

Jean-Baptiste de la Broise

Senior NLP Engineer
MDPI AG, Basel
(Switzerland)

34 Boulevard de l'Europe
68100 Mulhouse
France

✉ jeanbaptistedelabroise@gmail.com

in [Jean-Baptiste De la Broise](#)

github/ [JeanBaptiste-dlb](#)



Work experiences

- 2021–Today **Data Scientist/NLP engineer**, MDPI AG, Basel
Train, deploy, and evaluate models for many NLP applications regarding academic writing. Pre-Training of domain-specific language models based on the transformer architecture. Classification of unbalanced data.
Embedding representation of most of the academic works (referenced by Pubmed, Scopus, ...) to create a scalable recommendation system using a vector database.
Deployments of 72B parameters LLM on premise using vllm and kubernetes.
Work on using LLM and RAG to generate sourced descriptions of diverse academic topics (first using llama2, then mistral 7b, then mixtral).
Using NLP to develop ethical content check of submitted manuscripts.
Custom parsing and processing of internal guidelines to develop a RAG system using Qdrant and Qwen2.5-VL
Mentoring a master internship student on a Machine Learning project
- 2021–2021 **Internship**, MDPI AG, Basel
Pre-training of a BERT language model for many applications related to academic writing.
Topic and Journal recommendations for academic articles. Displaying the business value of using Language modeling for many applications in the publishing industry.

Schooling

- 2020–2021 **Exchange Student**, *Ecole Polytechnique Federale de Lausanne*, Lausanne
focus on applied data science, Natural Language Processing and much more...
- 2018–2021 **Completed Computer Science Engineering degree delivered by the CTI organization**, at *ENSISA engineering school*, Mulhouse France
- 2016–2018 **Preparatory classes for the "Grandes Écoles"** focus on Maths and Physics, at *Antoine de Saint-Exupery high school*

University projects

- 2020–2021 **semester project in data science**, *EPFL - Human Computer Interaction Group*, Lausanne
Work on interjections in order to improve users' perception of chatbots.
- 2020 **project in data science**, *ENSISA*, Mulhouse
Project in computer vision. Predict if a person has coronavirus based on an X-ray image of their lung.

Publications

- 2022 **How to pretrain an efficient cross-disciplinary language model: The ScilitBERT use case**, doi.org/10.1109/ICITR54349.2021.9657164
- 2026 **Cite Lens: An AI Tool for Detecting Out-of-Scope and Out-of-Context Citations**, doi.org/10.1007/978-3-032-06136-2_3

Skills

- Development Python + (Poetry, Pytorch, Faiss, Hugging Face suite, langchain, FastAPI, FastMCP, Pydantic, Polars, Matplotlib, Scikitlearn, NLTK, ...), Java, C, C++, Unix, SQL, Git, Kubernetes...
- Others AI, software engineering, project management, blog post writing

Languages

- | | | |
|---------|-------------------|----------------------------------|
| French | | <i>native language</i> |
| English | TOEIC 990/990 | <i>C1, main language at work</i> |
| German | Goethe-Zertificat | <i>B2</i> |

Areas of interest

- Magic the Gathering a favourite trading card game of mine; I have been playing it for 14 years.
- Current Events general, high-tech, hardware benchmarks, free software, Linux, security breach, economy
- Sport skiing, via ferrata, inline skating, hiking