Alpit Mali

Senior Data Engineer | AI Engineer (Python, PySpark, Databricks, LLM, RAG)
Surat, India | Open to relocation to the Netherlands, Germany and the wider EU | www.alpitmali.com
+91 7284067464 | alpitmali3434@gmail.com | github.com/malialpit | linkedin.com/in/alpitmali

PROFESSIONAL SUMMARY

Senior Data Engineer and AI Engineer with 5+ years of production experience across data pipelines, backend services and LLM applications. At Persistent Systems, build PySpark and Azure Databricks pipelines processing 50M+ retail records for NielsenIQ's store analytics platform (RTM-SES), and built the OCR text normalisation service and data pipelines (FastAPI, PostgreSQL, AKS) behind NielsenIQ's receipt-scanning rewards app in Germany and France. Earlier architected the backend of NueGO, India's first intercity electric bus platform (100K+ monthly active users). Build production LLM systems end to end: a WhatsApp AI agent over ERP data with closed-set intent routing and hallucination-safe templated answers, multi-tenant RAG and multilingual chatbots (LangChain, Qdrant, OpenAI and Anthropic APIs).

TECHNICAL SKILLS

- **Languages:** Python, SQL, TypeScript, JavaScript, C#
- **Data Engineering:** PySpark, Apache Spark, Azure Databricks, Apache Airflow, Snowflake, Hive, Pandas, Parquet, ETL, ELT, batch and incremental pipelines, job scheduling, data modelling, data warehousing, data quality
- **AI and LLM:** LangChain, OpenAI API, Anthropic Claude API, RAG, Qdrant, vector databases, AI agents, agentic workflows, LLM orchestration, intent routing, decision models (TypeSafe AI Jev), confidence gating, LLM evaluation, multilingual NLP, OCR (Tesseract, OpenCV), fuzzy text matching (Levenshtein, trigram, Soundex)
- **Backend:** FastAPI, Django, Django REST Framework, GraphQL, Flask, asyncpg, Pydantic, Celery, WebSocket, REST APIs, microservices
- **Databases:** PostgreSQL (pg_trgm, fuzzystrmatch, PostGIS), Snowflake, Redis, MongoDB, Neo4j, MySQL, MSSQL
- **Messaging and Streaming:** Apache Kafka, RabbitMQ, AWS SQS, AWS SNS, WhatsApp Business API
- **Cloud and DevOps:** Azure (Databricks, AKS, Blob Storage, DevOps, Service Bus), AWS (EC2, Lambda, S3, RDS, Athena), Docker, Kubernetes, Helm, GitHub Actions, GitLab CI/CD, Prometheus, ELK Stack
- **Ways of Working:** Agile, Scrum, Jira, Confluence, pytest, technical architecture, client-facing documentation

PROFESSIONAL EXPERIENCE

Senior Consultant, Persistent Systems (NielsenIQ, IQVIA) Mumbai, India
February 2025 to Present

- RTM-SES (NielsenIQ Store Execution Solution): Built PySpark pipelines on Azure Databricks processing 50M+ retail records per run, enabling CPG brands such as Bacardi, Coca-Cola, Unilever and Red Bull to benchmark in-store performance against competitors.
- Developed the RTM job services in FastAPI that launch and track Spark jobs for the Segmentation Builder and serve results to dashboards used by 200+ brand analysts daily.
- Implemented Spark logic for segmentation analyses including store ranking, store overlap, price bands, nested analysis, test and buyback, store matrix, store evolution, concentration curve and custom segmentation.
- Published reusable Python packages adopted by 3 engineering teams, cutting duplicate transformation code by 40%.
- CDAR Captivate (NielsenIQ receipt-scanning rewards app, Germany and France): Built a FastAPI and PostgreSQL service that normalises noisy OCR receipt text into clean master item descriptions.
- Designed Item Description Normalisation (IDN) as a staged fuzzy matching cascade (exact, character match, Soundex, Levenshtein, trigram) using PostgreSQL pg_trgm and fuzzystrmatch, returning the match stage and score for every item.
- Built Item Number Normalisation (INN), which corrects OCR-misread item codes using code frequency and median unit price from 13 weeks of live order history.
- Built the Captivate data pipelines and job services: Azure Blob parquet ingestion into PostgreSQL, a 52-week master reference build and weekly incremental updates, run as idempotent scheduled jobs with API triggers and run-level lineage.
- Deployed Captivate on Azure Kubernetes Service (AKS) with Docker, Helm and GitHub Actions CI/CD; added Prometheus metrics, structured JSON logging, a load-benchmark harness and 500+ pytest tests.
- IQVIA (healthcare): Created a data archival pipeline on Snowflake and PySpark for Sanofi EPA data, cutting processing time by 60% and SQL query calls by 45%.
- IQVIA: Replaced legacy Salesforce endpoints with a FastAPI microservice, reducing API latency by 40%, and added AWS SQS event-driven integration handling 10K+ messages per day.

Software Development Engineer, Neosoft Technologies Mumbai, India
February 2023 to November 2024

- Architected the backend of NueGO, India's first intercity electric bus platform (Django, GraphQL, PostgreSQL): 100K+

monthly active users, 200+ routes, 10M+ Play Store installs.
- Built the Automatic Fare Management System on Kafka event streams and Azure Blob Storage, processing 50K+ fare transactions per day in real time.
- Tuned PostGIS geospatial queries and added a Redis caching layer for live GPS tracking, cutting infrastructure costs by USD 3K per year.
- Integrated the WhatsApp Business API for automated ticket booking, increasing mobile conversions by 25% and reducing support tickets by 30%.
- Deployed the ELK stack for centralised monitoring across booking APIs, cutting mean incident response time by 45%; modelled journey master data in Neo4j for sub-second route queries across 500+ city pairs.
- Delivered EAlpha LMS microservices (Django REST Framework, AWS SNS, JWT) scaled to 15K+ students, and Webwerks data centre automation with one-click disaster recovery for 50+ endpoints via Nutanix APIs.

Software Engineer (Consultant), Techiemaya Hyderabad, India (remote)
November 2021 to December 2022

- Built data pipelines with Apache Airflow, PySpark and AWS Athena for P&G (Scotts Miracle), automating ingestion of 5M+ records and improving ETL throughput by 35%.
- Developed a FastAPI backend with a React and TypeScript frontend for an LMS, integrating the OpenAI API for assignment grading and reducing manual review time by 70%.
- Created Power BI dashboards with Python and SQL for the Morgan Stanley IMS platform across 3 data warehouses.
- Engineered a real-time chat system with Django Channels, WebSocket and Firebase Cloud Messaging supporting 5K+ concurrent users.

Python Developer, The Website Factory London Surat, India
July 2021 to November 2021

- Built Alwakeen, a global money transfer application (Django REST Framework, PostgreSQL) with role-based authentication and live exchange rate API integration.
- Delivered the Sudan Embassy UK government services portal (visa, passport, certificates) with Celery asynchronous processing and a custom SMTP service.

INDEPENDENT CONSULTING: SAMNITI TECHNOLOGIES

Lead Engineer and Technical Architect, Samniti Technologies (AI automation consultancy) Remote
2022 to Present

AI agents and automation for manufacturing clients in India, the EU and the GCC. Blog: www.alpitmali.com/blog

- **WhatsApp AI agent over factory ERP data:** Replaced manual pulling and reading of large reports from a local ERP with instant WhatsApp answers on sales, purchase orders, invoices and outstanding, asked in Hindi, Gujarati and English (Roman script). LLMs only make closed-set choices across 25 capabilities, and a confidence gate falls back to a menu instead of guessing.
- **Hallucination-safe answers:** The LLM never writes SQL or numbers and never sees ERP rows. Each capability maps to a reviewed, parameterised SQL query, party names are resolved by trigram search, and replies are rendered from templates using only query results.
- **Decision-model routing (TypeSafe AI Jev):** Evaluated the Jev decision model as the intent router against Claude Haiku and OpenAI GPT: typed closed-set output, calibrated confidence and routing decisions in milliseconds (70 to 500 ms), at the lowest cost per 1,000 decisions. Built a labelled multilingual test set to measure routing accuracy.
- **Multilingual WhatsApp commerce chatbot (JIJO Ventures):** Production RAG chatbot in 5 languages (Marathi, English, Gujarati, Hindi, Spanish) on LangChain, GPT-4o-mini, Claude and Qdrant. Customers browse, order and track deliveries in chat, with handling for code-mixed input and an admin dashboard to seed domain knowledge. Replaced manual order-taking for a full sales channel.
- **Multi-tenant RAG (SahuCosmetics):** Per-brand WhatsApp AI chatbots for 3 brands on one Django and Next.js platform with 5,000+ SKUs, using tenant-isolated Qdrant collections so each bot retrieves only its own catalogue. Added a React Native OCR business card scanner with GPT-4o-mini enrichment.

EDUCATION AND CERTIFICATIONS

- Software Engineering course, IIHT, 2018 to 2021
- Deep Learning certification, IIT Roorkee (NPTEL, MHRD)

WORK AUTHORISATION AND LANGUAGES

Indian national, age 26. Eligible on experience and salary, with no degree requirement, for the Netherlands Highly Skilled Migrant permit (under-30 threshold) and the Germany EU Blue Card route for IT specialists (3+ years of professional IT experience). Available within 30 days of visa approval. Experience letters available on request. Languages: English (full professional), Hindi, Gujarati.