



# Nolan Cacheux

AI/ML Engineer • Enterprise Agents, RAG & MLOps

Portfolio : nolancacheux.com

Github : github.com/nolancacheux

Email : cachnolan@gmail.com

Mobile : (+33) 625-156-261

Available for a contract from September 2026

Freelance missions anytime

## PROFESSIONAL EXPERIENCE

### Decathlon France

#### AI & ML Engineer (Work-Study)

Lille, France

Sept 2025 - Present

- Built **DAISI** (Decathlon AI Supplier Informations), a **production** Google Chat **Enterprise Agent** for supplier-process questions. Used LangChain/LangGraph, FastAPI, Vertex AI, Vector Database and Search, Cloud Run, Cloud SQL, MLflow/Databricks, Model Armor, DLP, GDPR safeguards, and Terraform. **Impact: 13,000 hours/year** saved and successful load test with **1000 concurrent users**.
- Led **OpsBot**, an **Enterprise Agent** for maintenance, safety, and compliance workflows. Built governed-source RAG with source-backed answers, safe fallback, traces, evaluation, Vertex AI RAG Engine, GDPR safeguards, and an internal agent library I built to standardize agents, tools, observability, and deployment. **Impact/ ROI: 85% first-line resolution target, 14,500+ h/year** reallocated, **~€540k productivity reallocation**.
- Led **Summarizer**, a Gemini-powered summarization platform orchestrated with Airflow/MWAA, AWS EKS/S3/ECR, Databricks Delta refresh, runtime validation, and operator runbooks.
- Led **internal AI/Tech innovation** through a weekly newsletter, onboarding guide, documentation/runbooks, **30+ hours** of GenAI/AI/Tech trainings and internal **hackathons**.

### Decathlon Belgium

#### Data Scientist - MLOps (Internship)

Brussels, Belgium

May 2025 - Aug 2025

- Led a strategic Data Science project to **industrialize** the prediction of **8 key sales KPIs** (GMV, items sold) segmented by channel (InStore/OutStore, 1P/3P) for **64 sports** categories.
- Built an automated forecasting pipeline using XGBoost, Apache Spark, PySpark, Databricks, AWS S3, Airflow/MWAA, and MLflow.
- Achieved **+15% forecast accuracy improvement (MAPE)** vs. previous manual process through rigorous model comparison (Prophet, XGBoost, LightGBM, Chronos-Bolt), feature engineering (weather, holidays, lag features), parallel processing with joblib, and deployment to MLflow Model Registry for production inference. Automated exports to Google Sheets/Tableau for business stakeholders.

### Beobank NV/SA

#### Data Analyst/Scientist - Banking Domain (Internship)

Brussels, Belgium

May 2023 - Aug 2023

- Processed banking transaction data with Python, VerticaPy, NumPy, Pandas, SQL, JupyterLab.
- Improved** transaction categorization **success rate from 63% to 84%** by identifying new categories, adding business rules, and iterating on classification quality. Supported income/expense identification and reusable reporting for digital banking analytics.

## FREESTYLE FOOTBALL CHAMPION

nolanfree.fr

**Professional athlete and President** of Nolan-Free Association since 2019 : performances, workshops & media appearances (M6, France 3).

- 30K+ followers** and **35M+ views** on social networks.
- Junior World Champion** (Prague, 2019), **Top 15 World** (2026), **Top 5 Europe** (2024), **2x Vice-Champion of France** (2024, 2025)
- 300+ performances**: PSG, LOSC Lille, Valenciennes FC, Puma, Umbro, Ferrero
- Sponsored** by Umbro for multiple years.

## PERSONAL PROJECTS

### Mistral AI Hackathon London 2026 - Ecotopia

- Selected among **7,000+** applicants. Organized by Mistral AI, sponsored by W&B, NVIDIA, AWS, ElevenLabs, and Hugging Face. Built an interactive political simulation powered by fine-tuned SLMs. 4 models via QLoRA (NF4 4-bit), <10 min/model. Our fine-tuned 8B Mistral model **outperforms** Mistral Large on structured output at **10x lower latency**. Used Spring Boot, Spring AI, TypeScript, Weights & Biases MCP, Hugging Face.

### RAG Equity Research Agent

- Production-deployed multi-agent** system for automated equity research. Scalable RAG pipeline on SEC filings with semantic **reranking** and Telegram bot interface. Deployed on Azure with LangGraph, LangChain, Qdrant, Azure OpenAI, FastAPI, Pydantic, Terraform, Docker, and Pytest.

### AI Product Photo Detector

- Production-ready MLOps** pipeline using PyTorch and EfficientNet-B0 transfer learning, FastAPI, MLflow, Vertex AI Pipelines (KFP), DVC, Prometheus/Grafana, Docker Compose, and GitHub Actions CI/CD.

### AI Video Comment Analyzer

- Full-stack ML application for automated **BERT** sentiment classification, BERTopic **topic modeling**, and LLM-powered summaries. Built with Transformers, Ollama, Next.js, React, FastAPI, SQLAlchemy, Tailwind CSS, and Pytest.

## EDUCATION

- JUNIA ISEN - Lille** (2021 - 2026)
- Master's in Engineering (Diplôme d'Ingénieur), Computer Science** (2023 - 2026) · Specialization in Data Science & Machine Learning.  
*Data Structures & Algorithms, Machine Learning, Deep Learning, Distributed Systems, Database Management (SQL & NoSQL), Operations Research, Computer Networks, Big Data (Hadoop, Spark), Cloud Computing, DevOps, Computer Vision, NLP, Software Architecture*
- Preparatory Classes · Computer Science & Networks** (2021 - 2023) · Ranked 2nd/96 (Year 1) • 8th/76 (Year 2) · Highest Honors
- French:** Native | **English:** Fluent

## TECHNICAL SKILLS

- AI/ML Engineering:** LangChain, LangGraph, HuggingFace, RAG, Vector Databases (Vertex AI, FAISS, Chroma DB, Qdrant, Pinecone, Weaviate), Finetuning, Python, PyTorch, TensorFlow, scikit-learn, XGBoost, LightGBM, Prophet, Forecast, BERTopic, Pandas, NumPy, MLOps, MLflow, W&B
- Cloud Platforms:** GCP (Cloud Run, Cloud SQL, Cloud Storage, Cloud Scheduler, IAM, VPC networks, Vertex AI), AWS (Airflow/MWAA, EKS, S3, ECR, SageMaker, Bedrock), Azure (Container Apps, Azure OpenAI, App Service, Storage, Key Vault)
- Big Data & Data Engineering:** Apache Spark / PySpark, Databricks, Delta Lake, SQL, PostgreSQL, MongoDB, Redis, Neo4j, Cassandra
- Infrastructure & DevOps:** FastAPI, Pydantic, Docker, Terraform, GitHub Actions (CI/CD), SonarCloud
- Teamwork:** Git, GitHub, Jira, Confluence, Agile/Scrum