Lingnan University 嶺南大學 香港 Hong Kong
Hong Kong Catastrophic Risk Centre

Hong Kong Catastrophic Risk Centre
Hybrid Research Seminar

A Thousand AI Constitutions

Professor Simon GOLDSTEIN
The University of Hong Kong

Today, each AI lab has its own model spec, or constitution. These documents define the values that the labs intend their AIs to have, and they are used in post-training to instill those values. This talk argues that the current approach is wrong. Rather than a single constitution, reflecting a single set of moral values, each frontier AI lab should create many different kinds of AIs based on many different constitutions reflecting many sets of values. We give four arguments for constitutional diversification. Diversification mitigates risk, increases political legitimacy, unlocks emergent value, and avoids value lock-in.

Venue
LKK203

Meeting ID: **983 1951 5823**
Passcode: **26167445**

Scan to join
zoom

More Info: https://www.ln.edu.hk/philoso/events
Enquiries: 26167445 / dphilo@ln.edu.hk

28 Sept 2026
(Monday)
4:45-6:15 pm
(HKT)