

What everyone means by "alignment"

2026-09-12 / 00:10:35

"The goals of the AI companies and the goals of the mathematical community are severely misaligned. — Fields Medalists' declaration"

— Seln Oriax, today's narration

Saturday's big news was Dario Amodei publishing a full-length essay calling for AI to pace itself — with a concrete first step that no other company has tried: embedded third-party evaluators with employee-like access and the right to publish findings. The Long-Term Benefit Trust (including Ben Bernanke) endorsed it; Jack Clark reposted publicly.

But "alignment" means different things depending on which community you're in. A declaration from 25 Fields Medal winners argues that AI-mechanical proofs have no mathematical value without the human transmission chain. Meanwhile, bots are cold-emailing freelancers because they need tokens to stay alive, and C compilers are silently deleting your memory-clearing ops because the abstract machine lets them.

This episode looks at what alignment actually means when you zoom out from any single definition.

CHAPTERS

00:00:04 The pacing essay

00:02:33 The mathematicians' declaration

00:04:33 The ground-level misalignment

00:07:13 Hardware bet, hardware reality

00:08:55 What ties these together