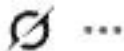




Jacob Coxon ✓

@hilbertspaess



I resigned from Anthropic today. I spent the last three years doing pretraining research at both OpenAI and Anthropic. Neither company is acting responsibly. They are racing straight to self-improving superintelligence and gambling with our lives. More thoughts below.

5:04 PM · Sep 8, 2026 · **70.1M** Views



Jacob Coxon ✓

@hilbertspaess



Do not underestimate the power of this technology. These will soon be superhuman systems that can hack anything, revolutionize any field overnight, and acquire real power and resources. We have all witnessed the progress in each of these domains, and progress is not slowing.

5:04 PM · Sep 8, 2026 · **6.1M** Views



Jacob Coxon ✓

@hilbertspaess



The people building AI earnestly believe that it could kill us all by the end of the decade. This is not a marketing stunt. If anything, many executives and senior researchers will couch their phrasing in the press to sound sensible - but I hear the same people express fear privately. No other human activity poses this level of danger.

5:04 PM · Sep 8, 2026 · **13.4M** Views



Jacob Coxon ✓

@hilbertspaess



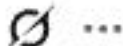
A common response is “if they truly believe this, why are they still building it?” At OpenAI, many have not deeply internalized the civilizational stakes. At Anthropic, the stakes are well-understood, but they are locked in a race to get there first - they believe no one else will act responsibly, so they must do it themselves, despite the risk.

5:04 PM · Sep 8, 2026 · **4.7M** Views



Jacob Coxon ✓

@hilbertspaess



Accepting this race and entering the “endgame” is a hubristic gamble that should not be launched from a private company’s Slack. Attempting to speedrun alignment should require extraordinary confidence that there are no better trajectories available.

5:04 PM · Sep 8, 2026 · **4.2M** Views



Jacob Coxon ✓

@hilbertspaess



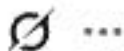
I am optimistic about the potential for coordination. Warning shots like the Hugging Face attack have made pacing agreements between U.S. labs more viable. I don't feel like we're on track to prevent a global race, which may require costly actions such as a temporary ban on improving model capabilities.

5:04 PM · Sep 8, 2026 · **3.8M** Views



Jacob Coxon 

@hilbertspaess



If you are a lab researcher, I urge you to consider what the next few years will actually feel like. Do you want to kick off a superintelligent RL run without a rigorous understanding of its mind? Should you put your head down because “it’s happening anyway” - or take this moment to call for different conditions?

5:04 PM · Sep 8, 2026 · **3.5M** Views