

YOSRI KHIARI

Software Engineer — Python, TypeScript, C#, Go, Java

Kelibia, Nabeul, Tunisia | +216 22-323-039 | khiariyosri02@gmail.com | github.com/yosriKhiari | linkedin.com/in/yosriKhiari | yosriKhiari.vercel.app

PROFESSIONAL SUMMARY

Software engineer who ships production systems across the full stack — from ML models and LLM pipelines to frontend applications and container orchestration. 13 months of internships across three companies in Python, TypeScript, C#/.NET, Go and Java, delivering a semantic search engine (6 custom services on a 29-container stack), a multi-tenant ERP, a demand-forecasting system and a warehouse optimization tool. Quick to ramp on new domains and codebases, with a track record of owning features end to end from database schema to user-facing UI.

TECHNICAL SKILLS

Languages: Python, TypeScript, JavaScript, C#, Go, Java, SQL, Bash

Backend & APIs: ASP.NET Core (.NET 8/10), Spring Boot, Go, REST APIs, Microservices, Event-Driven Architecture, Streaming (SSE, SignalR), RabbitMQ, Kafka, Hangfire, Entity Framework Core

Frontend: React, Vue 3 (Composition API, Pinia), Angular, Tailwind CSS, Vite, TipTap, Dexie/IndexedDB, Chrome Extensions (MV3)

Databases: PostgreSQL (Row-Level Security), SQL Server, MongoDB, Redis, OpenSearch / Elasticsearch, SQL / NoSQL

Cloud & DevOps: Docker, Kubernetes (k3s, Kustomize, Rancher), Azure, OpenStack, CI/CD (GitHub Actions, Jenkins), Linux

AI & ML: LLMs, Retrieval-Augmented Generation (RAG, self-verifying loops), GraphRAG, OCR/NER (Tesseract, Apache Tika, GLiNER), Prompt Engineering, Ollama, OpenAI / Anthropic / Gemini / Grok APIs, Prophet, XGBoost

Observability & Testing: OpenTelemetry, Prometheus, Grafana, Loki, Tempo, k6 Load Testing, Unit / Integration / E2E Testing (Vitest, Playwright, xUnit, pytest), SonarQube, Git, Agile, Scrum

PROFESSIONAL EXPERIENCE

Software Engineer Intern — Neolledge MEA

Feb 2026 – Aug 2026

Final-Year Engineering Project (PFE) · ASP.NET Core 8, OpenSearch, RabbitMQ, Redis, Ollama, GLiNER, Vue 3, Docker

Lead developer of a semantic document search engine — 6 custom services (ASP.NET Core API, 3 workers, Python NER service, Vue 3 frontend) on a 29-container Docker Compose stack — vector search, RAG and GraphRAG, from ingestion to observability.

- **Retrieval-Augmented Generation:** Built the full RAG backend — query classification, ACL-filtered chunk retrieval, reranking and grounded answer generation with citations — plus a GraphRAP path over GLiNER-extracted entities and relations.
- **Document ingestion pipeline:** Designed async Apache Tika to Tesseract OCR to GLiNER NER to format-aware chunking to OpenSearch indexing over RabbitMQ with a transactional outbox, idempotent uploads and dead-letter retry.
- **Vue 3 frontend:** Built a streaming search UI with real-time answer delivery over SSE, source-chunk citations, faceted filters and an admin console for document and user management.
- **Multilingual query understanding:** Implemented NLP query parser covering 300+ French, English and Arabic patterns at sub-millisecond latency, so non-technical users filter by date, author, amount and type in plain words.
- **Observability & CI:** Deployed OpenTelemetry with Prometheus, Grafana, Loki, Tempo and Alertmanager — 7 dashboards and distributed traces — plus an 867-case evaluation harness gating merges on quality thresholds.

Full-Stack Developer Intern — HorizonLux

Sep 2025 – Jan 2026

.NET 10, React / TypeScript, PostgreSQL, Redis, Docker, Kubernetes (k3s, Kustomize), Traefik, Rancher, GitHub Actions, GHCR

Erpero — multi-tenant ERP for Tunisian SMEs: owned expense, tax-compliance, payroll and accountant modules in a 5-person team.

- **Expense management:** Shipped the module end to end — approval workflows, receipt OCR with automatic vendor creation, FODEC withholding, split payments, bulk operations and accountant handoff queues.
- **Tax compliance engine:** Implemented Tunisian TEJ withholding-declaration exports with rectificative support, FODEC calculations, VAT rules and a fiscal-obligation calendar, replacing spreadsheet-built declarations.
- **Accountant platform:** Delivered a 15-ticket cabinet console — client engagements, RBAC presets with granular overrides, a deliverables/obligations engine and an isolation-aware audit log.
- **HR & payroll:** Built contract lifecycle, employee creation from a signed work contract via document extraction, and a Net-to-Brut payroll solver that inverts the Tunisian tax scale.

- **Multi-tenant isolation:** Extended the integration harness to prove 5-layer tenant isolation (PostgreSQL Row-Level Security, fail-closed EF Core filters), making cross-tenant data leakage a build failure.
- **Coppler (self-verifying RAG):** LinkedIn content-repurposing Chrome extension (MV3) plus .NET Capture/Content/Publishing modules — multi-provider adapter layer (Grok / xAI, OpenAI, Anthropic, Gemini) with automatic fallback; each request is embedded, the top-3 similar past outputs are retrieved for few-shot injection, and the result is embedded again and regenerated when cosine similarity > 0.92 flags a near-duplicate; multi-turn context cache, LinkedIn OAuth publishing and Hangfire-scheduled posts.

ML Engineer Intern — L-Mobile

Jul 2025 – Aug 2025

Python (Prophet / XGBoost), ASP.NET Core, Redis, Prometheus, Grafana, Docker, Azure

Hybrid time-series forecasting system for SKU-level demand planning.

- **ML pipeline:** Orchestrated Prophet feature extraction and XGBoost training from ASP.NET Core as one reproducible pipeline, so planners retrain from the UI instead of running notebooks.
- **Serving & monitoring:** Implemented Redis caching on forecast queries, Prometheus/Grafana dashboards on inference latency and uptime, Docker images delivered through Azure Container Registry.

Full-Stack Web Developer Intern — L-Mobile

Jul 2024 – Sep 2024

Angular, Spring Boot, MySQL

- **Warehouse routing optimization:** Built a forklift routing application from requirements gathering through deployment in an Agile cross-functional team, improving logistics efficiency by 25%.

PROJECTS

AgentOps — LLMops Benchmarking Platform

github.com/yosrikiari/AgentOps

Go, Docker, Prometheus, Grafana, k6, GitHub Actions, OpenAI / Anthropic / Gemini / Groq / Ollama

Production-grade benchmarking tool that stress-tests LLM providers under realistic traffic and monitors every metric in real time.

- **Go service & monitoring:** Dockerized Go binary with Prometheus metrics export, Grafana dashboards for latency percentiles, throughput, error rates and cost; k6 load-testing scripts simulating concurrent users.
- **Policy engine:** 11 configurable policies — rate limiting, retry with exponential backoff, circuit breakers, cost caps — enforced uniformly across OpenAI, Anthropic, Gemini, Groq and Ollama via a provider-agnostic Go interface.
- **CI/CD:** GitHub Actions for automated build, test and release; quality gates on every commit.

Versatile — Offline-First AI Fiction Writing Platform

github.com/yosrikiari/Versatile

Vue 3, Pinia, TipTap, Dexie/IndexedDB, LangGraph.js, .NET 10, Ollama / OpenAI / Anthropic / Gemini / Groq, AgentOps tracing, Docker

A local-first writing environment where the entire AI pipeline runs on the user's own machine.

- **Offline-first architecture:** Vue 3 + Pinia frontend with Dexie/IndexedDB persistence and a sync engine to an optional .NET 10 API; TipTap rich-text editor, chapter/scene/timeline management, story-network graph — all working offline.
- **Multi-agent AI:** LangGraph.js graph with Writer, Critic and Editor agents on separate GPU / CPU lanes, checkpointed per step, traceable end to end through an AgentOps gateway; one interface over Ollama, OpenAI, Anthropic, Gemini and Groq; approximately 3,100 automated tests (Vitest + Playwright), Storybook, SonarQube.

CERTIFICATIONS

Network Security — Cisco | Apr 2026

AWS Academy Graduate — Cloud Operations — Amazon Web Services | Nov 2025

Hashgraph Developer Course — The Hashgraph Association / Hedera | Nov 2025

CCNA: Introduction to Networks — Cisco | May 2024

EDUCATION

Engineering Degree in Computer Science — ESPRIT, Tunis2023 – 2026

Preparatory Diploma in Mathematics & Theoretical Physics — ISSAT Mahdia (ISSATMA)2021 – 2023

LANGUAGES

Arabic (Native) | French (B2) | English (B2)