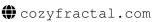
Diego DORN

✓ cv@ddorn.fr







EDUCATION

Master's in Communication Systems and minor in Mathematics at EPFL Lausanne 1 Focus on artificial intelligence, formal verification and complexity theory 2021 - present Summer school "Science and Policy – How to bridge the gap?" Interlaken 🕶 Topics: science for policy, science communication, open science, swiss policy landscape July 2023 ARENA, Alignment Research Engineer Accelerator (arena.education) London 6 weeks intensive training on interpretability, RL, training at scale + team projects May - June 2023 **Bachelor's in Mathematics at EPFL** Lausanne • Passed with a 5.42/6 average and top 5/100 of my year. 2018 - 2021

WORK EXPERIENCE

Research engineer at EffiSciences

Automated supervision of LLM-agents

February - present Cambridge **S**

Research assistant in the Machine Learning Group, Cambridge University

July – September 2023

Research on goal misgeneralisation with N. Alex and D. Krueger - Published "Goal Misgeneralization as Implicit Goal Conditioning" in the Goal-Conditioned Reinforcement Learning workshop at Neurips 2023

Teacher at ML4Good, a summer school on AI Safety

Delivery and improvement of 10 days of technical and conceptual content. 21 participants

Lead developper for SPRIG (sprigproofs.org)

Lausanne

Developing a distributed platform to increase confidence in mathematical proofs

2022 - 2023

Berlin ==

August 2023

Paris |

Semester projects in Mathematical Logic and Game Theory

Lausanne •

Guided research under Jacques Duparc's supervision

2021 - 2022

- "Infinite games in the Baire space", Bachelor thesis, Spring 2021

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

Teaching assistant at EPFL

Lausanne •

TA for 11 courses (8 different) for 1st, 2nd and 3rd year bachelors:

2019 - 2021

- Analyse I (real 1D analysis)

- Information, Calcul et Communication

- Analyse II (vectorial analysis)

- *Programation orientée objet* (C++ OOP)

- Analyse III (complex analysis)

- MOOC Introduction to C++

- Mathematical logic

MOOC Introduction to OOP in C++

Game development & small projects (cozyfractal.com/showcase)

Creation of 10+ small games under strong time constraints for jams, a 2-dimensional EsoLang (Asciidots)...

Volunteering

President of COFD

Founder of the Safe AI Lausanne student association

Lausanne **±**

Group for making AI beneficial and reducing related catastrophic risks

2022 - present

- Organisation of a 10-day bootcamp on AI Alignment, talks and a reading group

Vice-president, then Advisor of Effective Altruism Lausanne

Lausanne 🚹

Association aiming to find the best ways to help others and put them into practice

2022 - present

Co-founder of Chocopoly, the hot chocolate association of EPFL

Lausanne **±**

Followed by 400+ students, collaborated with 19 associations and served 1288L of chocolate 2021 - 2023

Lausanne +

The mathematics students association of EPFL

2020 - 2021

Member of the national organisation committee of the TFJM²

Many places

The french tournament of young mathematicians

2020 - 2021

- Coordination of 9 events across France, development of a new online infrastructure and external communication

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, 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 languagues LATEX (200h), 6502/NES assembly, Haskell, Matlab, Lean
- **Tools** Vim, Jetbrains IDEs, VS Code, git, Docker, slurm, runAI, inkscape, OBS, Google Suite, ArchLinux (i3wm/sway)...
- Languages: French (native), English (fluent), Italian & German (basics, willing to learn more)
- Others: climbing (since 2006), juggling (on a bike), cooking (vegan), hot chocolate