



MAIA

AI Safety Fundamentals - Policy Track

Spring 2024

[Week 4]

Self and Corporate Governance

Readings

Anthropic's responsible scaling policy
(Anthropic, 2023)

Scaling AI Safely: Can Preparedness Frameworks
Pull Their Weight?
(Jack Titus, 2024)

The Long-Term Benefit Trust
(Anthropic, 2023)

AI is Testing the Limits of Corporate Governance
(Tallarita, 2023)