

Methun Raj

AI & Data Engineer

Hamburg, Germany · +49 1577 3686620 · methunrajmani85@gmail.com

[linkedin.com/in/methun-raj-0335b617b](https://www.linkedin.com/in/methun-raj-0335b617b)

Permanent residence · Unrestricted work authorisation



PROFILE

Data engineer building **production LLM systems rather than prototypes**. Designed and still operate an AI extraction platform in daily use by ~200 people, routing across five model providers with per-project cost control. Four years at Statista — from forecasting the semiconductor, cloud and AI markets to engineering the ingestion and retrieval pipelines underneath — so I can hold both halves of the conversation. Current focus: agentic workflows, and how you evaluate what an agent produced before it ships.

PROFESSIONAL EXPERIENCE

Data Engineer

Jan 2026 – Present

Statista GmbH, Hamburg · Full-time

- Built the real-time news intelligence pipeline: GDELT ingest, two-stage deduplication and batched embedding into **sharded vector indexes**, running realtime and historical backfill off one code path with checkpointed, resumable state.
- Built a Python client library that made retrieval **self-service for other internal teams** — semantic search with date, country, language and source filters, streaming bulk export, and agentic retrieval with citations.
- Own pipeline reliability end to end: observability and alerting, so failures surface instead of passing silently.
- Turning the survey team's most time-consuming manual workflows into automated Python pipelines, with an internal operations app (React front end, Python back end) that runs them.

AI Consultant & AI Automation Engineer

Mar 2025 – Present

Swift ProSys · Freelance, Remote

- Sole engineer and client contact: scoped the problem, designed the system, shipped it, still maintain it in production.
- Platform used by **~200 people** across language-specific teams, processing **6,000,000,000 tokens** in five months.
- Route workloads across **five model providers** — OpenAI, Google Gemini, Qwen, Vector Engine and local Ollama — behind one interface, each billed correctly for reasoning, cached and thinking tokens.
- Template-driven extraction lets **non-technical operators add new document types** without engineering involvement; per-project cost analytics make spend predictable before a run.
- Built the reliability layer: a 12-category failure taxonomy and a four-attempt retry ladder that escalates prompt strictness and token budget. Provider credentials encrypted server-side with per-project isolation; 100-test suite (Vitest and Deno) with linting and dead-code analysis in CI.

Data Analyst

Apr 2025 – Jan 2026

Statista GmbH, Hamburg

- Owned market forecasts for four segments — Semiconductors, Public Cloud, AI and Data Center — published in Statista's **Technology Outlook** and sold to clients.
- Introduced new modelling methods with a real-time news indicator as the headline feature, so forecasts reflect what is actually happening in the news.
- Automated recurring analysis, improving turnaround across release cycles.

Junior Analyst

Dec 2023 – Mar 2025

Statista GmbH, Hamburg

- Published market size, revenue and growth KPIs across technology segments; accountable for accuracy from raw source to published figure.

Working Student, Data

Jul 2022 – Oct 2023

Statista GmbH, Hamburg

- First Python and SQL pipelines; web scraping and Common Crawl extraction; product matching across datasets with applied NLP (BERT).

Technical Intern

Jul 2020 – Jan 2021

Garuda Aerospace, Chennai, India

- Statistical analysis of drone telemetry; identified the drivers of battery performance.

EDUCATION

M.Sc. Smart Systems

Oct 2021 – Oct 2023

Hochschule Furtwangen University, Germany

- Thesis: a convolutional network identifying regional plant and insect species from photographs, served behind an API with continuous retraining.

B.E. Mechatronics Engineering

2017 – 2021

Agni College of Technology, Chennai, India · Grade 8.27 / 10

SKILLS

Programming	Python, SQL, TypeScript
AI & LLM	Large Language Models (LLM), Retrieval-Augmented Generation (RAG), vector search and vector databases, embeddings, AI agents and agentic workflows, model evaluation, LangChain, LangGraph, Agno, OpenAI and Gemini APIs, vLLM, Ollama, MCP
Data & Cloud	AWS (S3, S3 Vectors, Lambda), Docker, PostgreSQL, pgvector, Supabase, ETL and data pipelines, data modelling, observability, CI/CD
Delivery	Client scoping, solution design, production handover, stakeholder communication
Analysis	Market sizing and forecasting, KPI modelling, competitive benchmarking, pandas, NumPy
Tooling	Claude Code, OpenAI Codex, Cursor, GitHub Actions, Sentry
Languages	English — full professional proficiency · German — B1, working towards B2
Certifications	Python for Data Analysis, Coursera (2022) · Machine Learning with Python, Coursera (2022)