RAJ CHHAPARIYA

AI / Data Engineer · Data & LLM Systems

Bengaluru, India · rajchhapariya8@gmail.com · Portfolio ↗ · GitHub ↗ · X ↗ · LinkedIn ↗

PROFESSIONAL SUMMARY

AI / Data Engineer building and evaluating data pipelines, RAG systems and reproducible benchmarks in Python, SQL and TypeScript, with full-stack work across PostgreSQL, DuckDB, MongoDB and Next.js.

TECHNICAL SKILLS

Data & ML:	Python, SQL, DuckDB, Pandas, NumPy, Scikit-learn, PyTorch, Parquet
LLM Systems:	Retrieval-Augmented Generation (RAG), OpenAI API, BM25, RRF, Pydantic, Evaluation Design
Pipelines & Infra:	Kafka API (Redpanda), Terraform static analysis, PostgreSQL, Supabase (RLS), MongoDB, pytest
Full-Stack:	TypeScript, Next.js, React, Node.js, Streamlit, Git

AI & DATA PROJECTS

Agentic RAG Knowledge Assistant `[Python, OpenAI API, BM25, NumPy, Pydantic]` GitHub ↗ — **Aug 2026 – Present**

- Built a RAG pipeline that plans queries, retrieves, audits evidence sufficiency and answers with citations; weak evidence triggers retries or refusal. Dense and BM25 retrieval are fused with **Reciprocal Rank Fusion (RRF)**. 90 test functions, 87 run offline by default.
- Results on the **32-question held-out test set** (24 answerable, 4 unanswerable and 4 false-premise questions; accuracy is calculated over the 24 answerable questions; **frozen pre-fix** results; small sample):
- **Full agentic system:** answer accuracy 11.0/24 (45.8%); refused 4/4 unanswerable questions; 3/24 false refusals; average latency 17.69 s.
- **Naive dense baseline:** answer accuracy 14.0/24 (58.3%); refused 0/4 unanswerable; 0/24 false refusals; average latency 1.87 s.
- **Hybrid baseline:** answer accuracy 14.5/24 (60.4%); refused 0/4 unanswerable; 0/24 false refusals; average latency 1.86 s.

AI Data Analyst Agent `[Python, DuckDB, Streamlit]` GitHub ↗ — **Aug 2026 – Present**

- Built an agent that answers plain-English dataset questions through four fixed tools (query, chart, statistics, clarify), with a Streamlit UI for CSV upload, profiling, charts and report export. SQL is limited to single SELECT/WITH statements by regex keyword guardrails on in-memory DuckDB; a deterministic number-matching check compares each answer's figures with the tool output.
- Selected the correct tool in all 20 questions in a 20-question benchmark; answer correctness was 17/20. Single small benchmark.
- **Taxi Stream Benchmark** `[Python, Redpanda, Parquet, DuckDB]` GitHub ↗ — Compared at-least-once with effectively-once delivery (not Kafka exactly-once) on 2.87M NYC taxi rows: under injected SIGKILLs, 0 duplicate and 0 missing rows vs a mean of 93–3,531 duplicates. Single-machine benchmark (3 seeds), not a production guarantee.
- **EdgeFed-Bench** `[Python, PyTorch, DuckDB]` GitHub ↗ — INT8 gradient quantization cut per-client uplink by 75% with unchanged Multi-Krum accuracy under 20% sign-flip (84.90% vs 84.89%) on Fashion-MNIST (20 clients, 50 rounds, 5 seeds). Single host, small MLP.
- **tf-misconfig-scanner** `[Python, python-hcl2, SARIF]` GitHub ↗ — Static Terraform scanner for AWS (12 rules). 33-item independent TerraGoat evaluation: 100% precision and recall vs Checkov's 100% and 66.7% (small sample; TerraGoat is from Checkov's vendor). 73 scored items from locally authored fixtures, written alongside the scanner, are scored separately.

APPLIED WEB SYSTEMS

Founder & Independent Developer — Resume Roaster `[Next.js, TypeScript, Supabase, Razorpay]` Live ↗ — **Jul 2026 – Present**

- Independently designed, developed and deployed a full-stack AI-powered resume platform covering frontend, backend, database architecture, AI integration, payments and PDF/DOCX generation. Features are verified in the codebase; availability of individual features to users is not claimed.

OTHER PROJECTS

- **JanDarshan:** civic information platform for Indian legislators (Next.js, MongoDB, Three.js). GitHub ↗
- **Truly:** social video platform backend (Next.js, Neon PostgreSQL, Drizzle ORM); atomic conditional-update seat reservation.
- **MandateStack:** subscription billing with Razorpay UPI AutoPay and webhooks.
- **VyaparStack:** GST invoicing with a CGST/SGST/IGST tax engine and pdf-lib. Live ↗

EXPERIENCE

Restaurant Owner & Operator — Varanasi, India — **Jan 2025 – Jul 2025**

- Owned and operated an independent restaurant for six months: staff operations, budgeting, menu pricing and suppliers.

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)