

Harold Benoit ([email](#), [website](#))

EDUCATION

Swiss Federal Institute of Technology (EPFL)

MSc of Data Science

- 5.8/6.0 GPA (ranked 3rd in year)

2021 - 2023

Lausanne, Switzerland

Swiss Federal Institute of Technology (EPFL)

BSc of Computer Science / Communication Systems

- 5.5/6.0 GPA

2018 - 2021

Lausanne, Switzerland

WORK EXPERIENCE

Swiss Federal Institute of Technology (EPFL)

Research Associate

- 2024: Research and engineering of LLMs (throughput, SFT, eval) for the [Swiss AI](#) research initiative, co-led by [EPFL](#) and [ETH Zurich](#). Roughly summarized as training LLMs from scratch with 10 million GPU hours of compute, and conducting research along the way.
- 2023: Research on generative models to automatically create tailored training data for model fine-tuning @ [VILAB](#). (advised by [Amir Zamir](#))

October 2023 - Present

Lausanne, Switzerland

IBM Research

Research Intern (advised by [Mattia Rigotti](#) & [Amir Zamir](#))

- Research on data & algorithmic design for out-of-distribution (OOD) generalization of neural networks.

February 2023 - September 2023

Zurich, Switzerland

G-Research

Quantitative Research Intern

- Extracting meaningful predictive signals from financial data to predict movements in financial markets. Further details are under NDA. *Skills*: Scala, Spark, Python, ML, Algorithmic Design

July 2022 - September 2022

London, UK

PUBLICATIONS

[Controlled Training Data Generation with Diffusion Models](#)

T. Yeo*, A. Atanov*, H. Benoit[^], Aleksandr Alekseev[^], Ruchira Ray, Pooya Akhoondi, Amir Zamir

In review, 2024

[Unraveling the Key Components of OOD Generalization via Diversification](#)

H. Benoit*, L. Jiang*, A. Atanov*, O. Kar, M. Rigotti, A. Zamir

ICLR, 2024

SOFTWARE & PROJECTS

1st place at LLM training hackathon ([Github](#) / [Details](#))

I won \$1.5k and a trip to Huawei HQ in China. The competition was to train the best performing language model on [Sлимпajama](#) with 3 hours of compute on a A100 GPU.

April 2024

Integrated Comfort Controller Pipeline (ICCP) ([Github](#) / [Documentation](#))

Built *ICCP*, a PyTorch-based reinforcement learning library, for civil engineering and energy saving, during my time as a research assistant in [ICE Lab](#), advised by [Dolaana Khovalyg](#).

June 2022

SKILLS & INTERESTS

- **Skills**: Machine Learning; Deep Learning; "Generative AI"; Large Language Models; Statistics; NLP; Diffusion Models; Computer Vision; Big Data Engineering; Database systems; Advanced Algorithms;
- **Programming**: Python; PyTorch; Pandas; NumPy; SQL; Scala; Java; Go; Git; Linux; Bash; Docker.