Jakhongir **Saydaliev**

Data Scientist · NLP Engineer

Lausanne, Switzerland

(+41) 766038488 | jakhongir.saydaliev@epfl.ch | jakhongir0103.github.io | Jakhongir0103 | Jakhongir Saydaliev

Summary

I am currently pursuing an **MSc** in **Data Science** at **EPFL**, where I work as a **Research Student Assistant** in the **NLP lab** under the supervision of Prof. Antoine Bosselut. Prior to this, I served as a Research Student Assistant at the **DHLAB**, where I was supervised by Frédéric Kaplan. I hold a **BSc** in **Computer Engineering** from **Politecnico di Torino** and completed a 6-month internship as a **Data Analyst** at Fater, under the supervision of Salvatore Croce.

Education

École polytechnique fédérale de Lausanne (EPFL)

Lausanne, Switzerland

MSc in Data Science — *Sep. 2023 - Present*

- Avg. Grade: 5.3/6

Polytechnic University of Turin

Turin, Italy

BSc in Computer Engineering — *Sep. 2019 - Jul. 2023*

- ToPoliTo scholarship for being ranked 8th in admissions to the university.
- Avg. Grade: 109/110

Work Experience

NLP lab, EPFL

Lausanne, Switzerland

Research Student Assistant — *Jun. 2024 - Present*

- Contributing to the Multilingual Model Training project as part of the Swiss AI initiative.
- Developed a domain clustering pipeline for multilingual text data.
- Trained multilingual sentence embedding models using contrastive learning techniques.
- Expanded language support for multilingual sentence embeddings through model distillation.

DHLAB, EPFL

Lausanne, Switzerland

Research Student Assistant — *Feb. 2024 - Sep. 2024*

- Fine-tuned large language models (LLMs) for question-answering tasks.
- Designed and implemented a text-to-SQL system pipeline.
- Built an LLM-driven agent system for table-based question-answering tasks.

Fater

Pescara, Italy

Data Analyst — *Nov. 2022 - May. 2023*

- Contributed to the development of customer churn prediction model.
- Analyzed the customer data to extract insights.

LINKS Foundation

Turin, Italy

IoT intern — *Oct. 2021 - Feb. 2022*

- Developed an interface to visualize the positioning and other information of an IoT device..

Teaching Experience

Applied Data Analysis, EPFL

Lausanne, Switzerland

Student Teaching Assistant — *Fall, 2024*

- The most popular course in IC MSc in 2024 Fall with over 700 students.

Projects

LLM Training with SFT, DPO, and RAG

Lausanne, Switzerland

report — *Jun. 2024*

We first collected multiple-choice question-answering (MCQA) datasets from scientific fields, then fine-tuned the Galactica-1.3B model for the question-answering task, followed by DPO training. Next, we implemented RAG tuning, which integrates external knowledge and enhances performance by 11.52%. The resulting models are now available on Hugging Face.

Coin Detection and Classification *Lausanne, Switzerland*

CODE | SLIDES *May. 2024*

A Computer Vision project that consists of 2 parts: segmentation and classification. Given images of coins we first implemented the segmentation of the coins, and trained a classifier to label the coins with their values.

Reinforcement Learning on Mountain Car Environment *Lausanne, Switzerland*

CODE REPORT *May. 2024*

We implemented the DQN and Dyna Reinforcement Learning algorithms for the well known Mountain Car Environment. Also implemented the extension of the DQN algorithm with heuristic and non-domain specific reward functions to deal with the sparse reward environment.

YouTube Analysis *Lausanne, Switzerland*

CODE | DATASTORY | BLOG *Dec. 2023*

Causal Analysis of Tech channels' progress on YouTube using the videos published between May 2005 and October 2019. Through this analysis, we have identified several success factors of tech channels.

LLM Fine-Tuning *Lausanne, Switzerland*

REPORT | BLOG *Dec. 2023*

Fine-tuned 3 LLMs (Mistral-7B, Llama-2-7B, Phi-1.5) on a dataset from X for the stance detection task. Our trained model outperformed the baseline models on most of the datasets.

Cardiovascular Diseases Prediction *Lausanne, Switzerland*

CODE | REPORT *Oct. 2023*

Implemented the following standard ML algorithms using native python libraries and numpy for the classification task: Least Squares, Ridge and Logistic regressions.

Skills

Programming Languages: Python, C, Java, Shell, SQL

ML/AI: Pytorch, Transformers, Scikit-Learn

Languages: Uzbek (native), English (fluent)

Relevant Courseworks

Modern NLP, Machine Learning, Applied Data Analysis, Image Analysis and Pattern Recognition, Artificial Neural Networks and Reinforcement Learning, Operating Systems, Object Oriented Programming, Algorithms and Programming, Computer Architecture, Introduction to Databases.