Soufien HAJJI



Lead Software & AI Engineer · Copilotes d'entreprise en production · Azure

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

PROFIL

Lead Software & AI Engineer, 9 ans à concevoir et exploiter des systèmes métier critiques en C#/.NET sur Azure, de l'architecture à la production. Je fais passer les assistants IA **du prototype à l'exploitation** : logique déterministe dans du code typé, réponses ancrées et citées ou refusées, évaluations automatisées qui bloquent les régressions, et déploiements observables. Décisions d'architecture documentées, alternatives rejetées et critères de réversibilité. ENSTA ParisTech (Architecture & Sécurité des SI). **Microsoft Certified : Azure AI Apps and Agents Developer Associate.** TOGAF en préparation.

EXPÉRIENCE PROFESSIONNELLE

Ingénieur Logiciel & IA · Natixis Investment Managers

Oct. 2024 - Présent · Paris

Prise en charge des deux workflows de crédit, industrialisation de l'intégration du risque de marché, puis construction et déploiement de la couche IA sur les trois workflows. La plateforme couvre 8 portefeuilles d'investissement.

Phase 1 · Ingénierie de solutions : crédit réglementaire et industrialisation du risque de marché

Responsable de la conception, du développement et du support en production de deux workflows de crédit réglementaire au sein d'une plateforme couvrant 8 portefeuilles d'investissement.

- Conception et développement d'un workflow audité d'inventaire crédit couvrant l'ingestion, la normalisation, le rapprochement historique, l'identification des contreparties et les contrôles qualité, avec points de reprise reproductibles, rollback et données enrichies validées.
- Construction du workflow de reporting des expositions internes et titrisées en transparence, de l'enrichissement aux contrats, encours et titres jusqu'aux livrables réglementaires auditables.
- Livraison de l'application de bout en bout en **.NET 8**, des services de workflow et de traitement des données à l'expérience utilisateur Razor et SignalR, avec EF Core et SQL Server.
- Intégration d'un workflow Python spécialisé de risque de marché dans la plateforme derrière un service **FastAPI**, avec définition de la frontière .NET–Python, de l'orchestration et de l'isolation des défaillances, tout en maintenant les calculs quantitatifs sous la responsabilité du composant spécialisé.
- Mise en production de la solution sur IIS dans les environnements de développement, UAT et production, avec contrôle d'accès AD/LDAP, configuration par environnement, pipelines automatisés de build, test et déploiement, et logs Serilog corrélés.

`.NET 8 / ASP.NET Core MVC` `Razor / SignalR` `EF Core / SQL Server` `Python / FastAPI` `Livraison full-stack` `Tests automatisés` `CI/CD / IIS` `AD / LDAP` `Serilog` `RWA · FRTB/SBM`

Phase 2 · Ingénierie IA : copilote interne en production sur les workflows existants

Conception et mise en production d'un copilote interne sur trois workflows métier couvrant 8 portefeuilles.

- Guidage utilisateur, explication des pipelines et analyses en langage naturel sur les données de production. Conception d'un système de connaissance Document Intelligence et RAG, ancré sur les spécifications techniques et fonctionnelles, avec outils curés, citations obligatoires, accès analytique SELECT-only, garde-fous, revue humaine, évaluations d'exactitude et de fidélité, et observabilité de production.
- Le modèle orchestre et explique ; les services déterministes conservent la maîtrise de l'état et des calculs financiers.

`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

- Développement de X-One, plateforme de trading front-office sensible à la latence et à l'exactitude pour le desk de prêt-emprunt de titres, avec les équipes de trading et d'exploitation au sein d'une organisation d'environ 30 ingénieurs à Paris et 20 à Bangalore.
- Conception et développement du moteur de calcul de position des deals et des fonctions de valorisation du collatéral, données clés de l'exposition au risque. Livraison du prepay-offset avec stratégies sélectionnables à l'exécution pour les flux d'intérêts anticipés.
- Pilotage de la migration de .NET Framework 4.8 vers .NET 6, introduction de l'injection de dépendances et renforcement des pratiques SOLID et TDD. Rotation sur le support de production front-office et une task force qualité traitant les défauts structurels avant blocage des livraisons.

`C# / .NET 6` `Trading front-office` `Moteur de position` `Valorisation collatéral` `TDD / SOLID` `Jenkins`

Consultant .NET / Azure · SNCF Réseau Mai 2019 - Juin 2021 · Paris

- Construction d'une application opérationnelle de supervision des trains en temps réel reliant gares et centres de maintenance, avec une vue dynamiquement configurée par centre sur les trains, événements et workflows.
- Développement de l'application ASP.NET MVC 5 et des API métier ASP.NET Core 3.1, avec SignalR pour les mises à jour et le chat intégrés. Amélioration de la réactivité et de la capacité par traitement asynchrone, lazy loading et répartition de charge IIS, à la jonction des composants .NET historiques et modernes.

ASP.NET MVC / Core · SignalR · Systèmes temps réel · Répartition de charge IIS · Traitement asynchrone

Consultant Azure / Développeur .NET · Orange Business Services Janv. 2018 - Avr. 2019 · Massy

- Construction d'une plateforme interne de billetterie pour concerts, salons et événements d'entreprise, avec chaînage des billets et paiement sécurisé, sous forme de microservices .NET Core conteneurisés par Docker et exposés via Azure API Management.
- Mise en place de pipelines automatisés de build, d'intégration et de déploiement dans Team Foundation Server, ainsi que livraison de composants de gestion documentaire intégrés à l'API Box et Entity Framework.

Microservices .NET · APIM · Docker · Team Foundation Server · CI/CD · API Box / EF

PROJETS PERSONNELS · CONÇUS, DÉVELOPPÉS & DÉPLOYÉS EN SOLO

law.soufien.lu · Lex, le droit luxembourgeois & européen à une date donnée 2026 · en ligne

Droit luxembourgeois et européen à une date donnée, avec IA ancrée, provenance vérifiable et architecture éditeur extensible.

- Conception et exploitation de **2 647 œuvres et collections luxembourgeoises et européennes, couvrant 9 374 versions datées vérifiées**, sur une plateforme temporelle indépendante des éditeurs.
- Conception et exploitation d'un **assistant de recherche juridique ancré dans les sources** : il résout les noms professionnels des textes et les intentions ambiguës, planifie des appels bornés de recherche, d'état à date, de chronologie et de comparaison, puis demande une clarification plutôt que de sélectionner silencieusement un texte. Un registre de preuves et un juge conditionnel vérifient les affirmations du modèle avant l'affichage des citations et liens profonds.
- Publication d'un **serveur MCP public** exposant dix outils en lecture seule pour la recherche, l'état à date, l'historique, la comparaison, la provenance et la couverture.
- Soumission de la recherche hybride locale à des seuils publics de pertinence et de latence sur jeu de réserve ; le benchmark mesuré a rejeté la recherche sémantique candidate, la recherche lexicale restant donc le choix de production. Publication des preuves, alternatives rejetées et critères de réversibilité.

C# / .NET 10 · Assistant juridique IA · Orchestration d'outils · Juge de preuves · Recherche temporelle · Artefacts signés · Benchmarks publics

energy.soufien.lu · Co-Pilote de la Transition Énergétique du Luxembourg 2025 · en ligne

Prospection photovoltaïque ancrée combinant analyse géospatiale, services déterministes et assistant IA en production.

- Conception et exploitation d'un **assistant IA en production** sous la forme d'un agent borné sur cinq outils typés : NL-to-SQL, RAG documentaire, recherche juridique, calculs tarifaires et découverte d'appels d'offres. Le modèle planifie et explique ; les services déterministes restent responsables des calculs, de l'accès aux données en lecture seule et des actions affichées.
- Conception de l'orchestrateur comme hôte de sûreté : il impose les limites d'exécution et les contrats UI typés, tandis qu'un juge d'ancrage factuel confronte la prose du modèle aux documents et lignes réellement renvoyés. Les cartes de résultats directes restent déterministes et ne sont pas réécrites par le modèle.
- Mise en oeuvre d'un RAG hybride avec PostgreSQL, pgvector, recherche plein texte et fusion de rangs. Utilisation d'ADR et d'évaluations pour différer Azure AI Search et écarter un modèle premium plus coûteux sans gain de pertinence ni d'ancrage factuel.

Azure OpenAI · Assistant IA en production · Orchestration d'outils · Juge d'ancrage · RAG · pgvector / FTS / RRF · NL-to-SQL · PostGIS · Évaluations · Terraform / OIDC

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

Intégration LLM · IA générative & agentique · Planification Azure AI · Extraction d'informations · Analyse de texte · Vision par ordinateur

Obtenue en juil. 2026 · examen AI-103 · vérifier sur Microsoft Learn ↗

En préparation : TOGAF Enterprise Architecture.

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