

OSAMA MOFTAH

Applied AI Engineer | RAG, LLM Evaluation, Agent Workflows

Berlin, Germany | +49 152 9609337 | omoftah@protonmail.com | moftah.dev | linkedin.com/in/omoftah

SUMMARY

Applied AI Engineer building practical LLM and data intelligence systems: RAG pipelines, NLP-assisted document processing, agent workflows, semantic drift detection, and validation tools. Experienced in turning fragmented public and business data into structured information using Python, machine learning, automation, and production-oriented system design. Focused on systems that are auditable, evidence-based, and useful in real operational workflows.

EXPERIENCE

AI Engineer

Oct 2025 - Now

Prospecto | Pistoia and Berlin

Developed a national-scale procurement intelligence platform combining public data pipelines, NLP analysis, and automated tender discovery workflows.

- Built distributed Python pipelines aggregating procurement data from 20,000+ government websites for tender discovery, monitoring, and analysis.
- Developed NLP and LLM-assisted document processing workflows for tender classification, structured extraction from public procurement records, and automated report generation.
- Designed entity-resolution systems to unify fragmented suppliers, buyers, tenders, and procurement events across heterogeneous public sources.
- Implemented automated discovery workflows for continuous monitoring of new public tenders and faster review by procurement teams.

Data Analyst & Automation Consultant

Feb 2022 - Nov 2024

Copenhagen, Denmark

Designed and implemented data intelligence systems for technology companies across Germany and the Nordic region, including Siteimprove, Hedvig, ABC Labs, and BIIOC.

- Built Python ETL pipelines integrating multiple business data sources across HR, recruitment, sales, and marketing workflows.
- Automated manual extraction, cleaning, reporting, and operational workflows using Python, n8n, Zapier, Make, and Power BI.
- Developed analytics dashboards and structured reporting systems for recruitment, sales, and operational decision-making.
- Turned distributed business data into repeatable pipelines, dashboards, alerts, and automation workflows used by non-technical teams.

Consultant, Data & Infrastructure

Mar 2020 - Jan 2021

Formiko ApS | Copenhagen, Denmark

Designed validation pipelines into ESCO/ISCO accreditation frameworks and built data infrastructure supporting cross-institutional matching workflows.

- Designed data models and validation pipelines integrating global accreditation frameworks with Danish and Swedish national registries.
- Engineered automated sync pipelines to validate international credentials against real-time registry data in Scandinavia.

SELECTED PROJECTS

Compliance Agent - Compliance Screening Tool

Public GitHub project | Python, LangChain, ChromaDB, YAML Reasoner, RAG

- Built an open-source compliance screening tool that combines regulatory retrieval, semantic drift checks, and YAML rule packs for obligations, permissions, and prohibitions.
- Designed the system to produce evidence-backed Markdown/JSON reports for human review rather than unsupervised legal advice.

LegalDrift - Semantic Version Control for Legal Text

Public GitHub and PyPI project | Python, Embeddings, Statistical NLP, Audit Logging

- Built a semantic drift detection library for comparing legal-document revisions using embeddings, statistical tests, localized drift reports, and audit logs.
- Documented where embedding-based comparison breaks down, including legal wording changes such as "may" vs. "shall", and used those findings to guide future benchmark work.

DiffDDL - Safety Checks for Agent-Generated Database Changes

Public GitHub project | Python, AST Parsing, SQL Schema Validation, Agent Tooling

- Built a schema-diffing tool that parses DDL into structured operations, allowing agent-proposed database changes to be checked for destructive actions before SQL generation.

EDUCATION

MSc Data Science, AI & Digital Business - GISMA University of Applied Sciences, Germany | 2024 - 2025

TECHNICAL SKILLS

AI Engineering: RAG pipelines, agent workflows, LLM evaluation, semantic search, prompt engineering, fine-tuning concepts, model quantization, multi-modal systems

LLM / ML Stack: Python, PyTorch, LangChain, LangGraph, ChromaDB, RAGAS, Legal-BERT, NumPy, scikit-learn

Data & Automation: ETL pipelines, entity resolution, data modeling, Power BI, n8n, Zapier, Make, audit logging

Cloud & Engineering: Git, Docker, CI/CD for MLOps, AWS SageMaker, GCP Vertex AI, Azure AI Studio