

Akshay Prabhakant

akshayprabhakant@gmail.com | +91-9867573384 | linkedIn/akshay-prabhakant-438816190 | github/akshayDev17

EDUCATION

B.Tech in Computer Science, MS by Research in Computational Sciences India | August 2016 - July 2021
IIIT HYDERABAD

WORK EXPERIENCE

RUNWHEN | FOUNDING PRODUCT ENGINEER

Pune, India | July 2025 - Present

- Built a **production RAG retrieval service** over a **Neo4j vector store** with OpenAI **text-embedding-3-small** embeddings, **hybrid dense + BM25 retrieval fused via Reciprocal Rank Fusion**, and sentence-level chunking.
- Architected a **conversational agent on Google ADK** exposing the workspace as a **virtual filesystem (VFS) tool interface** (`ls/cat/grep/search` over a Neo4j-backed graph), replacing a prior 8-sub-agent / 74-tool design with a single tool-calling agent; with session-state memory and SSE streaming.
- Built **multi-provider LLM orchestration** via a **LiteLLM proxy** spanning **Gemini, GPT-4.1, and direct OpenAI**, with workspace-scoped model routing, Vault-backed credentials, and per-turn token accounting.
- Built **AI guardrails** - a **fail-closed PII/secret scrubber** (Presidio, memoized, streaming-safe) and a custom **ADK agent-eval harness** with rubric-based LLM-judge scoring and a user simulator for multi-turn regression checks.
- Cut LLM spend via an **incident novelty & summarization pipeline**: stacktrace hashing, issue-presence gating, and **severity-gating**, backed by 7-day **Redis** caching of completions; built interactive **SLI/SLO error-budget charts** (ECharts), correcting PromQL for accurate burn-rate math.

XPO LOGISTICS | DATA SCIENTIST - 2

Pune, India | July 2022 – March 2025

- Developed a **Gemini-flash** based **RAG-Chatbot** off of XPO's external help center to help potential or existing customers know all services offered and **enhance user experience**.
- **Collaborated** closely with the **Pricing team** to develop a **Gemini-flash** based financial **information retrieval** process from pricing-agreement *PDFs having diverse layouts* with an overall **accuracy of 88%**.
- Built an **end-to-end ML pipeline** (random forest regressors) to predict shipment **ETA**, reaching **90% accuracy** vs. **84%** prior; validated via a **confounder-controlled UAT** and **two-proportion z-test A/B test**, lifting on-time shipments from **83–87% to 88–92%**.
 - Derived per-SIC, per-hour P25/P50/P75 distributions of shipment volume and dwell time as congestion features, tuning outlier thresholds via MSE/MedAE analysis.
 - Constructed a **matched 12-SIC UAT cohort** across 4 states on hourly throughput, trailers, and workforce – avoiding **within-site interference effects**.
- **Mentored** incoming data scientists on **business knowledge** and supported them through **ideation** and **implementation** of their projects.
- Built a **Google BigQuery + DataStudio** dashboard to **quantify Price-Elasticity** via a **custom willingness-to-pay index**, **optimizing discount strategies** to drive a **9% shipment-volume increase** and **7% revenue boost**.

XPO LOGISTICS | DATA SCIENTIST

Pune, India | July 2021 – June 2022

- **Delivered qualitative and quantitative insights using Bigquery** on the **potential causes of Damaged Shipments** such as shipment weight, no. of human handlings, travel distance, NMFC class codes, Full-time and Part-time employees, and in-house vs 3rd party transportation, resulting in a **19% reduction in damaged shipments**.

SKILLS

Languages: Python, SQL

GenAI / Agentic AI: multi-agent systems, tool-calling, Google ADK, LiteLLM, guardrails (Presidio), multi-provider LLMs

Retrieval / Vector Search: RAG, vector DBs (Neo4j, Qdrant), embeddings, hybrid search (dense + BM25, RRF), chunking

ML / Data: scikit-learn, TensorFlow, Keras, A/B testing, PostgreSQL, Redis (Sentinel), Pandas, NumPy

Platform / MLOps: FastAPI, Kubernetes (GKE), Docker, Flux CD (GitOps), Helm, CI/CD (GitHub Actions), Vault, GCP