MOUHAMADOU LAMINE SOW

lamine.2004pro.sow@gmail.com | Paris, France
LinkedIn | github.com/Lamine2004Sow

AI Engineer – 6-month end-of-studies internship from March 2027

Final-year engineering student: **generative AI** (LLMs, RAG, agents), **machine learning**, **optimization** and **deployment to production**, with models evaluated before they are used.

EDUCATION

Engineering degree in Computer Science (Master's level) Sept. 2024 – 2027
Sup Galilée, Université Sorbonne Paris Nord *Villetaneuse, France*
Main courses: Machine Learning, Deep Learning, NLP, Text Mining, Operations Research

Integrated preparatory program (MPSI) Sept. 2022 – May 2024
ESIGELEC – ESMT *Dakar, Senegal*

EXPERIENCE

Machine Learning / Data Science Intern May 2026 – Jul. 2026
Sonatel (Orange group) *Dakar, Senegal*

- Developed an XGBoost model predicting fiber network risks 6 months ahead: **88 % fewer false negatives** than the existing business rule (PR-AUC 0.81 vs. 0.28).
- Calibrated and **explained (SHAP)** predictions, aggregated into risk maps by area (3 to 12 months) to help field teams prioritize their interventions.

Observation internship – Digital and DevOps Jul. 2024 – Sept. 2024
Sonatel (Orange group) *Dakar, Senegal*
Wrote a state-of-the-art review on AI-driven DevOps automation; introduction to Angular, LoopBack and MongoDB.

TECHNICAL PROJECTS

Legal LLM agent: guardrails, verified citations and automated evaluation 2026
Python, LangGraph, LLM (DeepSeek), FAISS, BM25, e5 embeddings, RAGAS, Streamlit, pytest, Docker

- **End-to-end RAG** over about 5,100 French law articles: ingestion (cleaning, deduplication, chunking), hybrid FAISS + BM25 retrieval (RRF fusion).
- LangGraph agent: query rewriting, a **guardrail** that treats the question as data (100 % of the 10 off-topic questions detected, including one **prompt injection** attempt), **deterministic citation checking** (92.3 % precision).
- **Golden set** of 60 labeled questions: deterministic metrics, LLM judge, RAGAS; a **regression test** fails if a prompt changes without re-running the evaluation. Retrieval 80 % → 90 %, 100 % correct refusals.

Self-healing MLOps platform on local Kubernetes 2026
Python, FastAPI, MLflow, Docker, Kubernetes (kind), Prometheus, Grafana, Evidently, GitHub Actions

- **FastAPI backend** serving the model on local Kubernetes; GitHub Actions **CI/CD** that fails if the model does not beat its baseline; Docker Compose.
- Prometheus and Grafana **observability**: degradation alert → automatic retraining, new model promoted only if it is 2 % better on unseen data (MAE 4.49 → 3.09 during the COVID shock).

Fine-tuning a legal embedding model (in progress) 2026
Python, LLM (DeepSeek), sentence-transformers, multilingual-e5-base, PyTorch, MLflow

- 11,088 question/article pairs **generated by an LLM** then filtered (rules and LLM judge) out of 12,849; 125-question test set frozen before training; CPU fine-tuning, MLflow tracking.

ROADEF/EURO 2020 challenge: maintenance planning of the power grid (RTE) 2026
Python, Gurobi, MILP, simulated annealing

- **Optimization under uncertainty**: simulated annealing within 0.52 % of the best known solution; Gurobi MILP: 8 optimal solutions proven out of 15.

SKILLS

Generative AI: LLMs, RAG, LangGraph, embeddings (e5, sentence-transformers), FAISS, BM25, fine-tuning, LLM-as-a-judge, RAGAS
ML and optimization: PyTorch, scikit-learn, XGBoost, SHAP, MLflow; MILP (Gurobi), metaheuristics
Development: Python, FastAPI, Git, CI/CD (GitHub Actions), Docker, Kubernetes, Grafana, SQL; Claude Code, Codex, Antigravity
Languages: French (native), English (B2, TOEIC 802) | **Certification:** IBM AI Engineering (Coursera, in progress)