

Ali Asfahani

Beirut, Lebanon | ali.asfahani.lb@outlook.com | +961 76 415 001

EDUCATION

MS – Artificial Intelligence

2025 – 2027

Université Saint Joseph (USJ) – ESIB

BS – Computer Science

2020 – 2023

CNAM University

EXPERIENCE

Freelance Software Developer

Feb 2023 – Dec 2024

- Delivered custom software and web solutions for clients, translating requirements into maintainable systems.
- Built and modified applications using Java and web technologies, focusing on reliability and practicality.
- Worked directly with clients to diagnose issues, propose solutions, and iterate based on feedback.

Web Development Intern

Nov 2022 – Jan 2023

- Developed a dynamic web platform for an education center using HTML, CSS, JavaScript, PHP, and async communication.
- Designed system architecture including actors, workflows, and database models.
- Implemented role-based access control and contributed within an agile development process.

PROJECTS

Maintainer's Copilot *NLP, Advanced RAG, LLM Agents, React*

- Built an AI assistant that classifies software issues, extracts entities, summarizes discussions, and answers questions using documentation and resolved issues.
- Compared classical ML, a fine-tuned transformer, and an LLM baseline; implemented hybrid retrieval, reranking, metadata filtering, and query rewriting.
- Developed a tool-calling chatbot with Redis and pgvector memory, an embeddable React widget, tracing, redaction, and CI-gated evaluations.

Document Classifier Service *ConvNeXt, FastAPI, SFTP Pipeline, HashiCorp Vault*

- Fine-tuned ConvNeXt Tiny/Small on RVL-CDIP (16 document classes, 320k images) achieving production-grade accuracy; shipped model weights, a SHA-256 model card, and a 50-image golden set.
- Built a layered FastAPI service (api, services, repositories, domain) with JWT auth, Casbin RBAC (3 roles), Redis-backed caching, and Alembic migrations.
- Delivered a full ingestion pipeline (SFTP to MinIO to RQ worker to inference) with Vault secret resolution, structured JSON logs, and CI gates enforcing golden-set replay.

Drift Triage Co-Pilot *MLOps, LangGraph Supervisor, Redis Queue, Self-Healing Pipeline*

- Built a self-healing MLOps stack: trained and registered a binary classifier with MLflow, served predictions via FastAPI, and computed drift reports (PSI, chi-squared) over rolling windows.
- Implemented a LangGraph supervisor with three sub-agents, Postgres checkpoint persistence, and human-in-the-loop approval gating production registry changes.
- Dispatched long-running tools through a Redis queue with idempotency keys, exponential backoff, and a dead-letter queue; agent routing covered by snapshot regression tests in CI.

Concierge AI SaaS *Multi-Tenant AI, RAG, FastAPI, PostgreSQL*

- Developed a multi-tenant AI concierge platform with CMS-managed knowledge, an embeddable website assistant, lead capture, and human escalation.
- Built a hybrid routing system handling predictable requests through classifier-driven workflows and delegating complex requests to a bounded tool-calling agent.
- Enforced tenant isolation using PostgreSQL RLS, tenant-filtered vector retrieval, signed widget tokens, configurable guardrails, and cross-tenant security tests.

Spectra AI — Screen-Aware Cross-Platform Copilot (In Development) *RAG, FastAPI, Computer Vision*

- Developing a multimodal AI copilot that combines screen understanding, browser DOM extraction, platform data, uploaded files, and official documentation to deliver contextual guidance, cross-platform analytics, and grounded conversational support.
- Integrating ML/DL models for visual interface understanding, semantic retrieval, intent routing, entity matching, and anomaly detection, alongside deterministic business analytics, cited RAG responses, authenticated chat with memory, secure secrets management, usage tracking, and automated Excel reporting.
- More features and details will be added.

SKILLS

Programming: Python, Java, SQL, OOP, Design Patterns, REST APIs, Agile Development

AI / Data: Probabilistic Modeling, Computer Vision, NLP, Fine-Tuning, RAG, Model Evaluation, Drift Detection, Scikit-learn, PyTorch, OpenCV

LLMs / Agents: Prompt Engineering, Embeddings, Semantic Search, LLM Prompt Chaining, Tool-Calling Agents

Backend / Infra: FastAPI, Redis Queue, PostgreSQL, pgvector, SQLAlchemy, Docker, HashiCorp Vault

Web: HTML, CSS, JavaScript, PHP, React

BOOTCAMPS

SE Factory – AI Engineering Bootcamp

Forward MENA – Full-Stack Development with AI

LANGUAGES

Arabic (Native), English (Fluent), French (B2)