



TOUATI ABDELATIF

📍 Lyon, France
@ abdel.touati@gmail.com
☎ 06 51 38 55 04
🌐 2501project.dev
🌐 [LinkedIn](#)
🌐 [GitHub](#)

AI ENGINEER - IA GÉNÉRATIVE & RAG

RAG hybride · Agents métier · Évaluation · Déploiement on-premise

Compétences

- **LLM & GenAI** : Ollama, vLLM, Qwen, Mistral, API OpenAI et Anthropic, sorties structurées (Pydantic, JSON Schema), benchmark de modèles sous contrainte CPU.
- **RAG & recherche** : hybride BM25 + vectoriel (RRF), chunking structurel, citation des sources, Qdrant, Elasticsearch, pgvector.
- **Agents** : routage déterministe puis LLM, agents métier connectés aux API ; LangGraph, LangChain.
- **Évaluation & sécurité** : golden sets, Precision@K, MRR, NDCG, LLM-as-judge ; anti-injection, contrôle d'accès.
- **Back-end** : Python, FastAPI, Celery, Redis, SQL, streaming SSE, intégration Symfony.
- **Qualité logicielle** : pytest, mypy, ruff, pre-commit, revue de code.
- **MLOps & DevOps** : Docker, GitHub Actions, Prometheus, Grafana, Helm, on-premise.
- **Data & ML** : dbt, Prefect, MongoDB, DuckDB, scikit-learn.

Formation

- **AI Engineer - Chef de projet en IA**
RNCP 36129 niveau 7 (bac+5)
ASCENCIA Business School · 2025
- **Data Analyst** - RNCP niveau 6
Wild Code School · 2024
- **Master 2 Aménagement du territoire, Licences de Géographie et d'Histoire**
Université Lumière Lyon 2

Méthodes & conformités

- Scrum : sprints, coordination dev / produit
- Cadrage des besoins avec les équipes métier
- Conformité RGPD et accessibilité RGAA
- Souveraineté des données, déploiement on-premise
- Vulgarisation et formation à l'IA

Langues

- Français
- Anglais - B2 (technique courant)
- Espagnol - B1
- Arabe - parlé

Expériences

AI Engineer (freelance) - DIR'AJ, plateforme de gestion de dossiers juridiques

Février – septembre 2026 · Lyon · équipe de 5 (tech lead, 2 développeurs, 2 chefs de projet)

Conception et mise en production d'Eidelia, l'assistant IA de la plateforme, 100 % on-premise sur serveur CPU (16 cœurs, sans GPU).

- **Architecture** : passerelle FastAPI intégrée au back-end Symfony ; Ollama (Qwen), recherche hybride Qdrant + Elasticsearch, Redis, Celery, streaming SSE.
- **Agents métier** : pré-routeur déterministe puis routage LLM vers des agents documentaire, statistiques multi-tenant et assistant.
- **Agent documentaire** : pipeline asynchrone avec reprise sur incident, chunking structurel, synthèse map-reduce en sorties typées et sourcées à la page.
- **Évaluation** : golden set de 30 questions et benchmark de 3 LLM pour le choix du modèle ; le pré-routeur déterministe choisit le bon agent dans 85 % des cas, contre 52 % avec le LLM seul.
- **Qualité et sécurité** : garde-fous anti-injection, contrôle d'accès par rôle, circuit breaker ; monitoring Prometheus / Grafana.

Bon agent choisi : **de 52 % à 85 %**

Golden set **77 % à 93 %**

1 000+ tests, **87 %** de couverture

RAGnarok-AI - framework open source d'évaluation RAG · créateur et mainteneur

Depuis janvier 2026 · github.com/2501Project/RAGnarok-AI · `pip install ragnarok-ai`

- Évaluation 100 % locale de pipelines RAG et d'agents : métriques de retrieval et LLM-as-judge (Prometheus 2) calibré sur annotations humaines.
- Industrialisation : CLI, seuils de qualité bloquants en CI/CD, monitoring de production, détection de dérive, chart Helm ; plus de 20 adaptateurs (Qdrant, pgvector, LangChain, LlamaIndex, Ollama, OpenAI).
- Qualité logicielle : pytest, mypy, ruff, pre-commit, publication PyPI signée.

v1.11 · 20 versions sur PyPI

AGPL-3.0 / commerciale

Contributeur : chercheur de l'**U. d'Ottawa**

Intervenant Data Science & ML - Sup de Pub (OMNES Education)

Octobre 2026 · Lyon

- Module de 14 h conçu de zéro et animé pour une quinzaine d'étudiants de M2 Data Marketing & IA non développeurs : arbres de décision, évaluation, segmentation sur Google Colab.

Teaching Assistant Data - Wild Code School

2024 – 2025 · Lyon

- Accompagnement de promotions Data (live coding, suivi de projets, cours et examens blancs en Python, SQL et ML)
- Création d'une plateforme de données « Iliad » : 4 sources unifiées, Prefect, dbt, MongoDB/DuckDB.

Projets

- **Plateforme EdTech multi-agents** : 9 agents tuteurs orchestrés avec LangGraph, RAG (BM25, cache Redis), recommandation par bandit contextuel LinUCB.
- **DontREADME** (analyse de documents PDF) · **Agilia** (agent agile, LangChain, PostgreSQL) · **MaiiA** (classification d'e-mails par NLP).