

Soufien HAJJI

Lead Software Engineer · AI Engineer on Azure

haji.soufien@gmail.com · +33 6 69 04 92 63 · Thionville, France, frontière luxembourgeoise ·
linkedin.com/in/hajji-soufien · github.com/SFHAJJI · energy.soufien.lu



PROFIL

Ingénieur logiciel avec 9 ans d'expérience en .NET / Azure, qui conçoit et livre aujourd'hui des **systèmes GenAI de production, de bout en bout**, agents LLM & appels d'outils, RAG, NL-to-SQL, qui restent **ancrés (grounded), gouvernés (AI Act / RGPD) et reproductibles**. **Certifié Microsoft : Azure AI Apps and Agents Developer Associate**. Chez Natixis Investment Managers, j'ai bâti une plateforme de risque financier (RWA), puis la **couche d'IA agentique** au-dessus. Seul, j'ai conçu, développé et déployé **energy.soufien.lu**, un produit RAG agentique en production qui **aide les chefs de projet à prospecter les opportunités solaires photovoltaïques** dans la transition énergétique luxembourgeoise, et je construis actuellement **law.soufien.lu**, la consultation du droit luxembourgeois & européen tel qu'il était à une date donnée, avec une provenance vérifiable, exposée via un endpoint MCP public. À l'aise du modèle de données au backend, à l'API, à l'UI et au cloud/DevOps.

PROJETS PERSONNELS · CONÇUS, DÉVELOPPÉS & DÉPLOYÉS EN SOLO

energy.soufien.lu · Co-Pilote de la Transition Énergétique du Luxembourg

2025 · en ligne

Un produit RAG agentique et ancré qui aide les chefs de projet à prospecter où installer du photovoltaïque (toitures & parkings) et répond aux questions de droit de l'énergie, de subventions & de tarifs, déployé sur Azure.

- **Orchestrateur agentique** (Azure OpenAI + **Microsoft Agent Framework**), un seul agent pilotant une **architecture à 5 outils** : langage naturel → SQL sur une carte/liste PostGIS de toitures & parkings, Q&R juridiques RAG citées via un **serveur de loi MCP** autonome, un moteur de tarifs et un chercheur d'appels d'offres PV, le LLM rédige, **les chiffres sont calculés de façon déterministe** (jamais inventés).
- **Azure AI Content Understanding** transforme les documents d'appels d'offres publics (HTML + PDF) en prédicats machine → **filtres d'UI dynamiques** + un « **appliquer l'appel d'offres** » **déterministe** reproduisant à chaque fois l'ensemble exact et certifié de candidats (3 373 toitures + 831 parkings) ; **vision multimodale zero-shot** pour classer le PV existant sur ~5 200 toitures et ~850 parkings à partir d'imagerie aérienne.
- **IA responsable par conception** (ancré & cité ou refusé, SQL en bac à sable SELECT uniquement, un « juge » d'ancrage, mémoire respectueuse du RGPD) et **maîtrise full-stack** : API C# · pipelines Python · React/MapLibre · PostgreSQL/PostGIS/pgvector · Terraform IaC (18 modules) + CI/CD GitHub Actions vers Azure Container Apps & Static Web Apps.

law.soufien.lu · Lex, le droit luxembourgeois & européen à une date donnée

EN COURS

2026 · en cours

Répond à une question que les portails officiels ne traitent pas : que disait exactement ce texte à cette date ?, 1 407 lois conservées en 4 693 versions datées remontant à 1849, chaque réponse traçable jusqu'aux octets de l'éditeur.

- **Architecture de données à deux couches** : une couche de preuve en ajout seul qui stocke les octets mêmes des éditeurs avec un sha256 par observation, et une **couche dérivée déterministe** (Markdown + JSON par article, chaînés par hachage) reconstruite de façon reproductible, toute affirmation remonte donc à la source officielle. Les profils d'extraction (Akoma Ntoso, Formex 4) sont immuables, verrouillés par des tests d'empreinte figée.
- **Serveur MCP public (9 outils)** sur un index SQLite/FTS5 signé dont la signature ECDSA couvre une empreinte du contenu lui-même, plus une couche IA ancrée sur ces mêmes outils, avec une **taxonomie de refus comme contrat d'API** : « aucune version à cette date » est une réponse, jamais une supposition. C# / .NET 10 · React · Azure Container Apps · ingestion nocturne automatisée avec veille active · 61 tests + evals notés par LLM bloquant la CI.

soufien.lu · assistant IA de portfolio (complément de ce CV)

2026 · en ligne

- Un assistant RAG agentique et ancré qui répond aux questions sur mon expérience & mon code et évalue une offre d'emploi collée face à mon profil. Un **orchestrateur** + **outils** sur **Microsoft Agent Framework** : recherche de connaissances (**Azure AI Search**, hybride vecteur+mots-clés), récupération du code source réel et un **scorecard d'adéquation déterministe**, ancré-ou-refus, Prompt Shields sur les entrées non fiables, Terraform IaC et CI/CD à eval bloquant.

EXPÉRIENCE PROFESSIONNELLE

Ingénieur Logiciel & IA · Natixis Investment Managers

Oct. 2024 - Présent · Paris

Conception & développement d'une plateforme interne de capital réglementaire (RWA), puis de la couche d'IA agentique au-dessus.

Phase 1 · Ingénierie Logicielle : la plateforme RWA

Calcule & reporte le capital réglementaire sur les risques de crédit & de marché · ASP.NET Core 8 (C#) ↔ Python (FastAPI) ↔ SQL Server, intégrant MSCI RiskMetrics · déployé sur IIS.

- **Trois modules métier**, un workflow audité d'**inventaire crédit** en 5 étapes (import Excel → normalisation → rapprochement avec la base historique → attribution des codes contrepartie → contrôles qualité → un jeu de données enrichi validé, avec points de reprise & rollback) ; le **reporting réglementaire crédit** sur deux axes d'exposition (interne + titrisé en transparence) via un pipeline enrichissement → contrat → encours → titres jusqu'à des livrables Excel auditable ; et un module de **risque de marché** calculant le capital FRTB (méthode par sensibilités) via **MSCI RiskMetrics**, transparisation, blends, sensibilités et données statiques.
- Architecture de la solution en couches **.NET 8 MVC** (DI, EF Core / SQL Server) avec points de reprise & audit pour la reproductibilité ; data engineering (ETL Excel/XML/SOAP, fuzzy matching, moteur de règles qualité), UX temps réel (SignalR),

CI/CD + tests unitaires.

- Conception de l'**intégration .NET ↔ Python**, un **sidecar FastAPI** co-localisé (HttpClient typé, health-checké) porte le SDK MSCI (Python uniquement) + le calcul de risque de marché. Diagnostic & résolution d'un blocage réseau tenace, whitelisting d'IP source derrière un proxy NTLM/Kerberos (**403 depuis le cloud**), avec les équipes réseau/sécurité + MSCI.
- Prise en charge des sujets de production : **authentification AD / LDAP** (claims/policies), journalisation **Serilog** avec corrélation IDs, résilience (**retries EF, fallback Redis→mémoire, sidecar dégradé proprement**), configuration multi-environnements.

.NET 8 / ASP.NET Core MVC **EF Core** **Python / FastAPI** **SQL Server** **MSCI RiskMetrics (SOAP)** **SignalR** **Razor / Bootstrap / Vite**
Serilog · IIS **RWA · FRTB/SBM · Bâle/CRR**

Phase 2 · Ingénierie IA : la couche agentique (PoC → production)

Construction & industrialisation d'un assistant IA au-dessus de la plateforme : les utilisateurs métier interrogent les données réglementaires en langage naturel, avec des chiffres déterministes et auditable.

- **Trois capacités**, un **assistant d'analyse NL** (questions en langage courant sur les données crédit et marché ; le modèle rédige, le moteur calcule), un **assistant de connaissances RAG** ancré sur les documents réglementaires & méthodologiques (FRTB/CRR, méthodologie MSCI, spécifications internes), et **Document Intelligence** (téléversement → extraction automatisée → analyse restituée dans l'application).
- Construction d'un **agent à outils** sur un ensemble d'outils curé (API REST de la plateforme, un outil analytique **text-to-SQL**, un outil de récupération **RAG**, l'extraction de documents), agissant via des services déterministes, sans jamais fabriquer de chiffres ; pipeline d'ingestion/enrichissement RAG (charger → enrichir → chunker → embedder → vector store).
- **IA responsable** (accès SELECT uniquement, ancrage + citations, prompt engineering, human-in-the-loop, garde-fous) + **evals** (exactitude/fidélité) et observabilité en production ; livré comme module natif via la CI/CD existante.

agents LLM · appels d'outils **RAG · embeddings · vector store** **text-to-SQL** **IA responsable · garde-fous** **prompt engineering** **evals**
Azure App Insights **Azure AI Document Intelligence** **Python · SQL · C#**

Ingénieur Logiciel .NET, Prêts & Titres · Société Générale

Juil. 2021 - Sept. 2024 · Paris La Défense

- **Plateforme de trading** front-office (client lourd, X-One), desk de prêt/emprunt de titres. Conception & développement du **moteur de calcul de position** des deals + **valorisation du collatéral** (données clés pour l'exposition du Risque), ainsi qu'un **module de prepay-offset** avec une stratégie dynamique à l'exécution pour les flux d'intérêts anticipés.
- Pilotage de la migration **.NET 4.8 → .NET 6** ; SOLID / DI / TDD. Rôles tournants : task-force qualité (anti-code-freeze) + support de production front-office. Équipe ~30 Paris / ~20 Bangalore, CI Jenkins.

Consultant .NET / Azure · SNCF Réseau

Mai 2019 - Juin 2021 · Paris

- Application web de supervision des trains en temps réel (gares ↔ technicentres) avec **généricité dynamique par technicentre**, **ASP.NET MVC 5** + APIs métier en **ASP.NET Core 3.1**, **SignalR** (temps réel + chat intégré). Mise à l'échelle pour la concurrence : load-balancing IIS, traitement asynchrone, lazy-loading.

Consultant Azure / Développeur .NET · Orange Business Services

Janv. 2018 - Avr. 2019 · Massy

- Développement d'une **plateforme interne de billetterie événementielle** (concerts, salons & événements d'entreprise, avec **chaînage de billets** & paiement sécurisé) en **microservices** (.NET Core 3.1, Docker) derrière **Azure APIM** ; parc Azure complet (AKS, App Services, Key Vault, SQL Azure) via **Terraform IaC** + CI/CD Azure DevOps ; également des composants de gestion documentaire (API BOX, EF).

COMPÉTENCES TECHNIQUES

IA / LLM	agents LLM & appels d'outils (function calling) · Microsoft Agent Framework · RAG (chunking, embeddings, pgvector) · NL-to-SQL · serveurs MCP (Model Context Protocol) · prompt engineering · evals · IA responsable / garde-fous · Azure OpenAI · Content Understanding · Document Intelligence
Langages	C# · Python · TypeScript / JavaScript · SQL
Backend	ASP.NET Core 8 / MVC · EF Core · FastAPI · SignalR · REST & SOAP
Données	PostgreSQL / PostGIS / pgvector · SQL Server · Oracle · SQLite / FTS5 · pandas · pipelines reproductibles & provenance adressée par contenu
Frontend	React · Vue 3 / PrimeVue · MapLibre · Vite · Bootstrap
Cloud / DevOps	Azure (Container Apps, Static Web Apps, OpenAI, AI Foundry, Functions, Key Vault, App Insights, APIM) · Terraform (IaC) · GitHub Actions & Azure DevOps · Docker · IIS
Pratiques	SOLID / architecture propre · TDD & tests unitaires · gouvernance AI Act & RGPD · mentorat & documentation

CERTIFICATION

Microsoft Certified : Azure AI Apps and Agents Developer Associate

Obtenue en juil. 2026 · examen AI-103 · [vérifier sur Microsoft Learn](#) ↗

Ensuite : AI-200 · Azure AI Cloud Developer Associate (en préparation) → AI-500 (bêta) · Multi-Agent AI Solutions Expert → à plus long terme, TOGAF EA Practitioner.

FORMATION

ENSTA ParisTech, Diplôme d'ingénieur (Bac+5)

Architecture & Sécurité des Systèmes d'Information · 2015-2018

LANGUES & CENTRES D'INTÉRÊT

Français, bilingue · Anglais, courant (TOEIC 895/990) · Allemand, notions

Défis de code (9e / 148, Spring Challenge 2020) · dev de jeux Unity · jeux de stratégie