Harold Benoit *([email](), [website](), [blog](), [twitter]())*

EPFL

WORK EXPERIENCE

Liquid AI **July 2024 - Present**
Member of Technical Staff - Research Scientist/Engineer *Remote*
- Hardware-aware model design, co-lead of the edge inference stack
- Training improvements, distributed training infrastructure optimization, synthetic data pipelines.

Swiss Federal Institute of Technology (EPFL) **October 2023 - June 2024**
Research Associate *Lausanne, Switzerland*
- 2024: Research and engineering of LLMs (throughput, SFT, eval) for the [Swiss AI]() research initiative, co-led by [EPFL]() and [ETH Zurich]().
- 2023: *"Controlled Training Data Generation with Diffusion Models"* paper @ [VILAB](). (advised by [Amir Zamir]())

IBM Research **February 2023 - September 2023**
Research Intern (advised by [Mattia Rigotti]() & [Amir Zamir]()) *Zurich, Switzerland*
- **ICLR 2024 first-author paper** - "*Unraveling the Key Components of OOD Generalization via Diversification*"
- Research on synthetic data & algorithmic design for out-of-distribution (OOD) generalization of neural networks

G-Research **July 2022 - September 2022**
Quantitative Research Intern *London, UK*
- Classic quant work. Predict things in financial markets. *Skills*: Scala, Spark, Python, ML, Algorithmic Design

PUBLICATIONS

[Controlled Training Data Generation with Diffusion Models]() **In review, 2025**
T. Yeo*, A. Atanov*, **H. Benoit^**, Aleksandr Alekseev^, Ruchira Ray, Pooya Akhoondi, Amir Zamir

[Unraveling the Key Components of OOD Generalization via Diversification]() **ICLR, 2024**
H. Benoit*, L. Jiang*, A. Atanov*, O. Kar, M. Rigotti, A. Zamir

SOFTWARE & PROJECTS

1st place at LLM training hackathon ([Github]() / [Details]()) **April 2024**
I won $1.5k and a trip to Huawei HQ in China. The competition was to train the best performing language model on [Slimpajama]() with 3 hours of compute on a A100 GPU.

Diffusion Transformer for Protein Structure Modeling ([Github]()) **May 2024**
Trained a protein structure diffusion model from scratch.

Misc: [RL library for civil engineering](), *contributed to an ecotoxicology paper, ...*

EDUCATION

Swiss Federal Institute of Technology (EPFL) **2021 - 2023**
MSc of Data Science *Lausanne, Switzerland*
- 5.8/6.0 GPA (ranked 3rd in year)

Swiss Federal Institute of Technology (EPFL) **2018 - 2021**
BSc of Computer Science / Communication Systems *Lausanne, Switzerland*
- 5.5/6.0 GPA

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; SQL; Scala; Java; C++; Go; Git; Linux; Bash; Docker.