

Milan Weibel

Mission

I've been following the development of LLMs and model alignment techniques for several years. I lead AI Safety @ UC, organizing reading groups each semester. I'm looking to contribute to the project of ensuring that the impact of transformative AI on humanity is positive. I see particular promise in the formal verification of structured CoT.

Education

- UC Chile — 2021-current — B.Sc. Computer Science Engineering, minors in management and political science.
- [ML4Good Bootcamp Brasil 2024](#)

Sample Courses

- IIC2613 Artificial Intelligence — 2023-1
- ICP0107 Public policy formulation — 2023-1
- IIC3697 Deep Learning — 2025-1

Teaching

- IIC1103 Introduction to Python programming — Teaching Assistant
- [AI Safety Fundamentals @ UC Chile](#) — Program lead, session facilitator

Writing Samples

- [ChatGPT understands, but largely does not generate Spanglish \(and other code-mixed\) text](#)
- [A systematic nomenclature for the typological classification of state strength profiles](#)
- [No-self as an alignment target](#)

Development Technologies

[Python](#), [PyTorch](#), [einops](#), [Selenium](#), [Git](#), [Nix](#), [Docker](#), [Bash](#), [C](#), [SQL](#), [Ruby on Rails](#), [React](#).

Languages

- English — Advanced. Native-equivalent proficiency.
- Spanish — Native.
- French — Intermediate.