

BAIM Mohamed Jalal

+33 765294231 — mohamedjalalbaim@gmail.com — in/mohamedjalalbaim/ — github.com/Jalalbaim — [Portfolio](#)

Experience

BNP Paribas

Mars. 2026 - Present

AI Research Intern

Paris

- Developing defenses against data poisoning attacks using causality-based methods to improve model robustness in production ML systems.
- Evaluating privacy leakage risk through membership inference attacks on fine-tuned LLMs, informing internal compliance and security guidelines.

CEA LIST / DILS - In-painting using Generative AI for eXplainable AI Methods Evaluation

Feb 2025 - Aug. 2025

AI Research Intern

Grenoble

- Designed a generative AI inpainting pipeline to benchmark XAI methods, providing a quantitative measure of saliency map correctness.
- Introduced a novel approach to assess the distribution of adversarial attacks in latent space.

L'Oréal Paris, Research & Innovation

June. 2024 - Aug. 2024

R&I Intern

Aulnay-sous-Bois

- Developed a classification model for chemical entities using SMILES notation, using RNN/LSTM architectures.
- Database management including data collection, organization, cleaning, and analysis.

Hubert-Curien Laboratory

Oct. 2023 - May. 2024

AI Research Intern

Saint-Étienne

- Explored a new approach for generating emotional facial expressions using latent space diffusion models to enhance emotion recognition performance.

Education

MSc, École Polytechnique — Data Science, Applied Mathematics & Statistics

Sep. 2025 – Oct. 2026

Relevant Courses: Reinforcement learning, Deep Learning, eXplainable AI, Generative Models, Statistical Learning.

National Yang Ming Chiao Tung University, Taiwan

Sep. 2024 – Jan. 2025

Exchange Semester in Artificial Intelligence. GPA: 4.24/4.3

MEng, Télécom Saint-Étienne — Computer Science

Sep. 2022 – Aug. 2025

Engineering Degree in Computer Science. GPA: 4/4

Projects

Capstone Project: L'Oreal's Protection market through predictive detection.

Jan 2026 – Mars 2026

- Built a two-stage market diversion detection pipeline, combining anomaly detection and order classification.
- Calibrated classification thresholds using a financial confusion matrix, weighing the business cost to align model decisions with operational investigator capacity.
- Delivered explainable model outputs, enabling investigators to audit flagged orders with interpretable risk drivers.

Trust RAG Orchestrator

Sep 2025 – Jan 2026

- Designed and implemented a hybrid RAG pipeline combining dense vector search and BM25 sparse retrieval with a reranking stage to optimize context quality fed to LLMs for downstream generation.
- Evaluated retrieval and generation quality using NDCG metrics and the RAGAS evaluation framework.
- Fine-tuned the underlying LLM using Reinforcement Learning from Human Feedback (RLHF - DPO), aligning model outputs with user preferences and improving response quality.

Skills

Languages: Python, C++, SQL, Java, Matlab Simulink, JavaScript

Pandas

Tools: Git, Power BI, Linux, Docker, Talend, Latex

Frameworks: PyTorch, Scikit-Learn, TensorFlow, Keras,

Languages English (C1), French (Native), Arabic (Native)