

Arun Joshi

Austin, TX | Open to NYC | joshiarun.com | github.com/arunjoshi9841 | linkedin.com/in/joshi-arun | arun_joshi1@outlook.com

SUMMARY

AI-focused software engineer with 6+ years of experience building production systems and applied AI products across RAG, agents, retrieval, evals, and applied ML. Strong across Python/FastAPI, TypeScript, AI SDK, vector search, cloud infrastructure, and end-to-end product delivery.

SKILLS

AI Systems: RAG, Agents, Tool Calling, Context Engineering, Hybrid Retrieval, Embeddings, Reranking, Evals, LLM-as-Judge

AI/ML: Python, FastAPI, PyTorch, Transformers, LoRA/Fine-Tuning, Qdrant, pgvector, OpenAI/Gemini APIs, AI SDK

Engineering: TypeScript, Next.js, React, .NET/C#, PostgreSQL/SQL, REST APIs, Redis/Celery, Docker, Azure, AWS, CI/CD

EXPERIENCE

Applied Dynamic Solutions | Software Engineer / Senior Software Engineer

May 2020 - Present | Remote

- Built a configurable AI support-agent platform adopted by 5 clients using RAG, product APIs, tool calling, grounded responses, live-agent handoff, and embeddable deployment.
- Built a permission-aware multimodal RAG workspace across documents, websites, audio, and video using Qdrant, metadata-aware retrieval, chunking, streaming, and background processing.
- Designed AI eval workflows and a testing playground covering retrieval quality, groundedness, citation quality, tool selection, response correctness, regression testing, and LLM-as-judge grading.
- Designed agent workflows and versioned configurations for prompts, tools, deployments, rollback, search/research orchestration, and production releases using AI SDK and model APIs.
- Owned architecture and delivery across frontend, APIs, backend, cloud, DevOps, and UX; established reusable patterns and CI/CD used across 20+ projects.

TheDrive.ai | Co-Founder, Product & Engineering

Mar 2023 - 2025 | Remote

- Co-founded an AI file workspace for natural-language search, Q&A, organization, and file actions, growing to 1,000 monthly active users.
- Built a TipTap-based collaborative file editor with multi-user editing, socket-driven synchronization, autosave, document state handling, and real-time updates.
- Designed reindexing pipelines and event-triggered synchronization so edited content remained searchable and available to downstream AI retrieval workflows.
- Shaped product architecture across grounded answers, source-aware retrieval, agentic file operations, sharing, permissions, and AI-assisted content generation.

Agile Academy | Software Engineer, Project-Based

2019 - 2021

- Built live and self-paced video learning, discussion workflows, and CDN-backed media delivery with caching for responsive course experiences.
- Implemented queue-driven workers, background jobs, and microservice-style processing for asynchronous and long-running workloads.

SELECTED PROJECTS

Lattice | Production RAG & Agent Platform

- Built a multi-tenant AI knowledge platform with FastAPI, PostgreSQL/pgvector, hybrid lexical-vector retrieval, rank fusion, reranking, context expansion, citations, and organization-scoped authorization.
- Designed asynchronous ingestion/reindexing with Celery/Redis plus a bounded agent runtime with tool controls, timeouts, retries, call budgets, duplicate-call handling, and SSE streaming.
- Built retrieval and response evaluation suites covering Precision@K/Recall@K, groundedness, citation quality, agent/tool behavior, and LLM-based graders.

Emergency Dispatch Transformer Classification | UT Austin

- Fine-tuned DeBERTa-v3-base for multitask 911 transcript classification on a 6,030-record real/synthetic dataset, achieving 0.8680 dispatch positive-label macro F1.
- Compared model variants using macro/micro F1, validation loss, confusion matrices, and error analysis to diagnose class imbalance and overfitting.

Local LLM Clinical De-identification | UT Austin

- Fine-tuned Gemma with LoRA for local PHI removal and evaluated privacy/clinical-fidelity tradeoffs across rules, NER, LLMs, and human review.

EDUCATION

The University of Texas at Austin | M.S. Artificial Intelligence | Expected Dec 2026

Caldwell University | B.S. Computer Information Systems | Minor in Mathematics and Business Administration | 2020