

# RAJ CHHAPARIYA

AI / Data Engineer · Data & LLM Systems · Machine Learning

Bengaluru, India · [rajchhapariya8@gmail.com](mailto:rajchhapariya8@gmail.com) · [Portfolio](#) · [GitHub](#) · [X](#) · [LinkedIn](#)

## PROFESSIONAL SUMMARY

AI / Data Engineer with hands-on experience building and evaluating reliable data pipelines, in-process analytical engines, and retrieval-augmented LLM systems. Proficient in Python, SQL, and TypeScript, with practical experience designing deterministic query guardrails, automated data ingestion scrapers, and predictive machine learning models. Supported by production full-stack engineering capabilities across PostgreSQL, DuckDB, MongoDB, and Next.js.

## TECHNICAL SKILLS

|                                   |                                                                                                                    |
|-----------------------------------|--------------------------------------------------------------------------------------------------------------------|
| <b>Data &amp; ML Core:</b>        | Python, SQL, DuckDB, Pandas, NumPy, Scikit-learn, XGBoost, Random Forest, SMOTE, Statistical Modeling              |
| <b>LLM Systems &amp; Agents:</b>  | Retrieval-Augmented Generation (RAG), OpenAI API, BM25, Multi-Agent Systems, RRF, Pydantic, Evaluation             |
| <b>Databases &amp; Pipelines:</b> | PostgreSQL, Supabase (RLS), MongoDB, DuckDB, Redis, Drizzle ORM, Star Schema, ETL / Ingestion Pipelines, REST APIs |
| <b>Full-Stack &amp; DevOps:</b>   | TypeScript, JavaScript, Next.js, React, Node.js, Streamlit, Docker, GitHub Actions, Power BI, Git                  |

## FEATURED AI & DATA SYSTEMS PROJECTS

**Agentic RAG Knowledge Assistant** [Python, OpenAI API, BM25, NumPy, Pydantic, RRF] [GitHub](#) Aug 2026 – Present

- Built a retrieval-augmented generation pipeline that plans queries, retrieves multi-source documents, checks evidence quality, and generates answers strictly grounded in retrieved evidence, with weak evidence triggering defensive retries or refusal.
- Combined dense semantic embeddings with BM25 keyword search using **Reciprocal Rank Fusion (RRF)** to optimize retrieval precision, adding contradiction detection and multi-part query decomposition.
- Built an automated evaluation framework measuring retrieval accuracy, citation quality, hallucination rate, and token cost: **87 automated test cases** pass deterministically without live API calls.
- Enforced strict Pydantic output schemas and token-budget context compression to prevent latency degradation.

**AI Data Analyst Agent** [Python, LLMs, DuckDB, Streamlit] [GitHub](#) Aug 2026 – Present

- Built an agent that queries datasets using read-only SQL via DuckDB in-process columnar engine, eliminating arbitrary code execution risks.
- Designed deterministic numerical verification checks validating model claims against ground-truth analytical outputs to prevent hallucinations.
- Evaluated systematically across **20 questions and edge cases**, verifying 100% tool selection accuracy and documenting boundary limitations.
- Created an interactive Streamlit UI supporting automated data profiling, query explain plans, dynamic charts, and automated report exports.

**JanDarshan — Civic Intelligence Platform** [Next.js, TypeScript, MongoDB, Puppeteer] [Live](#) · [GitHub](#) Apr 2026 – Jul 2026

- Architected an open-source civic intelligence platform aggregating, normalizing, and indexing Indian parliamentary dossiers (Lok Sabha & Rajya Sabha MPs, ministers) from official government portals.
- Engineered headless Puppeteer scrapers with exponential retry logic, upserting normalized document schemas into MongoDB via Mongoose with automated schema validation.
- Built interactive constituent search with fuzzy filtering, member attendance metrics, question hour tracking, and responsive UI in Next.js & TypeScript.

## PRODUCTION ENGINEERING & SAAS PLATFORMS

**Founder & Full-Stack Engineer** — Truly ([trulyapp.online](https://trulyapp.online)) [Live](#) Apr 2026 – Present

- Architected social video networking platform with structured 60-min circles, double-blind matching algorithms, Panda CSS, and Neon PostgreSQL with Drizzle ORM.
- Implemented pessimistic row-locking (`SELECT ... FOR UPDATE`) eliminating seat overbooking under concurrent booking bursts, UUIDv7 monotonic primary keys, and Transactional Outbox event processing.

**Founder & Developer** — Resume Roaster ([atsroast.com](https://atsroast.com)) [Live](#) Jul 2026 – Present

- Built and launched AI resume evaluation tool with Next.js, TypeScript, Supabase (PostgreSQL with RLS), and OpenAI API, featuring spatial ATS parsing, action verb rewriters, and ATS-safe PDF generation.
- Designed deterministic scoring rubric evaluating semantic density, quantifying achievements, and identifying syntax parsing bottlenecks.

**Full-Stack Engineer** — Aarohana Trails (Client Project) [Live](#) Aug 2026 – Sep 2026

- Engineered end-to-end trekking discovery and reservation platform for a commercial adventure tour operator with Next.js, custom CMS, and Prisma ORM.
- Designed normalized schema for slot inventory, automated booking confirmations, and customer enquiry workflows.

## DATA WAREHOUSING, ANALYTICS & MACHINE LEARNING MVPS

### Sales Analytics Data Warehouse & Power BI Dashboard [SQL, Star Schema DW, Power BI DAX] [GitHub](#)

Oct 2025

- Designed an enterprise hardware retail data warehouse across **150,000+ transaction records** using Star Schema modeling with optimized Fact and Dimension tables.
- Authored complex SQL views for currency normalization, profit margin calculation, and sales trend aggregations across multi-regional distribution channels.
- Developed an executive Power BI dashboard with dynamic DAX measures to surface sales performance bottlenecks and regional revenue growth.

### Customer Churn Prediction Engine [Python, Scikit-learn, Pandas, SMOTE, Random Forest] [GitHub](#)

Oct 2025

- Developed an end-to-end classification pipeline predicting subscriber churn in telecom customer datasets with severe class imbalance.
- Applied **SMOTE (Synthetic Minority Over-sampling Technique)** combined with feature engineering (tenure cohorts, monthly spend velocity, contract risk).
- Trained Random Forest and Gradient Boosted models, achieving high sensitivity on churners with SHAP-based feature importance interpretation.

### AvgEGT Industrial Predictive Maintenance System [Python, XGBoost, Scikit-learn, Matplotlib] [GitHub](#)

Jun 2026 – Jul 2026

- Built a predictive regression model forecasting gas turbine Average Exhaust Gas Temperature (AvgEGT) to detect thermal mechanical degradation early.
- Cleaned and normalized multi-sensor time-series telemetry data (compressor pressure, ambient temperature, fuel flow), eliminating sensor drift and collinearity.
- Trained regularized **XGBoost regression** models to forecast thermal anomalies, establishing early warning thresholds to mitigate unscheduled maintenance.

## SUPPORTING PRODUCTION SYSTEMS

- **SubsSync:** B2B SaaS subscription billing platform with Razorpay UPI AutoPay integration and webhook processing. [Live](#)
- **Namma Metro Guide:** Android transit navigation app with offline routing, transfer algorithms, and fare calculation.

- **VyaparStack:** Statutory GST compliance & invoicing SaaS with CGST/SGST/IGST tax engine, Neon PostgreSQL, and pdf-lib. [Live](#)

## EDUCATION

### Bachelor of Technology (B.Tech) in Information Technology

Dr. A.P.J. Abdul Kalam Technical University  
2020 – 2024

## CERTIFICATIONS

- Google Data Analytics Professional Certificate (May 2023)
- Cloud Computing Certification, Honeywell (Grade A, Oct 2023)