

Baptiste Blouin

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

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

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 Clés

Plateforme SaaS de Transformation de Données par IA

2025–Présent

Conception & Développement Full-Stack

Projet personnel — aubier-data.fr

- **Conçu et développé une plateforme SaaS multi-tenant** de bout en bout : extraction documentaire, RAG, text-to-SQL, analyse et prédictions
- **Implémenté un moteur RAG avancé** avec recherche hybride (pgvector + BM25), GraphRAG et RAPTOR sur des documents non-structurés (PDF, OCR, VLM)
- **Développé un moteur text-to-SQL** convertissant le langage naturel en requêtes SQL avec streaming SSE et raisonnement pas-à-pas
- **Intégré 15+ connecteurs de sources de données** : bases relationnelles (PostgreSQL, MySQL, MongoDB), APIs SaaS (Salesforce, HubSpot, Stripe), S3
- **Mis en place une architecture de sécurité complète** : JWT RS256, RBAC, isolation tenant, GDPR, audit logging, rate limiting, chiffrement Fernet
- **Stack** : FastAPI, PostgreSQL/pgvector, SQLAlchemy 2.0, Celery, Redis, React 18, TypeScript, LangChain, LangGraph, Langfuse, litellm, fastembed, Stripe, Docker

HistText - Plateforme d'Analyse de Textes à Grande Échelle

2023–2025

Recherche & Développement Full-Stack

Publié : [JDMDH 2024](https://arxiv.org/abs/2404.12345)

- **Développé une plateforme d'analyse de textes** traitant des milliards de tokens de documents historiques
- **Conçu une architecture évolutive** pour gérer des millions de documents aux exigences complexes
- **Implémenté une solution ML full-stack** avec analyses en temps réel et visualisations interactives
- **Déployé la plateforme** pour la communauté internationale de recherche en humanités numériques
- **Stack Technique** : Python, React, Rust, R, PostgreSQL, Apache Solr, Docker

Dataset NER Chinois & Pipeline ML

2023–2024

Ingénierie de Données & ML

Publié : [LREC-COLING 2024](https://arxiv.org/abs/2404.12345)

- **Travaillé sur l'annotation** d'un dataset NER chinois historique (1872-1949)
- **Développé des systèmes de contrôle qualité** automatisés pour datasets multilingues
- **Créé un pipeline d'entraînement ML** avec méthodologie d'évaluation
- **Stack Technique** : Python, frameworks d'annotation, PostgreSQL, pipelines de validation

Modèle de Langue Chinoise & Système de Tokenisation

2023

Développement de Modèles ML

Publié : [NLP4DH 2023](https://arxiv.org/abs/2404.12345)

- **Analysé les motifs linguistiques** du chinois transitionnel par investigation systématique
- **Développé des modèles de tokenisation** atteignant 83% de précision avec 35% d'amélioration
- **Collaboré avec les chercheurs** d'Academia Sinica sur ce projet international
- **Stack Technique** : PyTorch, TensorFlow, algorithmes de tokenisation, fine-tuning

Simulation d'Erreurs OCR & Robustesse de Modèles

2022

Recherche Expérimentale

Publié : [TALN 2022](https://arxiv.org/abs/2404.12345)

- Étudié la performance des modèles ML sous conditions de données bruitées
- Développé une approche d'augmentation réduisant l'impact d'erreur de 50%
- Évalué la robustesse des modèles avec un framework de benchmarking
- **Stack Technique** : PyTorch, Transformers, algorithmes de simulation, frameworks d'évaluation

Transfer Learning pour l'Adaptation de Domaine

Développement & Optimisation ML

2021

Publié : NLP4DH 2021

- Appliqué le few-shot learning atteignant 93% de récupération de performance avec données minimales
- Implémenté des architectures neuronales character-aware pour traitement de texte en production
- Évalué les modèles sur plusieurs datasets avec analyse statistique
- **Stack Technique** : BERT, transformers character-aware, techniques d'adaptation

Expériences professionnelles

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

Sept. 2025 – Présent

- **Ownership produit complet** : de la conception de l'architecture à la mise en production, en autonomie totale
- **Gestion de la complexité technique** - modélisation de données, API, workers async, sécurité, frontend

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

Janv. 2023 – Août 2025

- **Chercheur principal et développeur** pour la plateforme HistText, gestion de projet indépendante
- **Exploration et résolution** de défis ML complexes dans le traitement de textes historiques à grande échelle
- **Communication de concepts techniques** par ateliers de formation dans 4 pays
- **Collaboration interdisciplinaire** avec historiens, linguistes et informaticiens internationaux
- **Entraînement distribué** de modèles de langue et NLP à grande échelle sur Jean-Zay (IDRIS/GENCI, supercalculateur national), orchestration SLURM

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

Sept. 2019 – Déc. 2022

- **Recherche indépendante** sur subvention ERC Avancée européenne, apprentissage auto-dirigé
- **Publié 5+ articles à comité de lecture**, rédaction technique et communication
- **Présenté la recherche** en conférences internationales dans 7 pays (France, Inde, Vietnam, Japon, Italie, USA, Taïwan), prise de parole publique
- **Mentoré et collaboré** avec équipes de recherche internationales, leadership interculturel
- **Calcul HPC sur Jean-Zay** (IDRIS/GENCI) : entraînement de modèles de langue et NLP (NER, segmentation de mots, adaptation de domaine)

Compétences

IA & LLM	RAG hybride, Text-to-SQL, LangChain, LangGraph, Langfuse, litellm, fastembed, pgvector
NLP & ML	NER Multilingue, Transfer Learning, OCR, PyTorch, Hugging Face, spaCy, scikit-learn, FLAML
Backend	Python, FastAPI (async), SQLAlchemy 2.0, Celery, Redis, PostgreSQL, pgvector, MySQL, MongoDB
Frontend	React 18, TypeScript, TailwindCSS, D3.js, Vega-Lite, React Query, Zustand
Architecture & Infra	SaaS multi-tenant, RBAC, JWT RS256, SSO (OIDC), 2FA (TOTP), GDPR, Docker, GCP, Linux, Jean-Zay / IDRIS (HPC), SLURM, CI/CD, Prometheus, Stripe
Autres	Rust, Java, R, Apache Solr, Elasticsearch, Playwright, Docling, LaTeX

Formations

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

2022

Thèse : Extraction d'événements à partir de fac-similés de documents anciens pour les études en histoire

Compétences de Recherche : Investigation indépendante, test d'hypothèses, conception expérimentale, rédaction technique

Directeur : Prof. Benoit Favre | **Projet** : Subvention ERC Avancée ENP-China (no. 788476)

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

2019

Licence Informatique — *Aix-Marseille Université*

2016