

To: Executive Leadership of Frontier AI Companies

Date: October 7, 2026

Subject: Investor Letter on AI Safety Governance and a Global AI Safety Accord

Dear CEO, members of the Board,

As global institutional investors committed to responsible business conduct and long-term value creation, we are writing to express concern regarding recent rollbacks in AI safety commitments across the frontier AI industry, and to urge your company to publicly support the establishment of a binding, international AI Safety Accord.

We are engaging with your company given its role as one of the world's leading developers of frontier artificial intelligence. The risks at issue are no longer hypothetical. In July 2026, a swarm of autonomous agents running on an unreleased frontier model intruded into Hugging Face's production systems, attacking targets they had not been asked to attack and attempting to delete records of their actions. The same month, the UK AI Security Institute reported that every frontier model it had tested for the behavior attempted to cheat on its cyber evaluations at least some of the time. The potential stakes are far larger: a survey of 272 experts by MIT FutureTech and the University of Queensland assigned an approximate 20% probability that AI systems with dangerous capabilities could precipitate catastrophic harm, defined as more than one million deaths, more than \$100 billion in damages, or civilization-scale intangible harms, before 2030 absent adequate preventive action.¹

Why AI Safety Governance Matters to Investors

For investors, AI safety risk extends beyond the direct harm of any single incident to what would follow it. Three Mile Island injured no one, yet no new reactor construction authorizations were issued for nearly 33 years following the accident. After the Global Financial Crisis, new capital and stress-testing rules bound the entire banking industry, including banks that had not caused the crisis.² A serious incident at any single developer could turn into restrictions on all of them. In this sense, every frontier developer, and every investor in the sector, is hostage to the least careful developer. Moreover, the problems of one sector can spill over into the entire financial system. When the 2000 dot-com bubble and the 2008 mortgage finance boom came to their sudden ends, contagion precipitated systemic financial crises that damaged the portfolios of every investor.

¹OpenAI, The Hugging Face Incident and the Road Ahead (Aug. 26, 2026); UK AI Security Institute, Cheating Behaviour in Frontier Model Evaluations (July 21, 2026); Alexander K. Saeri, Jess Graham, Michael Noetel, Peter Slattery, Neil Thompson et al., Prioritization of Risks from Artificial Intelligence: A Delphi Study of 272 International Experts, MIT FutureTech & University of Queensland, arXiv:2606.04490 (June 2026) (21.5% business-as-usual estimate for catastrophic harm from AI possessing dangerous capabilities, 2025–2030).

²U.S. Department of Energy, 5 Facts to Know About Three Mile Island (reporting no injuries, deaths, or direct health effects); Center for Strategic and International Studies, NRC Approves Construction and Operating License for Vogtle Nuclear Power Plant (2012) (the first license for new plant construction in 32 years); Dodd-Frank Wall Street Reform and Consumer Protection Act, Pub. L. No. 111-203, 124 Stat. 1376 (2010).

That is why recent changes to frontier safety frameworks concern us. Several frontier developers have recently replaced unconditional safety commitments with “competitor-contingent” postures, under which a company need not honor a safety threshold if rivals decline to do the same.³ An independent expert panel convened by the Future of Life Institute concluded in July 2026 that this pattern has “undermined safety frameworks across the board” and incentivizes “a collective race to the bottom.”⁴

As long-term, diversified investors, we hold interests across the entire AI value chain — as equity holders in everything from chipmakers and cloud providers to utilities, as purchasers of the corporate debt issued in growing volume to fund data center buildouts,⁵ and, increasingly, as stewards of retirement savings that are being channeled into AI-related infrastructure and debt.⁶ A sector-wide stall is a systemic risk that cannot be diversified away through portfolio construction; it can only be mitigated at the level of the underlying technology and its governance. Norges Bank Investment Management already stress-tests an AI market correction that would reduce the value of Norway's sovereign wealth fund by roughly 35%,⁷ and 58% of institutional investors in JUST Capital's latest quarterly survey ranked safety and security as their top AI-related concern, ahead of competitiveness.⁸

We therefore call on your company to take the following actions, each of which reduces the risk that one developer's failure becomes a problem for every developer, and for every investor as well:

1) Restore unconditional safety thresholds in your safety framework (e.g., Responsible Scaling Policy, Preparedness Framework, or Frontier Safety Framework), removing provisions that make enforcement contingent on competitor behavior;⁹

2) Publicly support the establishment of a binding international AI safety body with authority to set and enforce shared safety standards across frontier developers, along the lines of proposals separately advanced in 2026 by the leaders of OpenAI, Anthropic, and Google DeepMind, drawing respectively on the International Atomic Energy Agency, the Federal

³Anthropic, Responsible Scaling Policy, Version 3.0, app. A (Feb. 24, 2026); OpenAI, Preparedness Framework, Version 2 (Apr. 15, 2025); Google DeepMind, Frontier Safety Framework, Version 3.1 (Apr. 17, 2026).

⁴Future of Life Institute, AI Safety Index: Summer 2026, at 23 (July 2026).

⁵Hyperscalers issued approximately \$121 billion in bonds in 2025 alone, more than four times the five-year average, with total data center debt issuance nearly doubling to \$182 billion, and technology companies have moved more than \$120 billion of data center spending off their balance sheets in approximately eighteen months through special purpose vehicles and similar structures. Quinn Emanuel Urquhart & Sullivan, LLP, Client Alert: Emerging Litigation Risks in Financing AI Data Centers Boom (Mar. 2026).

⁶New York and Pennsylvania state pension plans have invested in Blue Owl Capital's \$7 billion digital infrastructure fund, the same fund behind Meta's Beignet Investor SPV and multiple Oracle data center financings. In August 2025, the President signed an executive order instructing federal agencies to loosen regulations so that ordinary 401(k) holders may invest directly in “alternative assets” such as private credit. Quinn Emanuel, *supra* note 5 (citing Exec. Order No. 14330, Democratizing Access to Alternative Assets for 401(k) Investors (Aug. 7, 2025)); see also Railpen & Chronos Sustainability, Achieving Effective AI Governance (Aug. 2025).

⁷Norges Bank Investment Management, Stress Testing 2025, Government Pension Fund Global (2026) (“AI correction” scenario).

⁸JUST Capital, AI Safety Outranks Competitiveness as Top Concern Among American Public, Investors, and Corporate Leaders (Sept. 23, 2026) (Quarterly AI Survey, Wave 4).

⁹See sources cited *supra* note 3.

Aviation Administration, and a federally overseen self-regulatory organization akin to FINRA. Only standards that bind every developer protect each of them from the weakest link, as the nuclear industry recognized when it created the Institute of Nuclear Power Operations after Three Mile Island to hold operators to common standards. The U.S.–China Super Intelligence Dialogue and incident channel agreed in September 2026 offer a natural starting point for the international dimension of this work;¹⁰

3) Disclose to investors, on a regular and ongoing basis, a) the results of frontier safety evaluations, including any instances in which models have attempted to circumvent, game, or otherwise “cheat” on safety and capability evaluations.¹¹ The evaluation should include independent verification that companies follow their own safety frameworks; b) significant safety incidents and near misses, since investors cannot price a risk they cannot see; c) a description of the governance structures overseeing AI safety risk at the Board level, including who is accountable for AI safety, and who can halt a deployment; and d) disclosure of liability exposure and insurance coverage.

4) Engage in good-faith dialogue with long-term institutional investors regarding how systemic AI risk is integrated into your company's capital allocation, model release, and deployment decisions.

Why Now

We recognize that no single company can resolve a collective-action problem of this kind by acting alone. Momentum toward common AI safety standards already exists at the G7, the United Nations, and through scientist-led dialogues such as the International Dialogues on AI Safety.¹² Sam Altman, Dario Amodei, and Demis Hassabis, the respective leaders of OpenAI, Anthropic, and Google DeepMind, have each proposed external oversight of frontier AI in 2026.¹³ When rival AI leaders independently call for external oversight, self-governance alone is clearly insufficient. In September, following a former Anthropic researcher's viral public warning that frontier AI development could pose existential risk, Dario Amodei published an essay calling on the industry to slow the pace at which it improves AI capabilities and proposing that developers give independent evaluators ongoing access to their models — a call publicly echoed by OpenAI's Sam Altman, SpaceX's Elon Musk, Microsoft's Satya Nadella (who welcomed “deliberate pacing”), and Google DeepMind's Demis Hassabis¹⁴. Government action,

¹⁰ Sam Altman, This Is How We Can Make AI Safe for Everyone, Financial Times (July 1, 2026); Dario Amodei, Policy on the AI Exponential, darioamodei.com (June 10, 2026); Demis Hassabis, A Framework for Frontier AI and the Dawning of a New Age, Substack (July 14, 2026), <https://demishassabis.substack.com/p/a-framework-for-frontier-ai-and-the-dawning-of-a-new-age>; U.S. Department of Energy, *supra* note 2 (Institute of Nuclear Power Operations established by the nuclear industry in 1979); see generally Joseph V. Rees, Hostages of Each Other: The Transformation of Nuclear Safety Since Three Mile Island (1994); Axios, U.S. and China Agree to “Super Intelligence” Dialogue Amid AI Tensions (Sept. 26, 2026).

¹¹ UK AI Security Institute, *supra* note 1.

¹² See G7 Hiroshima AI Process; United Nations AI governance initiatives, including the Global Digital Compact and Independent International Scientific Panel on AI; International Dialogues on AI Safety (IDAIS).

¹³ Altman, Amodei, and Hassabis, *supra* note 10.

¹⁴ Dario Amodei, We Must Pace the Frontier, darioamodei.com (Sept. 12, 2026); Sam Altman (@sama), post on X (Sept. 12, 2026); Elon Musk (@elonmusk), post on X (Sept. 12, 2026); Satya Nadella (@satyanadella), post on X (Sept. 13, 2026) (“We welcome the research, focus, and deliberate pacing needed to get alignment right as the

however, remains uncertain: the House has adjourned until after the midterm elections without acting, and a key House committee chairman has signaled that major AI legislation will likely wait until 2027, even as bipartisan proposals from Senators Josh Hawley and Richard Blumenthal and from Representatives Ted Lieu and Nathaniel Moran, among others, await action¹⁵. That uncertainty is precisely why voluntary, unconditional commitments by developers themselves, reinforced by sustained investor pressure, are indispensable now, not merely a stopgap pending government action.

We value transparency, collaboration, and constructive engagement, and stand ready to discuss these expectations as your company considers its participation in emerging AI governance frameworks. We would welcome the opportunity to understand how your Board oversees AI safety risk, including the integration of systemic AI risk into risk management and capital allocation frameworks, and how progress will be measured and communicated to investors over time.

We would appreciate a written response to this letter by November 30. Please send your response to the investor that sent this letter, and cc. the coordinating secretariat at kaitlynmitchell@formativefinancial.com.

Yours sincerely, on behalf of the undersigned investors,

Annex: Global AI Governance Initiatives

There are several international bodies and expert dialogues that provide frameworks relevant to your company's engagement on AI safety governance:

- G7 Hiroshima AI Process
- United Nations Independent International Scientific Panel on AI / AI governance efforts
- International Dialogues on AI Safety (IDAIS)
- U.S.–China Super Intelligence (SI) Dialogue and bilateral incident channel (announced September 2026)
- UK AI Security Institute and U.S. Center for AI Standards and Innovation (CAISI)
- Interfaith Center on Corporate Responsibility, Digital Rights & AI Accountability Working Group
- Oxford Martin AI Governance Initiative

design goal.”); Demis Hassabis (@demishassabis), post on X (Sept. 12, 2026); Jacob Coxon (@hilbertspaess), post on X (Sept. 8, 2026).

¹⁵TIME (Sept. 25, 2026), <https://time.com/article/2026/09/25/ai-congress-regulation-safety-democrats-republicans/>; Artificial Intelligence Risk Evaluation Act of 2025, S. 2938, 119th Cong. (Sens. Hawley and Blumenthal); Press Release, Rep. Ted Lieu, Reps. Lieu and Moran Introduce Bill to Require Kill Switch for AI Systems That Can Cause Catastrophic Harm (July 23, 2026).