

Baptiste Blouin

Aix-en-Provence, France
✉ BaptisteBlouin@proton.me
🌐 baptisteblouin.fr
in [baptiste-blouin](https://www.linkedin.com/in/baptiste-blouin)
🔊 [BaptisteBlouin](https://www.youtube.com/channel/UCBaptisteBlouin)
Français, Anglais

Data Scientist & Ingénieur IA | NLP, LLM & Full-Stack

Résumé Professionnel

Data Scientist et Ingénieur IA avec un Doctorat en Informatique, je combine une expertise en recherche ML/NLP publiée avec une solide pratique d'ingénierie de produits IA en production. De l'architecture à la mise en production, je conçois et réalise des produits IA de bout en bout, en autonomie. Ma plateforme SaaS l'illustre : un système LLM réellement en production, multi-tenant, où j'ai autant soigné la valeur métier que la robustesse d'exploitation — maîtrise des coûts LLM et sécurité de bout en bout. Ma formation par la recherche (7+ publications, conférences internationales, subvention ERC) me donne une rigueur méthodologique que j'applique aussi bien à la conception d'architectures techniques qu'à la résolution de problèmes complexes.

Projets récents

Plateforme SaaS de Transformation de Données par IA

2025–Présent

Conception & Développement Full-Stack

Projet personnel — aubier-data.fr

- **Plateforme SaaS multi-tenant** de bout en bout : extraction documentaire, RAG, text-to-SQL, analyse et prédictions
- **Moteur RAG avancé** (recherche hybride pgvector + BM25, GraphRAG, RAPTOR) et **text-to-SQL** avec streaming SSE
- **Sécurité complète** : JWT RS256, RBAC, isolation tenant, GDPR, billing Stripe
- **Stack** : FastAPI, PostgreSQL/pgvector, Celery, React 18, TypeScript, LangChain/LangGraph, Docker

HistText - Plateforme d'Analyse de Textes à Grande Échelle

2023–2025

Recherche & Développement Full-Stack

Publié : [JDMDH 2024](https://arxiv.org/abs/2404.18244)

- **Plateforme d'analyse de textes** traitant des milliards de tokens de documents historiques, déployée pour la communauté internationale de recherche
- **Architecture évolutive** (millions de documents), analyses temps réel et visualisations interactives
- **Stack** : Python, React, Rust, R, PostgreSQL, Apache Solr, Docker

Expériences professionnelles

Développeur Full-Stack & Ingénieur IA — *Projet Personnel*

Sept. 2025 – Présent

Plateforme SaaS IA de bout en bout : architecture, backend async, sécurité multi-tenant, frontend React.

Data Scientist & Ingénieur ML — *Projet ENP-China, Aix-Marseille Université*

Janv. 2023 – Août 2025

Développeur principal de HistText ; ML sur textes historiques à grande échelle ; ateliers dans 4 pays.

Doctorant & Scientifique ML — *LIS / IrAsia, Aix-Marseille Université*

Sept. 2019 – Déc. 2022

7 publications à comité de lecture ; bourse ERC Avancée ; conférences dans 7 pays.

Compétences

IA & LLM	RAG hybride (pgvector + BM25), GraphRAG, RAPTOR, Text-to-SQL, LangChain, LangGraph, Langfuse, litellm
NLP & ML	NER Multilingue, Transfer Learning, OCR, PyTorch, Hugging Face, spaCy, scikit-learn
Backend	Python, FastAPI (async), SQLAlchemy 2.0, Celery, Redis, PostgreSQL, pgvector
Frontend & Infra	React 18, TypeScript, TailwindCSS, D3.js, SaaS multi-tenant, RBAC, JWT, Docker, GCP, CI/CD, Stripe

Formations

Doctorat en Informatique (Machine Learning) — *Aix-Marseille Université*

2022

Thèse : Extraction d'événements à partir de fac-similés de documents anciens (Dir. Prof. Benoit Favre, projet ERC ENP-China n°788476)

Master Informatique (IA/ML) — *Aix-Marseille Université*

2019

Licence Informatique — *Aix-Marseille Université*

2016