

Florian Bastin

LLM Engineer · Generative AI Lecturer · Freelance · 7+ years

bastinflorian1@gmail.com

+33 6 37 51 96 84

github.com/BastinFlorian



Freelance LLM Engineer and Generative AI Lecturer at Paris Dauphine-PSL, with a background in mathematics (ENS-PSL, Dauphine, ISI Kolkata) and 7+ years in applied AI, from data engineering and NLP to production-grade LLM systems. I operate with full-stack autonomy across Azure and GCP, delivering measurable ROI through RAG pipelines, agentic architectures, and structured LLM evaluation. Beyond production work, I am increasingly drawn to the deeper mechanics of how models learn: **RL-based fine-tuning, specialised LLM training**, and the intersection of language models with **robotics and embodied reasoning**.

EXPERIENCE

Université Paris Dauphine-PSL

2024 – Present

Lecturer · Generative AI & AI in the Cloud · M2 IASD

Designed and teach full-semester Gen AI course: prompting, Transformer architecture, RAG, agentic systems, LLM evaluation

Taught AI in the Cloud (GCP) in 2024, establishing Google Cloud partnership for free student compute resources

bastinflorian.github.io/teaching · [2025-2026 student project](#)

SUEZ via Octo Technology / Accenture

Mar 2024 – Present

LLM Engineer · Azure · 60% production / 40% POC

Full-stack autonomy: Terraform & Azure infrastructure to frontend delivery and LLM monitoring

30+ Gen AI POCs across HR, Research, Legal and Marketing: ReAct chatbots, document summarizers, NDA comparators, information retrievers

RAG pipelines: reranking, self-query retrieval, evaluators, async API · LangChain / LangGraph

Internal LangChain wrapper + LLM library (LangFuse, RAGAS, DeepEval) to standardize development

LLM & Embedding QLoRA fine-tuning on Azure to adapt models to internal business domains

5 use cases promoted to production: scalability, feedback loops, frontend refactoring, ownership attribution

LLM APIM layer for token governance, data privacy and internal access management

Delivered 2-day Gen AI training: 5 sessions, 80+ employees

Pernod Ricard via Octo Technology

Nov 2023 – Mar 2024

LLM Engineer · Azure

Production RAG chatbot on Microsoft Teams · GPT-3.5 Turbo, Ada 2, Azure App Service, PostgreSQL

LangChain: reranker, merge retriever, prompt engineering · Hexagonal architecture · **-45% Level 1 IT tickets**

Chanel via Octo Technology

Jun – Nov 2023

Deep Learning Engineer · Azure

Gesture recognition model: MediaPipe + TensorFlow, iOS & Android · TFLite & Core ML edge deployment

Casino via Octo Technology

Sep 2022 – Jan 2023

Data Engineer · GCP

Migrated large-scale goods pipeline Airflow → Dataflow · Apache Beam parallelism & OLTP SQL optimization

Club Med via Octo Technology

Mar – Sep 2022

Tech Lead Data · GCP

Customer Lifetime Value, Lead Scoring models · Agile best practices

Le Monde via Octo Technology

Nov 2018 – Mar 2022

Tech Lead Data · GCP

7 production solutions: recommender systems (statistical, content-based, user-based), churn & propensity models, NLP/NER news detection

End-to-end GCP architecture still in production: PubSub, Dataflow, BigQuery, Airflow, Vertex AI, Looker Studio

Lili.ai

Jan – Nov 2018

NLP Research Scientist

Unsupervised NLP for delay detection in industrial projects: LDA, HDP, DBSCAN, TF-IDF, LSTM/QRNN

Defense Research Internship

2017

Deep Learning Intern

7-layer CNN (MatConvNet) for tank detection in aerial imagery · 3 conv + 2 pooling layers

TALKS & PUBLICATIONS

CONFERENCE TALK

Jun 2023

How to Balance Power and Control When Using Dataflow with an OLTP SQL Database

Apache Beam Summit 2023 · New York

BLOG POST

Feb 2024

Comment utiliser le multimodal pour améliorer un chatbot RAG ?

Octo Technology Blog · Text, image & table extraction to enrich RAG chatbot performance

BLOG POST

Oct 2023

Construire son RAG grâce à LangChain : L'exemple de l'Helpdesk d'OCTO

Octo Technology Blog · End-to-end RAG pipeline with LangChain, Chroma & GPT

ML / DEEP LEARNING

PyTorch TensorFlow Keras Scikit-learn MediaPipe TFLite

CLOUD & INFRA

Azure OpenAI GCP Vertex AI Terraform Docker CI/CD
BigQuery Dataflow

CODE

Python SQL React R C++

EDUCATION

M2 MASH · AI & Statistics

Dauphine & École Normale Supérieure · 2018

🏆 1st / 36 · ENS Challenge Data

Exchange · Statistics

Indian Statistical Institute, Kolkata · 2017

BSc & M1 · Applied Maths

Université Paris Dauphine · 2013–2017 · 22nd / 150

CERTIFICATIONS

- ✓ GCP Professional ML Engineer
- ✓ GCP Professional Data Engineer

LANGUAGES

French Native
English Fluent
Spanish Advanced

AWARDS

GCP Gen AI Hackathon ×2

1st · Superior GenAI Business Value · 2024

ENS Challenge Data

1st / 36 · NBA challenge

Le Monde Hackathon

2nd place · 2024