Ibrahim Krayem

INGÉNIEUR IA GÉNÉRATIVE | LLM, RAG, AGENTS | DOCTEUR EN INFORMATIQUE

ibrahim.krayem@outlook.com +33 7 67 07 09 25 Le Plessis-Robinson, France

Google Scholar
linkedin.com/in/ibrahim-krayem
ikrayem.com

PROFIL

Ingénieur IA générative, docteur en informatique (Université de Rennes) et ingénieur diplômé (ENSTA Bretagne). Plus de 6 ans de R&D et d'ingénierie, dont 2 ans chez TALAN en mission chez Bpifrance : un agent financier conversationnel (LangGraph, Claude, MCP), ainsi qu'une extraction documentaire et une recherche sémantique en production. Ma formation de chercheur m'apporte une rigueur de modélisation et d'évaluation que j'applique à des systèmes d'IA traçables et observables.

EXPÉRIENCE

Développeur confirmé en IA générative
Oct. 2024 — Aujourd'hui

TALAN | Paris, France

- **SVI** (Bpifrance) — développeur IA générative d'une équipe agile de six : agent financier conversationnel construit avec LangGraph et Claude, avec gestion de l'historique et du contexte ; mise en production prévue très prochainement.
- Relie l'agent aux données financières via un serveur MCP et développe ses outils : analyse SWOT avec export PowerPoint, visualisations, analyse de documents financiers (dont les tables de capitalisation).
- Comparé des modèles open source (Qwen) aux modèles Claude pour l'extraction documentaire avec Docling ; intégré le streaming temps réel (SSE) dans le frontend Angular.
- **PICXEL** (Bpifrance, organisation SAFe d'environ 50 développeurs) — développé un outil d'analyse documentaire qui extrait les informations clés selon le type de document (signatures, dates, vérification SIREN/SIRET…), en production.
- Refait la barre de recherche PICXEL en recherche sémantique par embeddings, à la place d'une recherche stricte, en production.
- Autres projets TALAN : systèmes RAG (partie retrieval) et recherche vectorielle avec FAISS.

Ingénieur recherche et développement
Oct. 2023 — Mai 2024

IRISA | TARAN | Lannion, France

- Développé une méthode qui accélère l'analyse de performance des réseaux sur puce par fenêtrage de traces d'applications réelles (découpage et fusion de fenêtres, modélisation de Poisson).
- Amélioré la précision des prévisions de performance tout en réduisant la complexité de calcul, validé sur plusieurs types de trafic.
- Mis en œuvre la chaîne de traitement des traces en Python (Pandas, SciPy).

Doctorant en recherche et développement
Oct. 2020 — Sept. 2023

IRISA | TARAN | Lannion, France

- Thèse : « Méthodes pour l'exploration rapide d'architectures multicœurs basées sur les réseaux sur puce avec des technologies émergentes » (Université de Rennes, soutenue en février 2024).
- Conçu un modèle fondé sur la théorie des files d'attente qui évalue la latence de réseaux sur puce hybrides, électriques et sans fil : jusqu'à 500 fois plus rapide que le simulateur de référence Noxim, avec une erreur inférieure à 5 %.
- Présenté les travaux lors de conférences nationales et internationales.

ENSEIGNEMENT

Enseignant vacataire — Réseaux et virtualisation
Depuis 2026

UPEC | EPISEN | Paris, France

- Dispense de cours, TD et TP en réseaux et virtualisation (VMware vSphere).

Enseignant vacataire — Microcontrôleurs
2022 — 2024

ENSSAT | Université de Rennes | Lannion, France

- Encadrement de TD et TP en microcontrôleurs ainsi que d'un projet d'interfaçage.

PROJETS PERSONNELS

- **Aqarih** (aqarih.com) — plateforme immobilière bilingue avec assistant de recherche IA.
- **CalAL** (calal.ikrayem.com) — app Android et web pour ateliers d'aluminium : calculs, découpe des panneaux, stock et factures ; en production.
- **NewsAI** — digests d'actualité en arabe rédigés par LLM et publiés sur Telegram et WhatsApp.
- **Serveur personnel** — serveur Ubuntu auto-hébergé pour CalAL, NewsAI et ce portfolio : Caddy, PostgreSQL, déploiements scriptés, sauvegardes chiffrées.

COMPÉTENCES

IA GÉNÉRATIVE
LLM | RAG | agents LLM
MCP | appels d'outils
Claude | API OpenAI | Qwen
LangGraph | LangChain
Observabilité : Langfuse
Embeddings | Milvus | FAISS
Extraction documentaire : Docling

INGÉNIERIE LOGICIELLE
Python
TypeScript | React | Next.js
Angular | Java
API REST | SSE
PostgreSQL | MongoDB | SQL

CLOUD ET LIVRAISON
Azure | AWS | Vercel
CI/CD | Jenkins
Datadog
Git | GitHub | GitLab
Agile | Scrum

RECHERCHE
Modélisation des performances
Théorie des files d'attente
Pandas | SciPy
Rédaction scientifique et enseignement

FORMATION

Fév. 2024
Doctorat en informatique
Université de Rennes

Sept. 2020
Diplôme d'ingénieur en informatique et télécommunications
ENSTA Bretagne

Août 2017
Licence en génie des réseaux informatiques et télécommunications
Université libanaise | Faculté de technologie

CERTIFICATIONS

Microsoft Azure AI-900
Professional Scrum Master I
GitHub Copilot Certification

LANGUES

Français	Courant
Anglais	Professionnel
Arabe	Langue maternelle

ENGAGEMENTS

Membre de l'IEEE
Ordre des Ingénieurs - Liban
Bénévole à la Croix-Rouge française