

ABDELATIF DJEDDOU

AI Research Engineer (GenAI / LLMOps)

Paris, France • 06 01 14 44 00

abdelatifdjeddou@gmail.com

linkedin.com/in/abdelatifdjeddou •

github.com/BinkyTwin

QR CODE

« Si ce CV ne vous suffit pas, n'hésitez pas à aller discuter avec mon agent virtuel ! »

EXPÉRIENCE PROFESSIONNELLE

BNP Paribas Personal Finance - GenAI Catalyst

Septembre 2024 — Présent

Cadrage et qualification de plusieurs cas d'usage GenAI avec les métiers (besoins, contraintes data/sécurité, faisabilité technique). Développement de briques réutilisables (RAG, guardrails, parsing) et patterns d'architecture pour accélérer l'industrialisation des futurs POCs.

CIEC - Chargé Outils et Méthodes (IA/BI)

Sept. 2023 - Août 2024

Conçu et livré un pipeline de traitement automatisé de rapports (OCR, extraction, normalisation et classement automatique), réduisant d'environ 50% le temps de traitement opérationnel. Mise en place de contrôles qualité et de traçabilité des données (SAP/Coupa).

PROJETS DE RECHERCHE & D'IMPLÉMENTATIONS

Preprint en cours : Local Generative AI on Edge Devices in Regulated Banking Environments

Stack: MLX, GGUF, NIST AI RMF, Python, Docker

Proposition d'un framework d'évaluation et benchmarking systématique (inspiré de MT-Bench) sur tâches bancaires : classification d'intents (**Banking77**) et analyse de sentiment (**Financial PhraseBank**). Étude comparative MLX vs GGUF.

slm_rlm : Benchmarking Récursif sur SLMs

Stack: Python, Docker, LLM Benchmarking, Reasoning Evaluation

Reproduction scientifique du papier "Recursive Language Models" (2025). Évaluation de la robustesse et des capacités de raisonnement (MMLU-style) via le benchmark **RULER (S-NIAH)**. Analyse du trade-off performance/coût sur modèles compacts.

articlewriter : Orchestration multi-agents

Stack: LLMops, DeepSeek-V3, OpenRouter, Notion API

Pipeline complet de génération de contenu tech avec apprentissage de style via corpus. Gestion de sessions persistantes, streaming et RAG pour extraction d'exemples contextuels.

ReviewXiv : Plateforme IA de Revue Scientifique

Stack: Next.js 14, TypeScript, TailwindCSS, Vercel, LLM Integration

SaaS d'annotation et d'analyse de papers arXiv. Intégration de fonctionnalités IA (traduction contextuelle, résumé automatique, Q&A sur documents) pour accélérer la veille scientifique.

FORMATION ACADÉMIQUE

Ecole Management Leonard deVinci - Master Management des Systèmes d'Informations et des Datas Sept. 2024 — Août 2026

Master Programme Grande École. Mémoire sur l'orchestration multi-agents.

Université Paris 8 - Licence Informatique (En parallèle du Master)

Sept. 2025 — Présent

Sorbonne Paris Nord - BUT Gestion des Entreprises et des Administrations

Sept. 2021 — Sept. 2024

COMPÉTENCES TECHNIQUES

AI Architecture: LLMops, Orchestration multi-agents, RAG, SLMs, Prompt Engineering.

ML/Data: Scikit-learn, Vector Stores (FAISS/Pinecone), Computer Vision (CLIP/OpenCV).

AI-Native Dev: Claude Code, Codex, Cursor, Agentic Workflows, Automated Refactoring.

Recherche: Veille SOTA, Protocoles Expérimentaux, Synthèse Scientifique.

CERTIFICATIONS

Professional Scrum Master I (PSM I)