

MARTI SOURA VAMSEEKAR

Hyderabad, India • +91-9121661281 • martisoura@gmail.com • [LinkedIn](#) • [GitHub](#) • [Portfolio](#)
EU Blue Card Eligible • OPEN TO RELOCATION WITHIN THE EU

PROFESSIONAL SUMMARY

AI & Data Platform Engineer with end-to-end delivery of production RAG/LLM pipelines, event-driven microservices, EU regulatory compliance systems, and cloud data platforms. Built from first principles: a multi-tenant EU AI Act governance control plane, a SHA-256 hash-chained EU Pay Transparency analytics engine, a government-scale geospatial transport intelligence platform (56.5M population, 103 quality checks), and a cloud-native multi-service platform (6 Spring Boot microservices, 207 REST endpoints, EU VAT engine for 12 markets). Published research on SSRN, MPRA, and Zenodo. M.Sc. Data Science, University of Greenwich (Merit). Microsoft Azure Data Engineer Associate (DP-203). Targeting AI Engineer or Data Platform Engineer roles in the EU.

TECHNICAL SKILLS

AI & GenAI	Google Vertex AI, Gemini 2.5 Flash, RAG (FAISS, BM25, Cross-Encoder Reranking), Google ADK, DJL / ONNX Runtime, Evidence Graph (Evgraph), SARIF / OSCAL, scikit-learn
Data Engineering	Python, dbt, DuckDB, PySpark, ETL/ELT, Apache Spark, Parquet, Eurostat / public data pipelines
Cloud & DevOps	GCP Cloud Run, Docker, GitHub Actions, Azure Data Factory, Databricks, Data Lake Gen2, Synapse Analytics, Terraform
Backend & Systems	Java 17/21, Spring Boot 3, FastAPI, Spring Cloud Gateway, RabbitMQ, Flyway, WebSockets, Hash-chained audit logs
Databases	PostgreSQL, pgvector, Redis, MongoDB, DuckDB, Supabase, SQLite, MLflow
Governance & Reg.	EU AI Act, EU Pay Transparency Directive 2023/970, GDPR Art. 17, SHA-256 / HMAC-SHA-256 audit chains, multi-tenant isolation, EU VAT (12 countries)
Frontend	TypeScript, React 18/19, Next.js, React Native, Vite, Redux Toolkit, Tailwind CSS, shadcn/ui

EXPERIENCE

AI & Data Platform Engineer • Innosolv Private Limited (UK-based fintech, Remote) Sep 2025 – Present
Algorithmic trading & AI-powered equity research for public equity and derivatives markets

- RAG Infrastructure:** Architected production RAG across **305 annual report filings** covering multiple sectors; pipeline combines FAISS IndexFlatIP vector index, BM25 lexical retrieval, reciprocal rank fusion, and cross-encoder reranking into a **143K-chunk pipeline** with sector-aware routing; SSE-streamed responses with inline source citations.
- LLM Platform:** Owned end-to-end delivery of an institutional equity research system enforcing **10 Tier-1 quantitative screening filters** (ROE, ROCE, D/E, ICR, FCF thresholds) and a **100-point Tier-2 scoring model** with Supabase Auth and Postgres-persisted conversation history.
- Algorithmic Trading Engine:** Designed event-driven trading engine (Java, Spring Boot, MongoDB) with non-blocking async order processor and rate limiter; derivatives suite covers Iron Condor P&L, multi-leg spread margin engine, basket order aggregator, and participant-wise open interest tracker.
- Cloud Infrastructure:** GCP Cloud Run with WebSocket coordination, Redis caching for real-time market data, pgvector + Supabase persistence for AI conversations, **zero-downtime deployments** via GitHub Actions CI/CD and Docker.

Independent AI & Systems Engineer • Portfolio & Applied Research May 2022 – Aug 2025
Founder-mode product execution and applied research after M.Sc. — building production-grade, EU-regulated platforms from first principles

- Platform Delivery:** Designed, built, and maintained **5 platforms and libraries** (WorkforceGuard, Masova, Aequitas, EU AI Assurance OS, Evgraph) covering AI governance, compliance analytics, geospatial intelligence, and agentic operations — all with CI/CD, multi-tenant architecture, and regulatory audit trails.
- Research & Certification:** Published research on SSRN, MPRA (Paper No. 129330, University Library of Munich), and Zenodo; obtained Microsoft Azure Data Engineer Associate certification (DP-203, March 2025).
- Full-Stack Architecture:** End-to-end delivery spanning Java/Spring Boot microservices, Python data pipelines (dbt/DuckDB/PySpark), RAG architectures (FAISS/pgvector), and React/Next.js dashboards — deployed to GCP/Vercel with GitHub Actions.

PROJECTS & PLATFORMS

WorkforceGuard AI — EU Pay Transparency & Workforce Intelligence • [GitHub](#)
dbt • DuckDB • FastAPI • React • GCP • EU Pay Transparency Directive 2023/970/EU

- Data Pipeline:** End-to-end analytics platform ingesting **16 Eurostat datasets** (LFS, JVS, SES) across **27 EU member states and 13 NACE sectors**; **~31-model dbt pipeline** computing **4 composite indices** (HPI, LR, ERS, TR) with SHA-256 hash-chained audit log satisfying EU Pay Transparency Directive evidentiary standards; CI/CD to GCP + Vercel.
- Sector Analysis:** Finance-sector gender pay gap runs **~25%** versus a **~10.9%** all-sector panel mean; live research figures served from the production warehouse at /app/research (working paper, not peer-reviewed).
- Research Finding:** First quantitative evidence of labour market tightness positively correlated (**r ≈ +0.44**) with gender pay gaps across a **27-country Eurostat panel** — introduced the HPI × ERS Combined Risk Quadrant typology.

Masova — Cloud-Native Agentic Restaurant Management Platform · [GitHub](#)

Spring Boot · Java · GCP Cloud Run · GDPR Art. 17 · EU VAT (12 countries)

- **Multi-Client Delivery:** Shipped staff/customer web (**React 19**) plus customer and crew **React Native** apps on a single event-driven backend; optional Google ADK support agent for customer queries, with in-memory fallback.
- **Event-Driven Backend:** **6 Spring Boot microservices**, **207 canonical REST endpoints** across order management, payments, logistics, and BI; **11-state order state machine**, dual-write **PostgreSQL + MongoDB**, **3 RabbitMQ topic exchanges with DLQ**; PIN lookup latency reduced **200ms → 5ms (97.5% reduction)**; EU VAT engine (TSE, NF525, FDM, RT, NTCA) covering **12 markets** (DE, FR, IT, BE, HU, GB fiscal adapters) with Stripe SCA for EU payments.
- **Compliance:** GDPR Art. 17 compliant with 2-year customer / 7-year PCI retention workflows; Spring Cloud Gateway with JWT HS512 and per-route rate limits.

Aequitas — Public Transport Equity Intelligence Platform · [GitHub](#)

M.Sc. Dissertation → Production · FastAPI · DuckDB · FAISS RAG · React/Vite · GTFS-compliant (EU, UK, AU, NZ, USA)

- **Data Pipeline:** **7-stage validated pipeline** processing **1.75M GTFS trips**, **13,099 routes**, **274,719 stops**, **33,755 LSOAs (56.5M population)**; **103 automated quality checks**, **0 failures**; spatial join accuracy **99.9993%**; framework portable to any GTFS-compliant transport network globally.
- **ML & Equity Analytics:** Random Forest (**R^2 0.472**), HDBSCAN, Isolation Forest; 2SFCA accessibility scoring within 400m catchment; DfT WebTAG v2.03fc BCR at 3.5% HM Treasury discount; findings — **Gini 0.5741**, Palma **5.702**, **4,245 zero-stop LSOAs**, **612 triple-deprived communities**.
- **RAG Platform:** FAISS RAG chatbot powering a policy Q&A dashboard across **55 analytical sections** and **8 policy dimensions**; built for government transport authorities; Bus Services Act 2025 franchising readiness included.

EU AI Assurance OS — Governance Control Plane for EU AI Act · [GitHub](#)

Spring Boot 3.3 · Java 17 · Next.js 16 · pgvector HNSW · Multi-tenant · Regulated AI teams

- **API & Data Layer:** Spring Boot 3.3 (Java 17) backend with **Flyway migrations V1-V16**, multi-tenant JPA, JWT + API key auth — **64 REST endpoints** and **190 automated tests**.
- **Eval Gate & Audit:** Cited-evidence RAG (**DJL + ONNX Runtime**, all-MiniLM-L6-v2 embeddings, pgvector HNSW) grounds **PASS / REVIEW / BLOCKED** release gates; **HMAC-SHA-256** signed eval callbacks and a hash-chained append-only audit ledger.
- **Evidence & Reporting:** Sealed **Evidence Pack** exports (JSON + PDF); **Next.js 16** dashboard with lineage and readiness views, assisted obligation mapping, and certification readiness scoring; 3 sector overlays (insurance, HR, finance) as SPI packs, not live vendor connectors.

Evgraph — Evidence Graphs for AI System Artefacts · [GitHub](#)

Python · Evidence Graph · MLflow · SARIF · OSCAL · pytest · Library stack, local/CI only

- **Library Stack:** Shipped open-source monorepo **v0.1.0** (evgraph-core, evgraph-rules, evgraph, evgraph-cli) targeting **Python 3.10+**.
- **Evidence Model:** Typed Evidence Graph nodes/edges with bounded evidence levels; rules emit **cited, deterministic findings** without mutating the graph.
- **Integrations:** Adapters for Model Card / approval / deployment JSON, dataset manifests, and MLflow; reporters for **JSON, Markdown, SARIF, and OSCAL**; scan() / CLI for local and CI use with pytest; third-party rule packs via entry points.

RESEARCH & PUBLICATION

Why Tight Labour Markets Do Not Close Gender Pay Gaps: Evidence from a 27-Country Eurostat Panel

Marti, S. V. (2026) · [SSRN](#) · [MPRA](#) · [Zenodo](#) · ORCID: 0009-0005-4884-1292

- **Methodology:** Applied panel data econometrics + ML across **27-country Eurostat panel** (2019–2024, **11 NACE sectors**); found **$r \approx +0.44$** positive correlation between labour market tightness and gender pay gaps — contradicts competitive equalisation theory; introduced Combined Risk Quadrant ($HPI \times ERS$) as first integrated tightness-equity typology in literature. JEL: J31, J71, J16, J21, C38, C55, K31.
- **Implementation:** Findings implemented in WorkforceGuard open-source analytics system; **4 composite panel indices** (HPI, LR, ERS, TR) via PCA-informed weighting; fixed-effects regression, Hausman specification test, and heteroskedasticity-robust standard errors.

CERTIFICATIONS & EDUCATION

Microsoft Certified: Azure Data Engineer Associate (DP-203)

Mar 2025 · ID: CEA6999835BAD6E

M.Sc. Data Science — Merit · University of Greenwich

London, UK · Jan 2021 – Apr 2022

Dissertation: UK Bus Data Analysis (foundation for Aequitas) · Modules: Big Data, Applied ML, Data Visualisation, Cloud Grids

B.Tech. Electronics & Communication Engineering — 8.3 CGPA · GITAM Institute of Science and Technology

Visakhapatnam · 2016 – 2020

Dissertation: Hyperspectral Image Analysis with Deep Learning · Modules: Data Structures & Algorithms, Digital Signal Processing