

Rayane Bencharef

+1-579-421-2464 (CA) | +33-7-66-83-39-12 (FR) | rayane.bencharef.1@ens.etsmtl.ca | linkedin.com/in/rayane/ |
github.com/JayRay5 | huggingface.co/JayRay5 | jayray5.github.io/

SKILLS

Machine Learning: Data Science, Model Optimization (Distillation, Finetuning), Multimodality (VQA), Computer Vision (Classification & Segmentation), NLP (Tokenization, LLMs), Time Series (Forecasting), Distributed Training, Data Engineering

MLOps & Cloud Computing: CI/CD Pipelines (Git Hooks / GitHub Actions), Containerization (Docker), Model Serving (Hugging Face Spaces), Version Control (Git/GitHub), HPC Workload Manager (Slurm), Automated Testing (Pytest)

Software Development: Full-stack Web Development, Backend Architecture, Database Design

Programming Languages: Python, R, JavaScript, Java, SQL

Frameworks/Libraries: PyTorch, TensorFlow, Hugging Face (transformers, accelerate), OpenCV, FastAPI, Gradio, ReactJS, Node.js

Developer Tools: CUDA, Visual Studio Code, Android Studio, LaTeX

Languages: French (Native), English (Professional Working Proficiency)

EDUCATION

École de Technologie Supérieure de Montréal (ÉTS)

Montreal, QC

Master of Science (M.Sc) in Artificial Intelligence with thesis

Sept. 2023 – Nov. 2025

- Mention Excellent (Table of Honor)
- Jury recommendation for the Master's Excellence Award

ISIS Castres (INSA partner)

Castres, France

Master of Engineering (M.Eng) in Software Engineering (CTI-accredited degree)

Sep. 2019 – Nov. 2025

European University of Cyprus

Nicosia, Cyprus

Student Exchange in Software Engineering (Erasmus)

Feb. 2023 – June 2023

EXPERIENCE

Graduate Researcher - Multimodality & Model Optimization

Jan. 2024 – Nov. 2025

Synchromedia, ÉTS

Montreal, QC

- Reduced the computational cost of a **Large Vision-Language Model** in DocVQA by studying two **distillation** approaches between **heterogeneous architectures**, which halved the latency (**896ms → 446ms**).*
- Fine-tuned the **GEMMA** LLM decoder with a hierarchical visual encoder for DocVQA, using QLoRA, **improving the performance from 80.20 to 82.67 ANLS.***
- Investigated positional encoding in Vision Transformer (ViT) using 2D Fourier features, increasing performance **from 83 to 84 ANLS.**
- Studied how VQA models handle structure and layout understanding through document classification and layout analysis tasks (**interpretability**).*
- Adapted single-page Document Understanding VLM to process multi-page documents **without adding parameters** for industrial applications.
- Developed a lightweight OCR Transformer with a **new decoder approach** in this field. Presented at the 22nd Conference of the International Graphonomics Society (**IGS 2025**), at Montréal
- Read and wrote scientific articles.
- *Presented & published at the **VisionDocs workshop (ICCV2025)** and received the **best paper award**.

Intern Data Scientist

Jun. 2023 – Aug. 2023

Atout Majeur Concept

Toulouse, France

- Engineered and analyzed patient data for **feature selection**.
- Built an SVM model to predict hospital stay duration from patient symptoms and characteristics, achieving **78% accuracy with limited data**.
- Developed a full pipeline to **automatically process** new patient data and generate predictions.

Independent Data Analyst	Dec. 2022
<i>Linkypharm.fr</i>	<i>Remote</i>
<ul style="list-style-type: none"> • Cleaned and preprocessed large pharmacy statistics datasets for downstream analysis. • Created data-driven geographic visualizations of France to highlight pharmacy usage and distribution patterns. 	
Independent Data Engineer	Sep. 2022 – Nov. 2022
<i>TrainPredict</i>	<i>Remote</i>
<ul style="list-style-type: none"> • Designed and implemented a data model for cycling-related datasets. • Built an interactive web application for statistical data visualization using React and Redux. 	
Intern Data Scientist in Time Series	May 2022 – Aug. 2022
<i>CHU Toulouse</i>	<i>Toulouse, France</i>
<ul style="list-style-type: none"> • Engineered and preprocessed emergency call datasets from SAMU31 (emergency medical service). • Conducted exploratory feature analysis using geographic and statistical visualizations. • Built ARIMA and LSTM forecasting models (Keras) to predict call volumes, reaching 80% accuracy. 	
Front-End Developer	Sep. 2021 – Aug. 2022
<i>Horus HealthCare Systems</i>	<i>Castres, France</i>
<ul style="list-style-type: none"> • Built a Django web application for the Castres Olympique rugby club to manage training sessions, matches, and events. • Designed responsive, user-centric interfaces with HTML5, JavaScript, and Bootstrap. • Worked in a 15-member team using Trello for project coordination and GitHub for collaborative development. 	
Full Stack Developer	Jan. 2021 – Sep. 2021
<i>TrainPredict</i>	<i>Castres, France</i>
<ul style="list-style-type: none"> • Built full-stack web and mobile applications (React, React Native, Redux, Node.js) to assist cyclists during training sessions. 	
Back-End Developer	Jul. 2020 – Feb. 2021
<i>Horus HealthCare Systems</i>	<i>Castres, France</i>
<ul style="list-style-type: none"> • Built a web application with a 10-member team using Sails.js for the French National Cancer Institute (INCA), enabling psychologists to track patient progress during treatment. 	

PROJECT

DIVE-Doc Platform	Dec. 2025
Python, Docker, GitHub Actions, Hugging Face Space	
<ul style="list-style-type: none"> • Deployed a Large Visual Language Model (LVLM) for document information extraction (VQA) using FastAPI to build endpoints and Gradio for the web interface. • Engineered a robust MLOps pipeline by automating Docker image building via GitHub Actions and continuous deployment to Hugging Face Spaces (CI/CD). • Implemented comprehensive Quality Assurance (QA) with Unit & Smoke tests and security scanning to ensure production reliability. 	
Deployed demo: <i>huggingface.co/spaces/JayRay5/DIVE-Doc-docvqa</i>	
Code Repository: <i>github.com/JayRay5/DIVE_Doc_platform</i>	
Cyprus Fish Species Recognition App	April. 2023
Python, Docker, GitHub Actions, Hugging Face Space	
<ul style="list-style-type: none"> • Personal project that I built during my Erasmus in Cyprus. • Created and published a custom dataset of 5 fish species that live around Cyprus. • Finetuned a ConvNext on the dataset using k-fold cross-validation due to the small amount of data, achieving 95% accuracy on the test set. • Served the model using FastAPI for the REST endpoints and Gradio for the web interface. • Set up an MLOps pipeline using GitHub Actions for CI/CD and enforcing code quality via pre-commit hooks and automated testing. • Containerized the application with Docker and deployed it on Hugging Face Spaces. 	
Deployed demo: <i>huggingface.co/spaces/JayRay5/Cyprus-Fish-Recognition-App</i>	
Hugging Face Collection: <i>huggingface.co/collections/JayRay5/cyprus-fish-recognition</i>	
Code Repository: <i>github.com/JayRay5/cyprus-fish-classifier</i>	

International Conference on Computer Vision (ICCV), VisionDocs Workshop

Spotlight/Best Paper Award

Oct. 2025

Honolulu, Hawaii

- *DIVE-Doc: Downscaling foundational Image Visual Encoder into hierarchical architecture for DocVQA.*

Code Repository: github.com/JayRay5/DIVE-Doc

Model Weights: huggingface.co/JayRay5/DIVE-Doc-FRD