Jakhongir **Saydaliev**

Data Scientist · NLP Engineer
Lausanne, Switzerland
jakhongir.saydaliev@epfl.ch | jakhongir.me | Jakhongir0103 | jakhongir-saydaliev-0103

Summary

I am currently an MSc Data Science student at EPFL and a Research Intern at SwissAI (ETHZ & EPFL AI Centers). Prior to this, I worked as a Research Assistant at the NLP Lab supervised by Prof. Antoine Bosselut.

Education

EPFL (École polytechnique fédérale de Lausanne) — *Lausanne, Switzerland*
MSc in Data Science — *Sep. 2023 - Aug. 2026*
- 2 semester projects: Multilinguality and Multimodality
- Avg. Grade: 5.6/6

Politecnico di Torino — *Turin, Italy*
BSc in Computer Engineering — *Sep. 2019 - Jul. 2023*
- ToPoliTo scholarship for being ranked 8th in the university admission.
- Avg. Grade: 109/110

Publications

ConLID: Supervised Contrastive Learning for Low-Resource Language Identification — *2025*
Jakhongir Saydaliev, Negar Foroutan, Ye Eun Kim, Antoine Bosselut — *arXiv:2506.15304*

LLM-Powered Agents for Navigating Venice's Historical Cadastre Identification — *2025*
Jakhongir Saydaliev, Tristan Karch, Isabella Di Lenardo, Frédéric Kaplan — *arXiv:2505.17148*

Work Experience

Logitech — *Lausanne, Switzerland*
ML Research Intern (upcoming) — *Sep. 2025 - Feb. 2026*
- Computer use; LLM agents.

SwissAI, ETHZ & EPFL — *Lausanne, Switzerland*
Summer Research Intern — *Jun. 2025 - Sep. 2025*
- Working on Reasoning for vision language models through reinforcement learning.

Skills: GitHub, HPC, Python, PyTorch, HuggingFace, VeRL

NLP lab, EPFL — *Lausanne, Switzerland*
Research Student Assistant — *Jun. 2024 - Jun. 2025*
- Adapted large multimodal models (LMMs) for multimodal reasoning.
- Worked on Multilingual Model Training project 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.

Skills: GitHub, HPC, Python, PyTorch, HuggingFace

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.

Skills: GitHub, Python, PyTorch, HuggingFace, LangChain, LangGraph

Fater — *Pescara, Italy*
Data Analyst — *Nov. 2022 - May. 2023*
- Analyzed the customer data to extract insights.
- Contributed to the development of customer churn prediction model.

Skills: GitHub, Python, NumPy, Pandas

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

Technical Skills: Python, PyTorch, Hugging Face, Transformers, Scikit-learn, NumPy, Pandas, SQL, Git

HPC: Run:ai, SLURM

Artificial Intelligence: Fine-Tuning, Transfer Learning, Contrastive Learning, Model Distillation, Optimization Techniques

Natural Language Processing: Language Identification, (Multilingual) Sentence Embeddings, Text-to-SQL Systems, Question-Answering Systems, Stance Detection, Retrieval-Augmented Generation (RAG), Data Preprocessing (Tokenization, Lemmatization), Low-Resource Language Processing

Computer Vision: Image Segmentation, Image Classification, Image Generation

Data Analysis and Visualization: Causal Analysis, Data Storytelling, Visualization Tools (Matplotlib, Seaborn, Plotly, Tableau)

Languages: Uzbek (native), English (fluent)

Relevant Courseworks

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