Congressional Research Service
Informing the legislative debate since 1914

Legal Sidebar

Artificial Intelligence Safety Collaboration and Antitrust Law

September 28, 2026

On September 12, 2026, Anthropic chief executive Dario Amodei published a blog post arguing that the artificial-intelligence (AI) industry should slow the pace at which it develops frontier models to mitigate risks of catastrophic harm, including the loss of control of AI systems and misuse of AI for cyberattacks and bioterrorism. Amodei also proposed several measures to "pace the frontier," including coordination among AI companies to establish common safety standards and "limits on the rate of unchecked AI progress." He added, however, that "[s]ome forms of coordination that would be impactful for pacing are legally challenging, and will require government support," potentially including "waivers of antitrust restrictions."

Amodei's post—which was published amid heightened public discussion of potential AI safety risks—generated considerable commentary. Leaders of other AI labs endorsed his call to "pace the frontier." House Minority Leader Hakeem Jeffries called on Congress to take "decisive action" to slow AI development in the interest of safety. Other observers criticized Amodei's proposals. Some contended that antitrust law permits AI firms to engage in many forms of safety collaboration, making antitrust waivers unnecessary and potentially harmful. Amodei's post also generated antitrust litigation: On September 18, 2026, customers of several leading AI companies filed a putative class action alleging that the companies violated Section 1 of the Sherman Act by agreeing to slow the pace of AI innovation.

This Legal Sidebar discusses antitrust issues raised by the prospect of AI safety collaboration. It begins with an overview of the antitrust principles that govern competitor collaboration. Next, it applies those principles to evaluate the legal risks that may accompany certain types of safety collaboration between competing AI developers. The Sidebar concludes with considerations for Congress.

Congressional Research Service
https://crsreports.congress.gov
LSB11485

Competitor Collaboration and Antitrust Law: General Principles

Per Se Illegality, the Rule of Reason, and "Quick-Look" Review

Section 1 of the Sherman Antitrust Act prohibits "every" contract, combination, or conspiracy "in restraint of trade." Despite this categorical language, the Supreme Court has interpreted Section 1 to bar only *unreasonable* restraints of trade that harm competition. Applying this general standard, the Court has identified some types of agreements that are so likely to be anticompetitive that they are deemed per se illegal, meaning courts need not inquire into their effects in individual cases. Restraints in this category include agreements among competitors to fix prices, divide markets, and limit output.

While some types of agreements are per se illegal under Section 1, most restraints are evaluated using a standard called the rule of reason. Under the rule of reason, courts conduct fact-specific assessments of market power and the details of a challenged agreement to determine its competitive effects.

This inquiry typically proceeds using a three-step burden-shifting framework. In that framework, the plaintiff has the initial burden to prove that the challenged restraint has a substantial anticompetitive effect, such as higher prices, reduced output, or diminished innovation.

If the plaintiff makes a prima facie case of anticompetitive harm, the burden shifts to the defendant to establish a procompetitive justification for the challenged restraint. For example, a defendant might argue that the restraint increases output, creates operational efficiencies, makes a new product available, enhances product quality, or broadens consumer choice. Cognizable procompetitive justifications must involve a restraint's effects on *competition*; typically, the argument that an agreement promotes other social values will not qualify. In *National Society of Professional Engineers v. United States*, for example, an engineering trade association attempted to justify a provision in its canon of ethics prohibiting competitive bidding by arguing that price competition would lead to inferior engineering work and endanger public safety. The Supreme Court rejected that argument, explaining that the Sherman Act "reflects a legislative judgment that ultimately competition will produce not only lower prices, but also better goods and services." That statutory policy, the Court reasoned, "precludes inquiry into the question whether competition is good or bad" in particular cases.

If the defendant establishes a procompetitive justification for a restraint, the burden shifts back to the plaintiff to show that the procompetitive benefits could be reasonably achieved through less anticompetitive means. Some decisions have held that, if a plaintiff fails at this third step, the inquiry proceeds to a fourth step in which the court balances a restraint's anticompetitive and procompetitive effects.

While most agreements are evaluated under the rule of reason, courts have recognized a standard that lies between the full rule of reason and per se illegality. Courts have employed this intermediate standard—often called "quick-look" review—to evaluate restraints that resemble per se illegal conduct but involve factors that warrant additional examination. The principles for determining when quick-look review applies are not entirely clear, and courts have characterized quick-look review in different ways. The basic idea is that, under the quick-look standard, plaintiffs can discharge their initial burden without the detailed evidence of competitive harm—such as proof of market power—required under the full rule of reason. Defendants in quick-look cases have the opportunity to offer procompetitive justifications for their conduct, distinguishing quick-look review from per se rules. If the defendant in a quick-look case offers a sound procompetitive justification, the court proceeds to evaluate the overall reasonableness of the restraint.

Competitor Collaboration

As the above discussion suggests, the legal standard that applies to competitor collaboration varies based on the details of the collaboration. The federal antitrust agencies—the Department of Justice (DOJ) and Federal Trade Commission (FTC)—have traditionally taken the view that joint ventures (JVs) involving research and development (R&D) typically are subject to the rule of reason. In their 2000 *Antitrust Guidelines for Collaborations Among Competitors* (the 2000 Guidelines), the agencies adopted that position, explaining that most competitor collaborations involving R&D are procompetitive and that such collaborations may enable participants to more efficiently develop or improve goods, services, or production processes.

The antitrust agencies withdrew the 2000 Guidelines in December 2024. In doing so, the agencies explained that, while specific aspects of the 2000 Guidelines may accurately reflect the current state of the law, the guidelines no longer provided reliable guidance given subsequent legal and technological developments. While the 2000 Guidelines are thus no longer operative, case law supports the proposition that R&D JVs are generally analyzed under the rule of reason.

Congress has also taken steps to mitigate the antitrust risks raised by R&D JVs. The National Cooperative Research and Production Act of 1993 (NCRPA) provides that certain R&D-related JVs are not per se illegal and are instead to be evaluated based on their "reasonableness, taking into account all relevant factors affecting competition." Under the NCRPA, participants in covered JVs may further limit their antitrust exposure by filing a notification disclosing certain information about a JV with the DOJ and FTC. Parties that file notifications in conformity with the NCRPA limit their antitrust exposure under federal and state law to actual damages plus costs and attorneys' fees, as opposed to the treble damages typically available under the Sherman Act.

In addition to R&D JVs, the NCRPA applies to certain information-sharing collaborations, provided they do not involve competitors exchanging information relating to costs, sales, profitability, prices, marketing, or distribution if such information is not reasonably required to carry out the purpose of the JV. Under the case law, information sharing standing alone (i.e., unconnected to a broader conspiracy) is generally analyzed under the rule of reason, even if it involves the sharing of price information.

The enforcement posture of the antitrust agencies has varied based on the type of information shared among competitors. The 2000 Guidelines explained that the sharing of information regarding price, output, costs, or strategic planning is more likely to raise competitive concerns under the rule of reason than the sharing of less competitively sensitive information. The 2000 Guidelines also said that the sharing of information on current operating and future business plans is more likely to be anticompetitive than the sharing of historical information.

Information sharing regarding cybersecurity enjoys special statutory protection; that protection is set to expire on December 11, 2026, absent legislative extension. The Cybersecurity Information Sharing Act of 2015 (CISA) includes an antitrust exemption allowing private entities to exchange or provide cyber threat indicators or defensive measures or assistance relating to the prevention, investigation, or mitigation of a cybersecurity threat. CISA excludes from this protection agreements to fix prices, allocate markets between competitors, monopolize or attempt to monopolize, boycott, or exchange price or cost information, customer lists, or information regarding future competitive planning.

The antitrust agencies have also issued guidance regarding the sharing of cybersecurity information. In 2014, the DOJ and FTC released a policy statement explaining that antitrust is not—nor should it be—a "roadblock to legitimate cybersecurity information sharing." While cautioning that their analysis would be fact-driven, the agencies indicated that the sharing of cyber threat information appeared "unlikely in the abstract to increase the ability or incentive of participants to raise price or reduce output, quality, service, or innovation."

Although many activities undertaken by JVs are evaluated under the rule of reason, the JV label does not preclude application of per se rules of illegality. The U.S. Court of Appeals for the First Circuit, for example, has explained that the "talisman of 'joint venture' cannot save an agreement otherwise inherently illegal." Other appellate decisions have distinguished between restraints that are "core" to a JV's procompetitive purposes, restraints that are "ancillary" to such purposes, and restraints that are "nakedly unrelated" to such purposes. While core and ancillary restraints are reviewed under the rule of reason, restraints that are nakedly unrelated to a JV's procompetitive purposes are per se illegal.

Private standard setting represents another form of competitor collaboration that sometimes raises antitrust concerns. In 2004, Congress enacted the Standards Development Organization Advancement Act (SDOAA), which extended the NCRPA's protections to qualifying activities undertaken by certain standards development organizations (SDOs). Those protections, however, are available only to SDOs themselves; they do not apply to participants in an SDO. Participants in qualifying JVs, in contrast, may be eligible for the NCRPA's protections for activities carrying out the purpose of a JV. As discussed below, courts have generally evaluated private standard setting under the rule of reason, though certain agreements reached in connection with standard setting may be subject to more rigorous scrutiny.

AI Safety Collaboration and Antitrust Risk

Different types of safety collaboration among AI developers would raise different levels of antitrust risk under current law. The sharing of information regarding cybersecurity incidents and threats—standing alone—would likely involve low levels of antitrust risk. CISA currently offers an antitrust exemption for this type of information sharing, provided participants do not exchange specified categories of competitively sensitive information. While that exemption is slated to expire later this year absent legislative extension, courts have held that information exchanges are generally analyzed under the rule of reason. An arrangement in which AI developers share cybersecurity information may also constitute a JV entitled to rule-of-reason scrutiny under the NCRPA. While rule-of-reason review does not amount to an antitrust exemption, the sharing of cybersecurity information appears unlikely to raise significant concerns of anticompetitive harm, provided it does not involve the exchange of competitively sensitive information such as prices, costs, or production plans. The antitrust agencies' 2014 policy statement corroborates this point and suggests that enforcers are unlikely to view such conduct with suspicion.

Other forms of collaboration may involve the sharing of methods, metrics, and best practices regarding efforts to align the behavior of AI systems with human intentions. Like the sharing of cybersecurity information, alignment-related information sharing would likely be analyzed using the rule of reason under general antitrust doctrine and, in qualifying circumstances, the NCRPA. The adoption of safeguards to ensure that alignment-related information exchanges do not entail the disclosure of competitively sensitive information would likely bolster arguments for their permissibility under the rule of reason.

Some commentators have suggested that AI developers might engage in joint safety testing of frontier models before they are released. In 2025, OpenAI and Anthropic conducted joint safety and alignment evaluations of each other's publicly released models. Similar collaborations involving prerelease models may pose heightened antitrust risks to the extent that they entail the exchange of competitively sensitive information. As with information sharing, a joint testing regime standing alone (i.e., without a further agreement to withhold products based on test results) would likely be analyzed using the rule of reason under general antitrust doctrine and, potentially, the NCRPA.

The use of third-party evaluators could obviate concerns that safety testing would involve the exchange of competitively sensitive information among rivals. Both OpenAI and Anthropic have expressed support for proposals to embed independent safety evaluators within their companies. Unilateral commitment to such measures would not involve antitrust risk because liability under Section 1 of the Sherman Act requires an agreement between separate economic actors. An agreement among AI firms to use third-party safety

evaluators, however, could raise more nuanced legal issues; antitrust analysis would depend heavily on the details of a particular arrangement.

One possibility that would raise nontrivial antitrust risk would be an agreement not to release AI models unless third-party evaluators—potentially using a common set of standards—certify their safety. While this route could attract legal scrutiny, participating AI companies might defend it as a form of private standard setting. In *Allied Tube & Conduit Corp. v. Indian Head, Inc.*, the Supreme Court indicated that the promulgation of safety standards by private standard-setting organizations can have "significant procompetitive advantages," provided the standards are "based on the merits of objective expert judgments" and promulgated "through procedures that prevent the standard-setting process from being biased by members with economic interests in stifling product competition." Because of these potential procompetitive benefits, the Court explained, most lower courts apply the rule of reason to private standard-setting arrangements, even though such arrangements implicitly constitute agreements "not to manufacture, distribute, or purchase certain types of products." In a footnote, however, the Court added that "[c]oncerted efforts to *enforce* (rather than just agree upon) private product standards face more rigorous antitrust scrutiny."

While *Allied Tube* said that private standard setting is generally reviewed under the rule of reason even when it is tantamount to an agreement not to produce nonconforming products, it is uncertain whether the antitrust agencies share that view. In a July 2025 statement of interest, the DOJ distinguished between standard-setting arrangements that leave participants free to produce nonconforming products (which are evaluated under the rule of reason) and those that do not. Although the DOJ did not specify the appropriate mode of analysis for standard-setting arrangements that forbid the production of nonconforming products, the statement can be read to suggest that quick-look review may apply to such agreements.

An agreement to withhold noncertified AI models may prompt scrutiny regardless of the applicable mode of analysis. In 2025, the FTC investigated an arrangement between four truck manufactures and the California Air Resources Board in which the manufacturers agreed to abide by California emissions standards even if regulations implementing those standards were later invalidated. The FTC closed that investigation in August 2025 after receiving commitments from the truck manufacturers not to enforce the agreement against one another and to act independently and without regard for the agreement's restrictions in selling heavy-duty trucks. The antitrust agencies may be similarly suspicious of agreements between AI developers that restrict permissible product features.

The most dramatic type of AI safety collaboration currently under discussion is a developmental pause, which could take the form of an agreement among AI companies not to develop models above certain capability thresholds. Such an agreement would raise the highest level of antitrust risk of the possible measures reviewed in this Sidebar, as it would amount to a horizontal agreement between leading market participants not to engage in certain forms of quality and R&D competition.

One of the primary arguments for a coordinated pause—that competition will lead to unacceptable safety risks—appears to be foreclosed as an antitrust justification by *National Society of Professional Engineers*. Other defenses may turn on the precise details of a pause agreement. A pause arrangement that is reasonably related to a broader safety-evaluation regime may qualify as an ancillary restraint eligible for rule-of-reason scrutiny. A stand-alone pause arrangement unconnected to a broader venture, in contrast, may be viewed as a naked agreement not to compete on product quality and R&D, potentially resulting in per se condemnation. The lawsuit mentioned in the introduction to this Sidebar makes precisely this argument and alleges that leading AI labs have already entered into such an agreement.

Courts might deem a pause arrangement unlawful under the rule of reason. If a safety justification proved cognizable at step two of that inquiry, step three would allow a plaintiff to establish liability by showing that the procompetitive benefits of a pause agreement could reasonably be achieved through substantially

less restrictive alternatives. A plaintiff might argue that some of the other measures discussed in this Sidebar represent substantially less restrictive means of achieving the safety goals of a developmental pause.

Considerations for Congress

In his September 12 blog post, Amodei implied that antitrust exemptions for AI safety collaboration may be warranted to avoid a "race to the bottom" in which commercial incentives spur AI firms to pursue increasingly risky development of frontier models. Some commentators had discussed the possibility of AI-related antitrust exemptions before Amodei's post, contending that the threat of litigation may deter even unobjectionable forms of AI safety collaboration.

Opponents of AI-related antitrust exemptions have argued that AI firms have considerable scope to engage in safety collaboration under existing law, making such exemptions unnecessary and likely to facilitate anticompetitive collusion. Some have suggested that the leading AI companies may be seeking permission for a developmental pause to lower their massive capital expenditures, boost profits, and justify their valuations before going public, using safety as a pretext. Geopolitical competition with the People's Republic of China (PRC) is also a recurring issue in discussions of AI safety, with some officials worrying that a pause by U.S. AI developers would give PRC firms a competitive edge.

Antitrust exemptions for AI safety collaboration could take a variety of forms. In the 119th Congress, the Collaboration on Adversarial Threats and Security Risks Act (S. 5105 and H.R. 9914) would create antitrust exemptions for certain types of information sharing and agreements to delay or limit the release of AI models, provided certain conditions are met. The legislation would provide that it is not an antitrust violation for two or more nonfederal entities to

- "provide or exchange information or assistance relating to a covered artificial intelligence security risk in good faith for the exclusive purpose of a covered artificial intelligence security purpose"; or
- "coordinate or enter into agreements for the exclusive purpose of reducing covered artificial intelligence security risks via delaying or otherwise limiting the release, deployment, use, development, training, testing, or evaluation of artificial intelligence," provided the entities submit to the head of the DOJ's Antitrust Division a written notice detailing the specific "covered artificial intelligence security risk" and the scope of the proposed restriction.

"Covered artificial intelligence security risks" would include the potential that AI could

- be stolen or weaponized by certain foreign nations;
- substantially facilitate the development or deployment of a chemical, biological, radiological, nuclear, or offensive cyber weapon;
- disrupt, degrade, or cause a loss of operational control over critical infrastructure that is reasonably likely to result in a significant impact on security, national public health, or safety;
- substantially reduce the ability of certain persons to oversee, evaluate, control, or disable the AI, if the applicable persons have authority or responsibility to do so;
- autonomously improve itself in a manner that creates a substantial risk of the above consequences; or

-
- be vulnerable to unauthorized access that creates a substantial risk of the above consequences or unauthorized access that is for the benefit of, at the direction of, or under the control of certain foreign nations or an entity controlled by those nations.

In antitrust litigation, the relevant exemptions would be affirmative defenses that AI developers would have the burden of establishing by a preponderance of the evidence.

Other legislative options are narrower. Congress could, for example, enact an antitrust exemption limited to the sharing of certain information regarding AI safety. This type of measure could take the form of an extension of CISA's antitrust exemption for the sharing of cybersecurity information or the creation of an AI-specific exemption.

Alternative policy options to promote AI safety include ex ante regulation and ex post liability for harmful conduct. In the 119th Congress, H.R. 9925, the Frontier Risk Oversight, National Transparency, Independent Evaluation, and Reporting (FRONTIER) Act, would create a regulatory framework for frontier AI developers that would be administered by the Department of Commerce. Certain existing legal regimes—including tort law—may also provide compensation to persons injured by an AI developer's negligence and encourage developers to implement appropriate safety measures. For catastrophic risks that might render an AI developer insolvent, however, negligence law may be insufficient to induce developers to adopt socially optimal safety precautions.

Author Information

Jay B. Sykes
Legislative Attorney

Disclaimer