ERMIAS MIKAEL

PORTFOLIO: https://ermiasmikael.dev

SOFTWARE ENGINEER — AI APPLICATION DEVELOPMENT

Software engineer with 7+ years of building and shipping production-grade enterprise applications across federal and commercial environments, now focused on AI-integrated software. Designs and delivers backend systems end-to-end — Java 21 / Spring Boot microservices, event-driven architecture, and relational data layers — with LLM capabilities built in as production features: agentic tool-calling, retrieval-augmented generation, and human-in-the-loop guardrails. M.S. in Data Science; AWS Solutions Architect certified.

TECHNICAL SKILLS

Languages: Java 21, Python, JavaScript, SQL / PL-SQL, Bash

Backend & Architecture: Spring Boot, FastAPI, REST APIs, Microservices, Event-Driven Architecture, Domain-Driven Design, Hexagonal Architecture, Kafka

AI Engineering: LLM integration (LiteLLM, Ollama, self-hosted open-weight models), Agentic systems & tool-calling, MCP, RAG, Vector search & embeddings, Prompt engineering, Transformers / Hugging Face, LLM security (OWASP LLM Top 10)

Data: PostgreSQL, Oracle, SQL Server, MySQL, Redis, OpenSearch, ETL/ELT, Data modeling, Batch processing

Cloud & DevOps: AWS, Docker, GitLab CI/CD, Keycloak (OIDC/JWT), Prometheus, Grafana, Linux

ENGINEERING PROJECTS

PROVA — AI-Enabled Trade Compliance SaaS Platform *Complete — production-ready*

Solo-architected and built a full multi-tenant SaaS platform for trade sanctions screening. Source and architecture documentation: github.com/ermiasmikael

- Built the platform backend as event-driven Java 21 / Spring Boot microservices using hexagonal architecture and domain-driven design, with Kafka/Redpanda messaging and PostgreSQL persistence.
- Engineered an agentic AI tier (Python / FastAPI) that invokes platform APIs through a model-agnostic LLM abstraction layer (LlmPort + LiteLLM) and the Model Context Protocol, enabling the LLM to act on the system through defined tools rather than generating free text.
- Designed the agent security model around production risk: closed tool allow-lists, per-user token propagation, and mandatory human-in-the-loop approval gates on any state-changing action — addressing the OWASP LLM Top 10 threat classes.
- Implemented cite-or-abstain RAG over compliance regulations so the system either grounds every answer in a retrievable source or declines to answer, pairing LLM reasoning with deterministic rules to keep screening decisions explainable and auditable.
- Ran all inference on self-hosted open-weight models (Qwen2.5 7B via Ollama, nomic-embed-text embeddings) for zero-egress deployment, eliminating third-party data exposure for regulated customer data.
- Delivered the full operational stack: Keycloak OIDC/JWT multi-tenant authentication, containerized deployment, GitLab CI/CD, and observability via Prometheus, Grafana, Loki, and Tempo.

PsyAI — Clinical Decision Support System *M.S. Capstone, UMBC*

- Built an NLP system applying DSM-5-TR diagnostic criteria to clinical text, including the full pipeline from PDF criteria extraction through structured dataset construction to model serving.
- Fine-tuned BioClinicalBERT for psychiatric classification, achieving 93.35% accuracy and 93.0% F1.
- Added SHAP-based explainability so clinicians could inspect why the model produced a given output — a requirement for any decision-support tool in a clinical setting.

PROFESSIONAL EXPERIENCE

Software Developer | Spatial Front, Inc. *Oct 2019 – Present*

- Design, build, and maintain enterprise web applications supporting national statistical programs for federal clients, owning features from requirements through production deployment and support.
- Develop the application data layer in Oracle and PL/SQL — stored procedures, packages, and reusable database components — used across multiple applications in the program.
- Build and optimize batch-processing pipelines, tuning queries, creating and orchestrating automation scripts, and reducing runtime and operational overhead.
- Build front-end application functionality in JavaScript and integrate it against backend services and data APIs.
- Work within a federal SDLC: peer code review, release management, defect triage, and coordination with client stakeholders and QA.

Application Support Analyst | MPL Association *Apr 2019 – Aug 2019*

- Developed public-facing web applications and microsites; administered the iMIS database, RISE web application, and MS SQL / PostgreSQL environments; provisioned and maintained supporting infrastructure.

Technical Business Analyst | ACGI Software *Aug 2018 – Mar 2019*

- Translated client requirements into technical specifications for development teams and supported Oracle APEX and Business Objects reporting solutions, performing issue analysis and solution validation across enterprise applications.

EARLIER EXPERIENCE

Business Analyst | American Society of Appraisers *Jul 2017 – Aug 2018*

Led IT initiatives including website modernization; conducted business process analysis and authored system requirements.

Business Analyst II | Jacobs Technology *Aug 2016 – Jul 2017*

Supported federal application systems; performed cross-platform data and process analysis; produced 508-compliant technical documentation.

QA / Business Analyst | OCR Services, Inc. *Feb 2013 – Aug 2015*

Executed requirements analysis, produced technical specifications, and partnered with developers on implementation and QA validation.

Lead Communications Consultant | Sprint Nextel *Mar 2005 – Feb 2013*

Led a 26-person operations team managing KPIs, scheduling, training, and service delivery.

EDUCATION & CERTIFICATIONS

M.S., Data Science | University of Maryland, Baltimore County (UMBC) *2024*

B.S., Computer and Information Science | University of Maryland University College *2012*

Certifications: AWS Certified Solutions Architect – Associate | Project Management Professional (PMP) | Associate Systems Engineering Professional (ASEP)