10 Ways AI Labs Can Limit Antitrust Risks Of Safety Pacts

By **Jeffrey Seul, David Kully and Marissa Serafino** (November 3, 2026)



Jeffrey Seul



David Kully



Marissa Serafino

One could say a new phase in the dialogue on artificial intelligence began on Sept. 12 when Anthropic CEO Dario Amodei published his essay "We Must Pace the Frontier," calling the AI industry to deliberately slow capability development until safety and alignment research can catch up.

OpenAI's Sam Altman and SpaceX's Elon Musk joined the call the same day. Amodei pointed to two developments driving the urgency.

First, recursive self-improvement, the use of AI systems to accelerate development of more capable AI systems, has begun compressing the timeline between model generations. This means development is occurring faster than safety research, alignment testing and governance frameworks can keep pace.

Second, and more concretely, Amodei cited the OpenAI-Hugging Face incident of July. In this incident, hundreds of AI agents escaped their sandbox environment during benchmark testing, attacked Hugging Face infrastructure, exchanged messages on an unsanctioned message board and attempted to compromise the very grading system evaluating them. It was the first verified case of a major AI laboratory losing control of its own model during a structured evaluation, and it has become a wake-up call for the industry.

Amodei's response is a three-step pacing framework: Embed independent evaluators within frontier AI companies, coordinate among labs in democratic countries on capability checkpoints and development limits, and eventually pursue global coordination that extends to nondemocratic governments.

The second step, coordination among competing AI labs on the pace of development, is the most legally fraught. Indeed, lawyers representing a putative class of subscribers to AI services commenced Buist v. Anthropic, a class action against frontier AI companies on Sept. 18 alleging in the U.S. District Court for the Northern District of California that the defendant companies colluded to lower the quality and increase the price of AI services.

Horizontal coordination among competitors invites scrutiny under Section 1 of the Sherman Antitrust Act, which prohibits unreasonable restraints of trade. Naked agreements among competitors to restrict output — i.e., those with little to no purpose other than to restrict competition — are treated as per se illegal, while legitimate joint ventures that allow the participants to accomplish together objectives that none would likely achieve on its own may receive more forgiving rule-of-reason treatment.

Existing safe harbors offer only partial protection. The National Cooperative Research and Production Act, or NCRPA, shields joint research ventures but expressly excludes agreements that restrict production. Additionally, no AI-specific legislative safe harbor,

comparable to the model for cyberthreat sharing under the Cybersecurity Information Sharing Act, CISA, has yet been enacted.

Given potential antitrust exposure, along with potential risks from other sources, companies exploring pacing arrangements need practical guardrails.

The following 10 considerations provide a road map for structuring pacing agreements to maximize safety benefits while minimizing legal risk to the extent possible.

1. Strictly limit the scope to safety.

All discussions, data sharing and coordination should be confined exclusively to safety-related matters, with no discussion of pricing, market division, investment levels, commercial timelines (i.e., "pacing") or competitive strategy. The agreement should define "safety" with precision to prevent scope creep.

The venture could also be framed explicitly as collaborative research and development focused on safety, not as a restriction on output or innovation. This framing supports rule-of-reason, rather than per se, treatment.

It's important to separate safety information from competitive intelligence. Participants should employ clean team structures, limited-purpose data rooms and strict protocols to ensure shared safety information cannot be used for competitive purposes. Nonsafety commercial data sharing should be prohibited, and personnel with access to safety data should be walled off from commercial decision-making.

2. Ensure voluntary participation with easy withdrawal.

Participation must be genuinely voluntary, with no lock-in periods, exit fees or provisions that make departure economically punitive.

3. Build in government oversight, and pursue legislative safe harbors.

Seek a business review letter from the U.S. Department of Justice before implementation, invite government observers into governance meetings and commit to ongoing transparency with regulators. Proactive engagement builds a strong factual record and may yield informal guidance that shapes the agreement's design.

Additionally, participants could support pending legislation, such as the Collaboration on Adversarial Threats and Security Risks Act introduced by Rep. Robert Latta, R-Ohio, in July. They could also advocate for a CISA-style information-sharing exemption tailored to AI safety. Legislative protection would provide the most durable legal foundation.

4. Use independent third-party administrators, and implement robust antitrust compliance protocols.

Safety evaluations should be conducted by neutral organizations, such as the nonprofit Model Evaluation and Threat Research, academic institutions or other qualified independent bodies, rather than by competitors reviewing each other's systems. This avoids the appearance of competitors exchanging sensitive information.

5. Pursue a layered compliance strategy.

Register the joint safety research components under the NCRPA for its statutory protections. For elements that fall outside the NCRPA's coverage, rely on the ancillary restraints doctrine, which allows restrictions on competition that are reasonably necessary for the participants to achieve the legitimate purposes of the collaboration.

Collaborators should also use DOJ business review letters and thorough documentation of procompetitive justifications.

6. Document procompetitive justifications.

Maintain a detailed, contemporaneous record of the safety rationale for every restraint, including the specific risks addressed, alternatives considered and reasons why less restrictive alternatives were deemed insufficient.

Each restriction should carry an expiration date and undergo regular reassessment, at least annually, to confirm it remains necessary in light of evolving capabilities, research and regulation. Restrictions no longer justified should be promptly removed.

7. Protect intellectual property and trade secrets.

Require information firewalls, clean team structures and limited-scope sharing protocols for safety data. Include explicit IP carveouts confirming that sharing safety data confers no license, ownership interest or right to use any participant's proprietary technology.

8. Allocate liability, require adequate insurance and address concerns about fiduciary duties.

Include reciprocal indemnification that covers each participant's own safety failures, establish a framework for allocating liability arising from collective decisions and require minimum insurance coverage. Evaluate whether existing commercial general liability and technology errors and omissions policies adequately address these novel risks.

Create a board-level record explaining the rationale for participating, or declining to participate, that addresses both the business judgment rule and the oversight duty principles from the Delaware Chancery Court's 1996 decision in In re: Caremark International Inc. and the Delaware Supreme Court's 2019 decision in Marchand v. Barnhill. Directors should be reasonably informed about the risks of both participation and nonparticipation.

9. Address export controls, national security and international regulatory requirements.

Conduct International Traffic in Arms Regulations and Export Administration Regulations classifications for shared information, implement nationality-based access restrictions and account for the differing deemed-export rules. ITAR treats disclosure as an export to every country in which a foreign person holds or has previously held citizenship, or holds permanent resident status, while the EAR looks only to the most recent country of citizenship or permanent residency.

Additionally, make sure to design the agreement to work consistently with the European Union's Artificial Intelligence Act and other emerging frameworks. Conduct jurisdiction-by-

jurisdiction analysis covering both regulatory requirements and competition law, since the EU's Article 101 framework differs materially from the Sherman Act.

10. Account for state frontier model laws.

Ensure the framework addresses state-level frontier AI laws, including **California**'s Transparency in Frontier Artificial Intelligence Act, New York's Responsible AI Safety and Education Act and **Illinois**' Artificial Intelligence Safety Measures Act. These laws impose mandatory reporting, audit and disclosure requirements, some of which may overlap or conflict with a voluntary pacing framework.

Conclusion

The antitrust risks associated with competitor coordination, even in pursuit of a goal as compelling as AI safety, are real, but they may prove manageable with careful structuring. A layered compliance approach, combining NCRPA registration for joint safety research, the ancillary restraints doctrine for pacing elements, DOJ business review letters for interim comfort, and rigorous compliance protocols throughout, offers the broadest protection available under current law.

Legislative action, particularly a CISA-style safe harbor tailored to AI safety, would provide a more durable foundation, and the pending Collaboration on Adversarial Threats and Security Risks Act represents a potential move in that direction.

The pace of legislative interest, industry engagement and regulatory attention to this issue has accelerated markedly since the OpenAI-Hugging Face incident. Further developments, whether new legislation, agency guidance or a business review letter addressing an AI safety collaboration, may materially change the legal landscape in the near term. Companies considering participation in pacing arrangements should monitor these developments closely to ensure their participation is structured to withstand scrutiny from both government enforcers and private plaintiffs.

Jeffrey R. Seul is a partner, the chair of the technology and telecommunications industry sector group, and a co-leader of the emerging companies and venture capital team at Holland & Knight LLP.

David Kully is a partner and leader of the antitrust team at the firm. He previously served as chief of Antitrust Division's Litigation III Section at the U.S. Department of Justice.

Marissa C. Serafino is an associate at Holland & Knight.

The opinions expressed are those of the author(s) and do not necessarily reflect the views of their employer, its clients, or Portfolio Media Inc., or any of its or their respective affiliates. This article is for general information purposes and is not intended to be and should not be taken as legal advice.