based QA over cloud documentation (NLP) Oct 2018 – Mar 2019

SELECTED PROJECTS

- **Spectrogram K-means Classifier** — 95% blind-test accuracy with robust preprocessing and evaluation. github.com/mertall/audio-classification
- **Multimodal Image–Text Vector Search** (CLIP · HNSW · FastAPI · SageMaker) — chose embedding retrieval over generative RAG for latency and determinism. github.com/mertall/VisionSearch
- **Quantum Computing Hackathon** — 8th of 600 teams; optimized 60-qubit tensor-network contraction under hardware constraints. github.com/mertall/bluequbit_hackathon

SKILLS

AI / ML: LLM agents & orchestration, RAG, vector search (HNSW), evals, fine-tuning, CLIP/embeddings, PyTorch, scikit-learn, HuggingFace, LlamaIndex, DSPy, LangGraph, Triton

Cloud & Backend: AWS (Bedrock, SageMaker, Lambda, EC2, S3, DynamoDB, CloudFormation, IAM), FastAPI, Node.js, GraphQL, Flask, React, MongoDB, OpenSearch/Elasticsearch, QuickBooks API

Languages & Practices: Python, TypeScript/JS, C/C++, SQL, Bash · Docker, CI/CD, IaC, TDD, Git

EDUCATION & CERTIFICATIONS

University of California, Davis — B.A.S., Computational and Scientific Mathematics

HuggingFace AI Agents (2025) · Databricks GenAI Engineer (2024) · CU-Boulder Advanced Algorithms (2024) · IBM Variational Algorithms (2024) · Stanford Machine Learning (2019)