FLORENT LIN

 Paris, France +33 7 81 63 66 63 1florentlin@gmail.com

 github.com/fllin1 linkedin.com/in/florentlin

EDUCATION

ENSAE Paris - IP Paris   Palaiseau, France
Diplôme d'Ingénieur (MSc equivalent) in Data Science & Applied Math   *Sep 2021 - Jun 2024*

· **Focus:** Advanced Probability, Bayesian Stats, Optimal Transport, Time Series.
· **AI Specialization:** Deep Learning, NLP, Online Learning, High-dimensional Stats.

Lycée Saint-Louis   Paris, France
CPGE - PC (Intensive preparation in Math & Physics)   *Sep 2019 - Jun 2021*

PROFESSIONAL EXPERIENCE

Aubay Data & AI (Client: SNCF Réseau)   Boulogne-Billancourt
Data Scientist Consultant - Predictive Maintenance   *Jan 2025 - Present*

· **ML Production:** Deployed predictive models for railway defects (XGBoost/Random Forest).
· **Engineering:** Built robust ETL pipelines for multi-source geospatial & temporal data.
· **Optimization:** Tuned models for Recall/F1-Score to prioritize safety risks; SHAP for interpretability.
· **Decision Support:** Designed interactive dashboards and translated complex indicators into actionable KPIs for operational stakeholders.

Aubay Data & AI   Boulogne-Billancourt
R&D Data Scientist Intern - LLM Machine Unlearning   *Aug 2024 - Dec 2024*

· **Research:** Conducted SOTA review on Machine Unlearning (Concept Erasure, DPO).
· **Implementation:** Fine-tuned **LLaMA** (LoRA) and trained RoBERTa classifiers for knowledge scrubbing.
· **Metrics:** Benchmarked efficacy vs. degradation using custom metrics (Truth Ratio, Conditional Probabilities).
· **Innovation:** Devised cross-lingual techniques to prevent knowledge leakage across languages.

Deloitte Conseil AIM (via ENSAE)   Paris
Operations Research Data Scientist   *Nov 2022 - May 2023*

· **Modeling:** Formalized nurse scheduling constraints into Mixed-Integer Programming (MIP).
· **Solvers:** Developed constraint programming algorithms for multi-objective routing.

SELECTED PROJECTS

Multi-modal Document Analysis (RAG)   *link*
Client: MEWE Architects   *Apr 2025 - Present*

· Built production RAG workflow (**Mistral OCR** + **Gemini**) for thousands of complex regulatory PDFs.
· Engineered context-aware injection strategies to reconstruct layouts, boosting retrieval accuracy.
· Implemented Pydantic for deterministic structured outputs from vision inputs.

Lead-to-Franchise Matching Engine   *link*
Client: Franserve Consultants   *Jul 2025 - Present*

· Architected semantic search system using embeddings to match leads to inventory.
· Built scraper-to-DB pipeline (Supabase) integrating CRM logs (GoHighLevel) for real-time enrichment.

Academic Research Projects
ENSAE   *2023 - 2024*

· **Multi-Armed Bandits:** Benchmarked UCB/Thompson Sampling in online learning settings.
· **Bayesian Stats:** Applied Gibbs Sampling and Bayesian Lasso for high-dimensional variable selection.

TECHNICAL STACK

Deep Learning	Python, **PyTorch**, **JAX**, HuggingFace, Scikit-learn
Engineering	SQL (PostgreSQL), AWS, Docker, Git, Firebase, Supabase
Dev Tools	TypeScript, Golang, C, R, Cursor (AI-assisted dev)
Languages	French (Native), English (C1+), Chinese (B2)

CONTINUOUS LEARNING

Backend Engineering (Python & TypeScript)   🌐 *link*
Boot.dev   *Sep 2025 - Present*

· **Core CS:** Advanced Complexity Analysis (Big-O), Memory Management, OOP and Dynamic Programming.
· **Systems:** Scalable HTTP Servers, Pub/Sub Architectures, Docker Containerization.
· **RAG from Scratch:** End-to-end pipeline implementation without frameworks.
· **Search Algorithms:** Built TF-IDF, BM25, and Hybrid Search (Vector + Keyword) systems.