



## EXPERIENCE PROFESSIONNELLE ET DE RECHERCHE

### CATIE, Data Scientist

Jan. 2021 – présent

Membre de l'équipe Algorithmes et Données

Bordeaux, France

#### Missions :

Accompagnement des entreprises dans leur développement technologique sur des sujets de science des données.

En pratique, participation à une vingtaine de projets différents pouvant être classés de trois façons différentes :

- Etudes de faisabilité/de pré-industrialisation et encadrements.

- Clustering de données non annotées
- Expérimentation de méthodes de l'état de l'art
- Encadrement d'étudiants et acculturation en IA

- Aide à la conception et au prototypage de solutions d'apprentissage profond en image/texte, d'apprentissage machine ou de séries temporelles.

- Contrôle qualité (ex : détection de défauts dans des images)
- Optimisation de process (ex : recherche de documents similaires, analyse satisfaction client, etc.)
- Maintenance en temps réel (ex : détection d'anomalies en embarqué)

- Recherche & Développement du CATIE notamment en NLP.

- Entraînement à partir de zéro de manière efficiente d'un modèle encodeur-décodeur en français de 147M de paramètres (le Flash Attention T5) : <https://github.com/catie-aq/flashT5>
- Amélioration de l'état de l'art en français en Question Answering : <https://hf.co/CATIE-AQ/QAmembert> et en Reconnaissance d'Entités Nommées : <https://hf.co/CATIE-AQ/NERmembert-base-3entities>
- Jeux de données de prompt en français : <https://hf.co/datasets/CATIE-AQ/DFP>

### INSERM U1219 « Bordeaux population health », Data Scientist

Nov. 2019 – Déc. 2020

Membre de l'équipe IETO (Injury Epidemiology Transport Occupation)

Bordeaux, France

#### Travail effectué :

- Entraînement sous PyTorch de *transformers* (BERT/GPT2) pour classifier automatiquement des anamnèses du service des urgences du CHU de Bordeaux. Publication : <https://ai.jmir.org/2023/1/e40843/>
- Désidentification des anamnèses du service des urgences du CHU de Bordeaux (mix de techniques de type *transformers*, Bi-LSTM et de règles). Publication : <https://journals.flvc.org/FLAIRS/article/view/128480>
- Communications dans des congrès (Dataquitaine 2020 , PFIA 2021)



## EDUCATION

### Université de Bordeaux, Master MAS parcours MSS mention bien

2017 – 2019

Master Mathématiques appliquées et statistique, parcours Modélisation Statistique et Stochastique

Bordeaux, France

#### Principaux cours :

- *Machine Learning* et *Deep Learning*, *Data mining*, Séries temporelles, Analyse de données en grande dimension
- Algorithmes stochastiques, Optimisation convexe, Chaînes de Markov, Martingales, Analyse de survie et de fiabilité
- Ethique, Management, Théorie des organisations

## COMPETENCES

<b>Python</b>	Machine Learning/Deep Learning : PyTorch, Scikit-learn, WandB	Data Science : Pandas, Numpy
	NLP : Hugging Face (transformers, sentence-transformers, datasets)	Demo : Matplotlib, Seaborn, Streamlit
<b>Langues</b>	Anglais	

## DIVERS

<b>Blog</b>	Mon blog personnel dédié à l'apprentissage profond : <a href="https://lbouardois.github.io/blog/">https://lbouardois.github.io/blog/</a> Cité dans le rapport « Les nouveaux développements de l'intelligence artificielle » de l' Office parlementaire d'évaluation des choix scientifiques et technologiques <a href="https://www.senat.fr/notice-rapport/2024/r24-170-notice.html">https://www.senat.fr/notice-rapport/2024/r24-170-notice.html</a>
<b>Traduction</b>	Traduction du cours d'apprentissage profond de Yann Le Cun et Alfredo Canziani de l'Université de New York : <a href="https://lbouardois.github.io/cours-dl-nyu/">https://lbouardois.github.io/cours-dl-nyu/</a>
<b>Programme</b>	Membre du programme <a href="https://huggingface.co/hugging-fellows">Hugging Face Fellows</a> : <a href="https://huggingface.co/hugging-fellows">https://huggingface.co/hugging-fellows</a>
<b>Formation</b>	TP de <i>deep learning</i> (25H/an) à destination d'étudiants de l'ENSAM et de l'IUT de Bordeaux Cours d'initiation au <i>machine learning</i> (8h/an) à destination d'étudiants de l'école YNOV

# Loïck BOURDOIS

✉ Given only by email

✉ [loick.bourdois@outlook.com](mailto:loick.bourdois@outlook.com)

in [Loïck Bourdois](#)

🌐 [lbourdois.github.io](https://lbourdois.github.io)

## 👤 PROFESSIONAL AND RESEARCH EXPERIENCE

### CATIE, Data Scientist NLP

Jan. 2021 – present

Member of the Algorithms and Data team

Bordeaux, France

#### Missions:

Support companies in their technological development on data science topics.

In practice, participation in about 20 different projects that can be classified in three different ways:

- Feasibility/pre-industrialization studies and education.

- Clustering of non-annotated data
- Experimentation of state-of-the-art methods
- Student mentoring and AI acculturation

- Assist in the design and prototyping of image/text deep learning, machine learning or time series solutions.

- Quality control (e.g. image defect detection)
- Process optimization (e.g. search engine for similar documents, customer satisfaction analysis, etc.)
- Real-time maintenance (e.g. on-board fault detection)

- Research and development at CATIE, especially in NLP.

- Efficient training from scratch of a 147M parameters encoder-decoder model in French (the Flash Attention T5):

<https://github.com/catie-aq/flashT5>

- Improving the state-of-the-art in French in Question Answering: <https://hf.co/CATIE-AQ/QAmembert> and in Named Entity Recognition: <https://hf.co/CATIE-AQ/NERmembert-base-3entities>

- Datasets of French prompts: <https://hf.co/datasets/CATIE-AQ/DFP>

### INSERM U1219 « Bordeaux population health », Data Scientist

Nov. 2019 – Dec. 2020

Member of the IETO (Injury Epidemiology Transport Occupation) team

Bordeaux, France

#### Work performed:

- Train NLP transformers models (BERT/GPT2) in PyTorch to automatically classify emergency records from the emergency department of the Bordeaux University Hospital. Publication: <https://ai.jmir.org/2023/1/e40843/>
- De-identification emergency records in the emergency department of the Bordeaux University Hospital (mix of transformers, Bi-LSTM and rules techniques). Publication: <https://journals.flvc.org/FLAIRS/article/view/128480>
- Communications in French conferences (Dataquitaine 2020, PFIA 2021)

## 🎓 EDUCATION

### University of Bordeaux, Master MAS specialization MSS with mention

2017 – 2019

Master Applied Mathematics and Statistics specialization Statistical and Stochastic Modeling

Bordeaux, France

#### Main courses:

- Machine Learning & Deep Learning, Data mining, Time Series, Big Data
- Stochastic algorithms, Convex optimization, Markov chains, Martingales, Survival and reliability analysis
- Ethics, Management, Organization theory

## ⚙️ SKILLS

### Python

Machine Learning/Deep Learning: PyTorch, Scikit-learn, WandB  
NLP: Hugging Face (transformers, sentence-transformers, datasets)

Data Science: Pandas, Numpy  
Demo: Matplotlib, Seaborn, Streamlit

Languages French, English

## MISCELLANEOUS

### Blog

Personal blog dedicated to deep learning (in French): <https://lbourdois.github.io/blog/>  
Quoted in the report “Les nouveaux développements de l'intelligence artificielle” by the “Office parlementaire d'évaluation des choix scientifiques et technologiques” (a working group of French deputies and senators): <https://www.senat.fr/notice-rapport/2024/r24-170-notice.html> (in French)

### Translation

French translator of the *Deep learning course* by Yann LeCun and Alfredo Canziani from New York University  
<https://lbourdois.github.io/cours-dl-nyu/> (in French)

### Fellowship Course

Member of the [Hugging Face Fellows](https://huggingface.co/hugging-fellows) program: <https://huggingface.co/hugging-fellows>  
Deep learning practical work (25H/year) for students at ENSAM and IUT of Bordeaux  
Introductory course in machine learning (8h/year) for YNOV students