Muhammed Rasin O M

AI and Data Engineer · Python, LLMs, data pipelines

rasinbinabdulla@gmail.com · Kozhikode, Kerala, India (open to relocation)
rasinmuhammed.tech · linkedin.com/in/rasinmuhammed · github.com/rasinmuhammed

SUMMARY

AI and data engineer with 1.5 years of industry experience building LLM backends and data pipelines in Python: remote backend work for DearPriceHunter, a Swedish startup, and data engineering at JMAN Group. Hands-on with document ingestion, RAG, LLM evaluation and REST APIs. Founder of Misata, an open-source synthetic data tool with 50+ users and one pilot running.

SKILLS

AI and LLMs	Large language models (LLMs), RAG, LangGraph, LangChain, OpenAI, Claude and AWS Bedrock APIs, Ollama, Langfuse, Prompt caching, MCP
Data	Python, SQL, pandas, NumPy, PySpark, ETL pipelines, Databricks, Snowflake, dbt, Azure Data Factory, DuckDB, PostgreSQL, Splink, Power BI
Machine learning	scikit-learn, LightGBM, PyTorch, NLP, Model evaluation, Statistics and A/B testing
Engineering	FastAPI, REST APIs, SQLAlchemy, pytest, Git, Docker, GitHub Actions (CI/CD), TypeScript

EXPERIENCE

DearPriceHunter AB · Software developer, contract · Sweden (remote) Dec 2025 – Aug 2026

- Built a document ingestion pipeline as FastAPI services that turns uploaded CSV, Excel, PDF and JSON files into structured spending data.
- Built a detection engine that flags unusual spending, with LLM-written reports in English and Swedish. Prompt caching cut inference cost by about 90%, and Langfuse tracing showed what the models were doing.
- Ran models on AWS Bedrock in an EU region so client data stayed within data-residency requirements.

Independent · Backend and AI developer · Remote Mar 2025 – Present

- Onepane AI: built a report generator with LangGraph, FastAPI and the OpenAI API that summarises cloud billing data.
- Peer reviewer for Springer Nature's Network Modeling Analysis in Health Informatics and Bioinformatics.

JMAN Group · Software engineer, data engineering · Chennai, India Jul 2024 – Feb 2025

- Built a probabilistic record-linkage model with Splink and PySpark on Databricks for an entity-resolution product, tuning blocking rules and match thresholds for precision and recall.
- Generated synthetic datasets so consultants could test ideas without touching production data.
- Built an LLM lead classifier and Power BI dashboards; trained in Snowflake, dbt and Azure Data Factory.

PROJECTS

Misata · Synthetic data product, founder · Python, NumPy, DuckDB, MCP, pytest misata.studio

- Describe a business in plain English and get connected tables where the numbers you state (revenue curves, churn, fraud rates) come out exactly, every key between tables is valid, and each dataset ships with a certificate proving it.
- Open source with 70+ GitHub stars, an MCP server for AI coding agents and an arXiv preprint (2606.08736).
- Misata Studio has 50+ users, free and paid tiers, and one pilot running. Selected for Claude for Startups.

Ratio · Legal search · Python, BM25, Embeddings, Reranking ratio-legal-rag.vercel.app

- Answers legal questions with citations across 10,588 Indian court judgments; finds the exact case asked for 89% of the time on 313 hand-labelled questions, up from 59% with keyword search.

Entify · Duplicate detection · Splink, DuckDB, TypeScript github.com/rasinmuhammed/entify

- Finds duplicate customer records without specialist tuning; on the FEBRL benchmark every pair it flagged was a real duplicate, and it found 98.7% of them.

Harara · Heat safety · Python, Weather data, LLM evaluation harara-one.vercel.app

- Turns heat forecasts into hourly work and rest plans under Qatar's heat-stress rules; found that ECMWF misses 12 to 18% of stop-work hours where NOAA GFS misses 40 to 45%.

Open-source libraries · indicfuzz, rag-tui, opsdeck, pytest-misata, mangograd, node-canon.

EDUCATION

Mar Athanasius College of Engineering, Kerala · B.Tech in CSE (Data Science), CGPA 8.52 2020 – 2024