

# JULIEN RABAULT

## Applied AI / ML Engineer

+33 7 81 16 46 29  
julienrabault@icloud.com  
linkedin.com/in/julienrabault  
github.com/JulienRabault  
Toulouse, France  
julienrabault.github.io

4 ans au CNRS a entrainer des modeles deep learning en production (PyTorch, multi-GPU, Jean Zay), 2 publications peer-reviewed. Aujourd'hui chez Berger-Levrault : conception d'Athena, plateforme agentique multi-agents (LangGraph, RAG, MCP).

### COMPETENCES TECHNIQUES

**Agentique & RAG** : RAG / GraphRAG, Multi-agents, MCP Protocol, LangChain / LangGraph, Prompt Engineering, Structured Outputs, Embeddings, LLM Routing, Fine-tuning

**Deep Learning** : PyTorch, Transformers, Computer Vision, Diffusion Models (DDPM), NLP, CNNs / U-Net / YOLOv5, Modeles generatifs (GAN, VAE)

**MLOps & Infra** : Docker, AWS, CI/CD, MLFlow, Airflow, Kubernetes, Celery, Langfuse, HPC / Slurm, Linux

**Developpement** : Python, FastAPI, HuggingFace, Qdrant / pgvector, Mistral / OpenAI API, Git, SOLID / Architecture, C#, SQL

### EXPERIENCES PROFESSIONNELLES

#### AI ENGINEER

Janv. 2026 - present

##### Berger-Levrault | Toulouse | Equipe R&D IA, 12 personnes

Conception et developpement d'Athena, plateforme agentique de Berger-Levrault : multi-agents, routage intelligent, connexion aux documents et APIs metier pour tous les clients (collectivites, industrie, maintenance). ~30 utilisateurs pilotes. Equipe cross-fonctionnelle (designer, dev front, DevOps), ateliers clients, observabilite Langfuse.

##### Architecture multi-agents

Conception de l'architecture multi-agents (LangGraph), routage par type de question, orchestration d'agents RAG et d'agents sur APIs MCP, integration du sourcing des reponses. **Plateforme en production**, reponses sourcees et actionnables sur plusieurs domaines metier.

##### Content-extractor - OCR/PDF/DOCX

Refonte complete du service d'extraction : OCR, images, PDF, DOCX. Batch async (Celery + API batch Mistral), architecture factory/registry pour ajouter facilement de nouveaux formats. **-50% sur les couts d'extraction.**

##### Pipelines Airflow

Reprise et evolution en equipe des pipelines d'ingestion documentaire (PDF, notices, work orders, docs machine). **5 DAGs operationnels.**

##### MCP Builder

Un LLM analyse les specs OpenAPI pour pre-traiter les endpoints (regroupement, masquage, generation de descriptions), puis revue human-in-the-loop pour valider routes, doc et connaissance metier. **120+ APIs internes cartographiees**, integration progressive au runtime.

#### MACHINE LEARNING ENGINEER

Dec. 2021 - Janv. 2026

##### CNRS - Programme National de Recherche en IA (PNRIA) | Toulouse

Accompagnement d'equipes de recherche en France. 2 projets en parallele (6-12 mois), livres a Meteo France, CNES, CEA, INEE. Entrainement et fine-tuning sur Jean Zay (multi-GPU DDP, 8 GPUs, Slurm). Profilage PyTorch.

##### GENS / MetScore - Meteo France

Evaluation de modeles meteo en production, optimisation multi-GPU et fine-tuning d'un modele de diffusion (DDPM) en PyTorch sur Jean Zay. Conception de MetScore (configuration YAML, librairie Python). **Librairie toujours en prod + POC diffusion a -20% de calcul** a qualite equivalente. Co-auteur du papier AMS 2025.

##### DeepFaune - CNRS/INEE

Fine-tuning YOLOv5 sur dataset custom (1,5M images, 24 classes), entrainement multi-GPU, gestion du desequilibre de classes, optimisation precision/vitesse CPU. **93% sur 24 especes, 3x plus rapide.** Publication peer-reviewed.

##### Autres contributions

AUTOFILL (CEA, PairVAE, generation de donnees nanomateriaux, MAE 0,98), BIGSF (CNES, tech lead et refacto d'architecture, librairie d'analyse d'images de filaments galactiques), MORPHOGAN (Univ. Lorraine, refonte code StyleGAN2, pipeline automatisee).

Formation dispensee : *Introduction aux LLMs* (3h, ~25 doctorants/chercheurs CNRS).

#### INGENIEUR LOGICIEL (alternance)

Aout 2020 - Sept. 2021

##### Agileo Automation | Montauban

Framework de supervision de machines robotisees (semi-conducteurs). C#, architecture OO, IHM, CI/CD. Equipe de 5 ingenieurs, Agile/Scrum.

### FORMATIONS

**MASTER - IA ET RECONNAISSANCE DES FORMES (IARF)** | Universite Paul Sabatier Toulouse III - IRIT | Deep Learning, Vision, NLP 2019 - 2021

**LICENCE INFORMATIQUE** | Universite Paul Sabatier Toulouse III 2016 - 2019

**Langues** : Francais (natif) | Anglais professionnel (redaction scientifique, documentation technique)

### PUBLICATIONS

"Enriching Operational High-Resolution Ensemble Forecasts with StyleGAN-2" - AIES, 2025. **3e auteur**, peer-reviewed.

"The DeepFaune initiative: automatic identification of European fauna" - **Julien Rabault** et al. **Co-auteur**, peer-reviewed.

### PROJETS OPEN SOURCE

**LLMock** (PyPI) - Mock server LLM, retries/fallbacks, 10+ providers, OpenAI-compatible. Python, FastAPI.

**DDPM-weather** - Modele de diffusion pour debruitage d'images meteo, -20% ressources. PyTorch, Meteo France.

**BananaML** - Pipeline ML end-to-end sur AWS. Computer Vision + API REST. FastAPI, Docker, CI/CD.