Mahdi Zarour
mahdi.zarour.main@gmail.com
514-999-3492 · Canada
Linkedin.com/in/MahdiZarour · https://mahdi-zarour.netlify.app/

SUMMARY

Machine Learning Engineer · Data Scientist · Data Engineer

- **With 5 years practicing in Data Science and Machine Learning,** dedicated to continuous learning and staying up to date with the latest advancements in machine learning, data science, and MLOps. Enjoy tackling new problems without being afraid to explore solutions directly from research papers to deliver the most effective results.
- **Technically Strong with a focus on Mathematics, Statistics, and Probability**.
- **Effective collaborator and Listener** with experience working in cross-functional teams. Proficient in understanding requirements, conducting comprehensive data analysis, and providing insightful recommendations to help make decisions.
- **Proactive and helpful team player** who contributes and shares his ideas, whether when asked or when a need is perceived. Always ready to provide assistance and make meaningful contributions to the team.

EDUCATION

Université de Montréal and MILA — Montréal, QC
Master of Science, Computer Science — 2020-2023

- **Coursework**: Data Science · Fundamentals of Machine Learning · Probabilistic Graphical Models · Representation Learning · Theory of Deep Learning.
- **GPA**: 3.7.
- **Scholarships**: International Students Scholarship $27,000 CAD.

Université Oran 1 Ahmed Ben Bella (IGMO) — Oran, Algeria
Licence (Bachelor of Science), Mathematics & Computer Science — 2015-2018

- **Grade**: A (Top 5% in the program).

EXPERIENCE

AAVAA — Montréal, QC
Machine Learning Engineer — Jan/2024 - Now

- Refactored and Optimized Python code for Data Processing and Feature extraction pipelines.
- Built Python pipelines to experiment and finetune Machine Learning models for Blink Detection.
- Developed and trained machine learning models for enhanced blink detection, achieving a 10% improvement in F1 (XGBoost) score without constraints and a 7% increase (LSTM) under size limitations, over the existing model.
- Collaborated within the ML team to implement coding best practices and enhance productivity with **Git** and **Docker**.
- Participated in daily stand-ups as part of an **Agile** framework, to enhance team motivation and communication.
- **Skills: MLFlow, Tensorflow, XGBoost, Pandas, Scikit-Learn, MangoDB, Agile, Signal Processing, Machine Learning, Deep Learning, Docker, Git**

Machine Learning Engineer / Software Engineer — Montréal, QC
Skylar — Jan/2023 - Dec/2023

Q&A with LLMs and RAG:

- Built a tailored chat bot app for a startup, with the large language model **ChatGPT** and **Flask**.
- Used OpenAI **Embeddings** and **Vector Databases** to augment ChatGPT prompts with relevant context (retrieval-augmented generation).
- Integrated a real-time speech-to-speech interface via the Microsoft **Azure** API to enable users to verbally pose questions and receive spoken responses.
- Collaborated within a **cross-functional** team as part of an **Agile** Framework to ensure continual client engagement through regular progress demonstrations and feedback integration.

Discord Digital Store:

- Designed an online store on the Discord Platform using UML and implemented it with **Python** and **Django.**
- Employed **OOP** concepts including **Inheritance** and **Polymorphism** with **Overloading** and **Overriding**.
- Managed data effectively using **PostgreSQL (CRUD)** and **SQLAlchemy**, including users, transactions, orders, and other information.
- Successfully deployed the store on **Heroku**, then actively monitored the application, to identify and rectify flaws and **bugs**.

- **Skills: Agile, CI/CD, Big Data, AWS, EMR, SageMaker, Lambda, S3, Python, R, PyTorch, Spark, Hadoop, XGBoost, Supervised Learning, Data Cleaning, Data Preparation, Object Detection, Computer Vision, NLP, Model Training, Deep Learning, LangChain, OpenAi, Vector Database, Microsoft Azure, Embeddings, Jupyter Notebook.**

Beam Me Up Inc. Montréal, QC
Data Scientist / Unity Developer Intern Jan/2022 - Dec/2022

- Implemented a python library (with **NumPy, Pandas,** and **Matplotlib**) to process raw Eye Tracking and EEG data by applying the latest feature extraction algorithms to data captured from consumer grade VR and EEG devices, and to visualize the results.
- Developed a virtual environment in **Unity** using **C#** language and successfully conducted experiments on 31 individuals at BMU labs.
- Facilitated **Agile** product development with BMU team, and implemented **CI/CD** practices to increase deployment release frequency and decrease deployment time.
- Developed a state-of-the-art distraction detection system in terms of accuracy (84%, best among those present in the literature) by combining features from EEG and Eye Tracking and applying a machine learning classification algorithm, SVM with **Scikit-learn**.
- Developed the first (in the literature) numerical indicator of the levels of distraction by combining linear dimensionality reduction techniques (Linear Discriminant Analysis) and interpolation for effective Supervised Learning.
- **Skills: Agile, CI/CD, Data Analysis, Data Visualization, Data Acquisition, Data Cleaning, Data Preparation, C#, Unity, OOP, Python, R, EEG, Eye Tracking, Scikit-learn, PyTorch, Numpy, Pandas, Matplotlib, Virtual Reality, Supervised Learning, Unsupervised Learning, Signal Processing, Jupyter Notebook.**

PROJECTS

- **Object Detection:** Implementation in **Pytorch** of the famous "You Only Look Once" **CNN** based object detection algorithm used in real time computer vision systems. Repo
- **Fraud Detection:** Detection of frauds using ML with Apache Spark for distributed processing and XGBoost. Repo
- **Tutorial System Management:** A Spring Boot, Angular, and PostgreSQL project to implement CRUD operations. Repo
- **DentalGPT :** Utilizing Retrieval Augmented Generation (**RAG**) to answer questions about dental aligners with ChatGPT, everything with a beautiful Streamlit UI. Repo
- **PicAI:** Django based web application for basic image processing using Pandas, Sklearn, and Numpy.
- **Semi-Supervised Image Classification:** Combining Unsupervised Representation Learning and only 100 labels to maximize classification accuracy on MNIST Dataset.
- **River Flow and Water Level Forecasting for HydroQC:** Using Time Series models such as Prophet and Arima, ranked second at HackQC 2022.
- **Plant Disease Detection:** Using Deep Learning CNN model, won first place at CodeML 2022 at Polytechnique Montreal.
- **Movie recommendation system:** with NLP Sequence BERT based on the user's movie history. Repo

SOFT SKILLS

- Strong problem-solving and analytical skills.
- Excellent communication and collaboration skills.
- Good mathematical modeling.
- Proficient in English, fluent in French and Arabic.