Anas Belfathi

+33 7 45 47 90 70 | anasbelfathi247@gmail.com | linkedin.com/in/anas-belfathi | anas-belfathi.vercel.app/

Skills

Languages: Python, C/C++, Java, JavaScript, TypeScript, SQL

Technologies & Tools: Hugging Face (Transformers/Datasets/PEFT), PyTorch, FAISS, Docker, vLLM, Kubernetes, MLflow, AWS SageMaker, Azure MLOps, TensorFlow

Work Experience

Nantes Laboratory of Digital Sciences (LS2N) — Oct 2023 – Present

LLM Research Engineer (Ph.D. Candidate) — *France*

- Engineered **LLM** pipelines for long-document **classification** in legal and medical domains, including **BERT** and **DeepSeek**, achieving **+15% macro-F1** vs baseline.
- Analyzed **20+ summarization** datasets using **GPT-4**, **Qwen-3**, and **DeepSeek** with the **Hugging Face** stack; added retrieval and constrained decoding for **+4.6 ROUGE-L**, **+8.9** factuality points, and **−22**% hallucinations on held-out legal and medical cases.
- Collaborated with **legal scholars** to build an annotated corpus of **U.S. Supreme Court** cases, **30,000+** long opinions and **1.2M** sentences; achieved inter-annotator agreement $\kappa = \mathbf{0.72}$ and released baseline models and training scripts.

ADVANCE.AI — Jun 2023 – Oct 2023

Data Scientist Intern — *Singapore*

- Designed **text + time-series** pipeline combining headline embeddings with **OHLCV** features for next-day movement classification; **+3.1** macro-F1 vs text-only baseline.
- Shipped **entity & event extraction** (tickers, guidance changes); **P@1 = 0.83**, reduced analyst triage time by **32%**.
- Increased revenue by packaging the pipeline as a premium signal for clients, generating **$120k ARR**.

PROCORE — Feb 2022 – Jul 2022

Software Engineer Intern — *United States*

- Built an internal web app to manage **projects, staff availability, and site-visit reports**; automated timesheets and leave tracking, cutting admin workload by **30%**
- Implemented document versioning for **drawings & RFIs** with role-based access; reduced email back-and-forth by **40%** and improved handoff between architects and engineers
- Deployed dashboards for project KPIs (deadlines, hours, costs) used by **50+** employees; decreased late-reporting incidents by **25%**

Education

Nantes University — Expected Jan 2027

Ph.D. Candidate in Artificial Intelligence — *France*

- Research focus on LLMs for long legal and medical documents: classification, summarization, NER, retrieval-augmented generation, evaluation and serving
- **7+** publications on LLMs with **4** under review; **2** Best Paper Awards; recent submissions to TALN, ICAIL and ACL

Sorbonne Paris Nord University — 2023

M.Sc. in Data Science (Artificial Intelligence), Valedictorian — *France*

- Program recognized among France's leading AI/Data Science tracks
- Core coursework: Machine Learning, Deep Learning, Natural Language Processing, Probabilistic Modeling, Optimization

Mediterranean Machine Learning Summer School (by DeepMind) — 2025

Participant (selective summer school) — *Croatie*

- Team project on reproducible ML and responsible AI practices; poster/demo at the end of the school.
- Selected participant in a competitive international cohort; collaborated with peers from diverse research backgrounds.
- Completed intensive modules in deep learning, representation learning, and RL; hands-on sessions with research mentors.

Publications & Awards

[1] Coupling Local Context and Global Semantic Prototypes via a Hierarchical Architecture for Rhetorical Role Labeling. **Anas Belfathi**, Nicolas Hernandez, Laura Monceaux, Mary C. Lavissière, Warren Bonnard, Christine Jacquin, Richard Dufour. **EACL 2026 (Under Review)**.

[2] A Simple but Effective Context Retrieval for Sequential Sentence Classification in Long Legal Documents. **Anas Belfathi**, Nicolas Hernandez, Laura Monceaux, Richard Dufour. **ACL 2025**.

[3] Is Selective Masking a Key to Improving Domain Adaptation for Masked Language Models? **Anas Belfathi**, Ygor Gallina, Nicolas Hernandez, Laura Monceaux, Richard Dufour. **ICAIL 2025**.

[4] The Role of Context in Sequential Tasks for Long Documents. **Anas Belfathi**, Nicolas Hernandez, Laura Monceaux, Richard Dufour. **TALN 2025**, **Best Paper Award**.

[5] Adapting Language Models to Specialized Domains via Genre- and Topic-Based Selective Masking. **Anas Belfathi**, Ygor Gallina, Nicolas Hernandez, Laura Monceaux, Richard Dufour. **TALN 2024**.

[6] Harnessing GPT-3.5-Turbo for Rhetorical Role Prediction in Legal Cases. **Anas Belfathi**, Nicolas Hernandez, Laura Monceaux, Richard Dufour. **JURIX 2023**, **Best Paper Award**.

[7] Enhancing Pre-trained Language Models with Sentence Position Embeddings for Rhetorical Role Recognition in Legal Opinions. **Anas Belfathi**, Nicolas Hernandez, Laura Monceaux, Richard Dufour. **ICAIL 2023**.