

DATA SCIENTIST - AI ENGINEER

As a research data scientist and machine learning specialist with hands-on experience across the full MLOps pipeline, Yessine works with large-scale, real-time research data to build reliable and interpretable AI solutions. Yessine has expertise in designing and deploying end-to-end ML systems, covering data ingestion, preprocessing, modeling, CI/CD automation, and scalable deployment using cloud-native tools such as Docker, Kubernetes, Jenkins/GitLab CI, and KServe. Yessine’s research focuses on time series forecasting and explainable AI, with applied projects in the field of healthcare (blood glucose prediction for type 1 diabetes) and decision support systems. At the same time, Yessine is closely interested in large language models (LLMs) and their integration into multi-agent frameworks such as MCP, exploring how agents can improve automation, reasoning, and adaptive decision-making in complex environments.

SKILLS

Programming	: Python, C/C++, Bash, Java
Machine Learning	: PyTorch, TensorFlow, NumPy, Pandas, Dask, Spark, Kafka, Matplotlib, Seaborn, scikit-learn, Keras, Deep Learning, Computer Vision, NLP, RL/DRL
Large Language Models (LLMs)	: Hugging Face Transformers, Prompt Engineering, RAG (LangChain, LlamaIndex), Vector DBs, Fine-tuning (LoRA/QLoRA with PEFT), Serving (KServe)
MLOps	: Kubeflow, Docker, Git/GitLab, Jenkins, Airflow, CI/CT/CD, DVC, MLflow, TFX, Neptune, Vertex AI, XAI
Mathematics	: Statistics, Probability, Optimization, Graph Theory, Game Theory
Platforms	: Linux, Azure, GCP
Data	: ETL/ELT, Data Integration, SQL, MongoDB

PROFESSIONAL EXPERIENCE

Research Assistant — Explainable AI (XAI) — Full-time <i>CIRRELT Lab, Université Laval</i>	May 2025 – Present Québec, Canada
<ul style="list-style-type: none">Designed a novel instance-based counterfactual explanation framework for CGM forecasting.Built a graph attention mechanism applied per timestamp to capture temporal feature interactions and lagged causality.Connected the first graph to a second graph to construct a shapelet-based graph model, enabling dual-graph training to identify specific patterns corresponding to hypoglycemia, normal, and hyperglycemia cases.Trained a residual LSTM model using actionable features (bolus, basal insulin, physical activity, carbohydrate intake, and steps taken) to accurately extract hypo- and hyperglycemia episodes.Developed a counterfactual engine that extracts patterns from corrected unsafe sequences and applies top-k shapelet combinations to generate minimal changes required to transform unsafe forecasts into safe ones.	
Research Assistant — AI Engineer — Full-time <i>CIRRELT Laboratory, Laval University</i>	January 2025 — April 2025 Quebec City, Canada
<ul style="list-style-type: none">Preparation and annotation of large educational datasets using GPT-4 for automated labeling and content generation.Adaptation and fine-tuning of the DeepSeek-8B LLM to support intelligent question-answering assistance for students, improving answer accuracy and contextual reasoning.Application of parameter-efficient fine-tuning techniques, such as QLoRA and low-rank adaptation methods, to optimize LLM training efficiency while maintaining strong performance.Design and deployment of a RAG system, fully integrated with the I4EvoSim education platform to improve knowledge retrieval and adaptive interaction.	
Research Assistant-Deep Learning-Full-time <i>CIRRELT Laboratory, Laval University</i>	Sept. 2024 – April 2025 Quebec City, Canada
<ul style="list-style-type: none">Development of a Physics-Informed Neural Network (PINN) framework to solve Hamilton–Jacobi–Bellman (HJB) equations with discrete control in the event of system failure.Transformation of the classical HJB formulation into a differentiable deep learning method, enabling gradient-based optimization and scalability.Integrating the PINN approach with reinforcement learning methods to assess performance in high-dimensional	

control tasks.

- Comparison of results with numerical dynamic programming solutions, validating accuracy and consistency in several scenarios.
- Validation of the framework against low-dimensional analytical solutions to ensure theoretical robustness and interpretability.
- Contribution to the preparation of a scientific article documenting the methodology, benchmarks, and theoretical knowledge.

Research Assistant – Data Science & Deep Learning – Full-time

Sept. 2023 – Aug. 2024

CIRRELT Lab, Université Laval

Québec, Canada

- Processed and engineered features from the **T1DEXIP** datasets (pediatric), including cleaning, 5-minute harmonization, alignment, and construction of lagged, window-based, and statistical features.
- Developed reproducible pipelines for multi-source ingestion, quality control, per-patient normalization, and dataset versioning to ensure transparency and replicability.
- Designed and benchmarked advanced forecasting models for Continuous Glucose Monitoring (CGM), including LSTMKAN, Residual LSTM, and Transformer-based architectures, with tailored strategies for **T1DEXI** adult vs. pediatric cohorts (independent splits, transfer learning, fine-tuning).
- Trained models on **Compute Canada** clusters using distributed GPU resources, leveraging mixed-precision training and adaptive learning-rate scheduling for efficiency and scalability.
- Conducted clinical/statistical evaluations (RMSE, MAE, hypo/hyperglycemia zones) and applied explainable AI methods to quantify feature influence (insulin, meals, physical activity) on blood glucose trajectories.

Research Intern –Data Science – Full-time

Feb. 2023 – Aug. 2023

CIRRELT Lab, Université Laval

Québec, Canada

- Built an automated pipeline for **T1DEXI** adult patient selection and multi-source time-series integration, ensuring consistent CGM and clinical data alignment.
- Developed **GMSE** as a custom loss function for model training and integrated it into the XGBoost model and deep learning model
- Designed an initial 30-minute CGM forecasting framework using custom neural architectures with transfer learning strategies.
- Benchmarked traditional machine learning models, including **XGBoost** and **Random Forest**, against LSTM and Seq2Seq models for both population-level prediction and transfer learning fine-tuning approaches.
- Applied **Optuna** for hyperparameter tuning and **Keras Tuner** for deep learning models to obtain the best per-patient results in short-term blood glucose prediction (30 minutes).
- Incorporated the **SHAP** method for XGBoost explainability to provide an overview of model behavior with respect to feature importance.

AI Intern – Full-time

Jul. 2022 – Aug. 2022

TALAN Tunisia

On-site

- Designed and deployed a data pipeline on Azure to support large-scale voice broadcasting workflows.
- Implemented real-time event streaming with Apache Kafka and parallel data processing using Apache Spark, ensuring low-latency and fault-tolerant data flow.
- Optimized data storage and retrieval by integrating MongoDB as the persistence layer, enabling efficient access for analytics and reporting.

AI Intern – Full-time

Jul. 2021 – Aug. 2021

Digital Research Center of Sfax

Remote

- Web application for stock forecasting and an analytical dashboard.

EDUCATION

Université Laval

Jan. 2024 – Dec. 2025

Master's with Thesis in Management Sciences — Operations and Decision Systems

Québec, Canada

- **GPA 3.9:** Courses: **GIF-7010, GLO-7027, MQT-7014, MQT-7015, MQT-6021, ADM-7900** (Time Series & XAI).

National School of Computer Science (ENSI)

Sept. 2022 – Sept. 2023

Research Master's in Intelligent Systems

Manouba, Tunisia

- Courses: Data Mining, Data Engineering, Robotics.

National School of Computer Science (ENSI)

Engineering Degree in Computer Science

Sept. 2020 – Sept. 2023

Manouba, Tunisia

- Courses: ML/DL, Distributed Databases, Distributed Systems, Networks, Security, Parallel Architecture, IoT.

Preparatory Institute for Engineering Studies

Physics and Chemistry

Sept. 2018 – Jun. 2020

Sfax, Tunisia

- National exam rank: **85 / 1200**.

Québec Comparative Evaluation (MIFI)

Ministry of Immigration, Francisation and Integration (MIFI)

Issued: Aug. 22, 2025

Québec, Canada

- **Baccalauréat (2018, Tunisia)** → Québec equivalent: Secondary School Diploma (DES) + one year of CEGEP.
- **Engineering Degree in Computer Science (2023, ENSI)** → Québec equivalent: Bachelor's + DESS in Computer Engineering.
- **Research Master's in Intelligent Systems (2023, ENSI)** → Québec equivalent: Master's (Maîtrise) in Computer Science.

PROJECTS

Deep Reinforcement Learning for Traffic-Light Control *DQN, DDQN, DRQN (Single/Multi-Agent), SUMO, Docker, Jenkins*

- Built a SUMO-based simulation environment to model real-world traffic flows and evaluate reinforcement learning policies.
- Developed a traffic optimization framework by benchmarking DQN, DDQN, and DRQN under single-agent and multi-agent settings.
- Applied MLOps practices (Docker, Jenkins) to automate training and ensure reproducibility of reinforcement learning workflows.

Kubeflow Pipeline “Digits-Recognizer” (MNIST) *Kubeflow Pipelines, TensorFlow/Keras, Docker, GKE, KServe, MinIO*

- Designed an end-to-end ML pipeline with Kubeflow for dataset preprocessing, model training, and evaluation.
- Trained a CNN on GKE with automated metric tracking and model versioning in MinIO.
- Deployed inference endpoints via KServe and validated predictions through Python notebooks and REST APIs.

Computer Vision Pipeline “Apple-Classifer” *CNN, Data Augmentation, TensorFlow/Keras, ONNX, FastAPI, Jenkins*

- Created a robust vision pipeline by applying data augmentation techniques (flip, rotation, zoom, brightness) to apple image datasets.
- Trained and optimized a CNN with EarlyStopping and evaluated performance using accuracy, F1-score, confusion matrix, and learning curves.
- Exported the best-performing model in ONNX format and deployed a real-time inference service using FastAPI.
- Integrated the pipeline into a CI/CD workflow (Docker, Jenkins) for continuous testing and deployment.

Hospital Management Platform

Laravel, Bootstrap, MySQL, jQuery

- Online solution to support medical staff (COVID-19).

PUBLICATIONS

- [1] Anas Neumann, Yessine Zghal, Marzia Angela Cremona, Adnene Hajji, Michael Morin, and Monia Rekik. A data-driven personalized approach to predict blood glucose levels in type-1 diabetes patients exercising in free-living conditions. *Computers in Biology and Medicine*, 190:110015, 2025.

AWARDS & CONFERENCES

PCD Challenge — 2nd Place

LessJam: traffic-light optimization

Tunisia Digital Summit — 2nd Place

Hack4Impact

JOPT 2025

[link](#)

CORS 2025

[link](#)

PROFESSIONAL DEVELOPMENT

Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow — Aurélien Géron

Deep Learning — Ian Goodfellow, Yoshua Bengio, Aaron Courville

Designing Machine Learning Systems — Chip Huyen

Machine Learning Design Patterns — Valliappa Lakshmanan, Sara Robinson, Michael Munn

Applied Machine Learning Explainability Techniques — David Foster, Danilo Sato