

# Xavier Theimer-Lienhard

Lausanne, Switzerland

[xavier.theimer-lienhard@epfl.ch](mailto:xavier.theimer-lienhard@epfl.ch)

+41799006846

## Formation

---

### École Polytechnique Fédérale de Lausanne (EPFL)

Master in Data Science

Bachelor in Communication Systems (completed in 2022)

sep 2018 - now

Lausanne, Switzerland

Lausanne, Switzerland

## Research / Work Experience

---

### Post-Graduate Scholar @ [Yale](#) (sep 2024 - now)

Synthetic data lead in the Meditron project. **Team size:** 20.

### Research Assistant @ [EPFL MLO Lab](#) & [Yale LIGHT Lab](#) (feb 2024 - now)

Working on Meditron-ddx, a Medical LLM assistant, next iteration of open source breakthrough [Meditron-70B](#). Created a multiagent LLM framework. Exploring Tree of thoughts and other methods. Models finetuning for proof of work. **Team size:** 7-10.

### Research Assistant @ [EPFL MLO Lab](#) & [Yale LIGHT Lab](#) (sep 2023 - jan 2024)

Worked on Mini-Meditron, a low scale Medical LLM assistant for low resource settings. Created a new dataset using LLM generation for finetuning. Finetuned Pre-trained models (Mistral-7B) with generated data and DPO. **Team size:** 3-4.

### Software dev & Data scientist Part-time @ [Irbis Consulting SA](#) (sep 2023 - now)

#### Intern (jul 2023 - aug 2023)

Created a scraper for data retrieval that uses Deep Learning with Keras and TensorFlow and archives projects to help employees and associates for future projects.

Created an application for assisting creation of appels d'offres (documents) that helps employees gain time via automation.

Listening to the feedback of the users and heavy collaboration with the team to understand needs for the project. **Team size:** 1-3.

## Publications

---

### Meditron paper (TBD)

### GPoet: a Language Model Trained for Rhyme Generation on Synthetic Data

@ LaTeCH-CLfL 2023 ([paper](#))

Fine-tuned a GPT-2 model to address text constraints in the context of poem generation.

Achieving up to 60% rhyming lines in generated samples compared to 11% prior finetuning.

## Skills

---

**Machine Learning :** Natural Language Processing (NLP), Transformers, Reinforcement Learning.

**Software Engineering** : Python, Pytorch, Spark, Scala, Docker, Database, Git, C++, Java, Electron framework, Tensorflow, Linux

**Web Dev** : HTML, CSS, Javascript.

**Engineering** : Signal Processing, Computer Security, Communication Systems