

Diego DORN

 cv@ddorn.fr

 cozyfractal.com

 github.com/ddorn

Diego finishes his master in August 2024, with professional expertise in **software engineering** and **teaching**, especially in the measurement and mitigation of **risks** from **artificial intelligence** systems.



WORK EXPERIENCE

Paris 

Research engineer at EffiSciences
Feb. 2024 – now *Automated supervision of LLM-agents, design a benchmark to evaluate detection of out-of-distribution failure modes by monitoring systems.*

Cambridge 

Research assistant, Machine Learning Group, Cambridge University
July – Sep. 2023 *Research on goal misgeneralisation with N. Alex and D. Krueger*
“Goal Misgeneralization as Implicit Goal Conditioning” in the GCRL workshop at Neurips 2023

Berlin 

Teacher at ML4Good, a summer school on AI safety
August 2023 *Delivery and improvement of 10 days of technical and conceptual content. 21 participants*

Lausanne 

Lead developer for SPRIG (sprigproofs.org)
2022 – 2023 *Developing a distributed platform to increase confidence in mathematical proofs*

Lausanne 

Teaching assistant at EPFL
2019 - 2021 *TA for 8 courses for 1st, 2nd and 3rd year bachelors: Analysis (real, vectorial, complex), C++, mathematical logic, computer science basics*
Game development & small projects (cozyfractal.com/showcase)
Creation of 10+ small games under strong time constraints for jams, a 2D EsoLang (Asciidots)...

VOLUNTEERING

Lausanne 

Founder of the Safe AI Lausanne student association
2022 – 2024 *Events on reducing systemic and catastrophic risks from AI. Organisation of a 10-day bootcamp, talks and a reading group. Moderation of two round table discussions.*

Lausanne 

Vice-president, then advisor for Effective Altruism Lausanne
2022 – 2023 *Association aiming to find the best ways to help others and put them into practice*

Lausanne 

Co-founder of Chocopoly, the hot chocolate association of EPFL
2021 – 2023 *Followed by 400+ students, collaborated with 19 associations and served 1288L of hot chocolate*

Lausanne 

President of CQFD
2020 – 2021 *The mathematics students association of EPFL*

Many places 

Member of the national organisation committee of the TFJM²
2020 – 2021 *The french tournament of young mathematicians. Coordination of 9 events across France, development of a new online infrastructure and communication*

EDUCATION

Lausanne 

Master's at EPFL in Communication Systems, minor in Mathematics
2021 – present *Focus on artificial intelligence, formal verification and complexity theory*

Interlaken 

Summer school “Science and Policy – How to bridge the gap?”
July 2023 *Topics: science for policy, science communication, open science, Swiss policy landscape*

London 

ARENA, Alignment Research Engineer Accelerator
May – June 2023 *6 weeks intensive training on interpretability, RL and training at scale*

Lausanne 

Semester projects in Mathematical Logic and Game Theory
2021 – 2022 *Guided research under Jacques Duparc's supervision*

“Infinite games in the Baire space”, Bachelor thesis, Spring 2021

“Between decidable logics: ω -automata and infinite games”, Master's semester project, Spring 2022

Lausanne 

Bachelor's in Mathematics at EPFL
2018 – 2021 *Passed with a 5.42/6 average and top 5/100 of my year.*

SKILLS

- **Programming** Main hobby for the 10 last years. Many projects can be seen at cozyfractal.com/showcase
 - **Python (6000h)** Some of the modules I enjoyed using in more than 2 projects each include: asyncio, click, einops, fastAPI, flask, jaxtyping, joblib, huggingface, kivy, matplotlib, moderngl, mypy, numba, numpy, pillow, plotly, poetry, pre-commit pygame, pytest, pytorch, stable_baselines3, streamlit, transformer_lens, typeguard
 - **Rust (300h), Scala (200h) and C/C++ (300h)**
 - **JavaScript / CSS / HTML (500h)** Also using, VueJS, TailwindCSS, typescript
 - **Other languages** LaTeX (200h), Typst, 6502/NES assembly, Haskell, Matlab, Lean
 - **Tools** Vim, Jetbrains IDEs, VS Code, git, Docker, slurm, runAI, inkscape, OBS, Google Suite, ArchLinux (i3wm/sway)...
- **Collaboration:** Non-violent communication
- **Languages:** French (native), English (fluent), Italian & German (basics, willing to learn more)