

# Draft Report of the Joint California Policy Working Group on AI Frontier Models — Liability and Insurance Comments

Gabriel Weil, Mackenzie Arnold | Institute for Law & AI | April 8, 2025

*Any opinions expressed in these comments are those of the authors and do not reflect the views of the Institute for Law & AI.*

**Comment 1: The draft report correctly points to insurance as a potentially useful policy lever. But it incorrectly suggests that insurance alone (without liability) will cause companies to internalize their costs. Insurance likely will not work without liability, and the report should acknowledge this.**

Insurance could advance several goals at the center of this report. Insurance creates private market incentives to more accurately measure and predict risk, as well as to identify and adopt effective safety measures. It can also bolster AI companies' ability to compensate victims for large harms caused by their systems. The value of insurance is potentially limited by the difficulty of modeling at least some risks in this context, but to the extent that the report's authors are enthusiastic about insurance, it is worth highlighting that these benefits depend on the underlying prospect of liability. If AI companies are not—and do not expect to be—held liable when their systems harm their customers or third parties, they would have no reason to purchase insurance to cover those harms and inadequate incentives to mitigate those risks.

Passing state laws that require insurance doesn't solve this problem either: if companies aren't held liable for harms they generate (because of gaps in existing law, newly legislated safe harbors, federal preemption, or simple underenforcement), insurance plans would cease to accurately track risk.

In section 1.3, the draft report suggests efforts to:

reconstitute market incentives for companies to internalize societal externalities (e.g., incentivizing insurance may mold market forces to better prioritize public safety)."

We propose amending this language to read:

reconstitute market incentives for companies to internalize societal externalities (e.g., clear liability rules, especially for harms to non-users, combined with incentives to acquire liability insurance may mold market forces to better prioritize public safety).

**Comment 2: Liability can be a cost-effective tool for mitigating risk without discouraging innovation, especially under conditions of uncertainty. And many of the report’s transparency suggestions would improve the efficiency of liability and private contracting. The report should highlight this.**

Overall, the report provides minimal discussion of liability as a governance tool. To the extent it does, the tone (perhaps) suggests skepticism of liability-based governance (“In reality, when governance mechanisms are unclear or underdeveloped, oversight often defaults largely to the courts, which apply existing legal frameworks—such as tort law...”).

But liability is a promising tool, even more so given the considerable uncertainty surrounding future AI risks—a point that the authors correctly emphasize is *the* core challenge of AI policy.

Liability has several key advantages under conditions of uncertainty. Liability is:

- **Proportional:** Automatically adjusting incentives based on the severity and breadth of harm that actually occurs or that AI companies expect to occur
- **Flexible:** Creating general incentives to mitigate risks and act “reasonably” rather than prescribing the exact steps companies must take—this has several advantages, among them:
  - Allowing those closest to the ground (the companies themselves) to identify effective mitigation measures and decide whether the cost is merited
  - Requiring less political consensus than prescriptive regulations
  - Enabling incentives that are conditional on disputed risks actually materializing
  - Avoiding the need to prescribe specific mitigations before best practices emerge.
- **Fact-bound:** Deferring key decisions until after harm has occurred, which means that key decisions are made based on a fuller factual record, when we know more, with greater confidence

Ex ante regulations require companies to pay their costs upfront. Where those costs are large, they depend on a strong social consensus about the magnitude of the risks that they are designed to mitigate. Prescriptive rules and approval regulation regimes, the most common forms of ex ante regulation, also depend on policymakers’ ability to identify specific precautionary measures early on, which is challenging in a nascent field like AI, where best practices are still being developed and considerable uncertainty exists about the severity and nature of potential risks.

Liability, by contrast, scales automatically with the risk and shifts decision-making regarding what mitigation measures to implement to the AI companies, who are often best positioned to identify cost-effective risk mitigation strategies.

Concerns about excessive litigation are reasonable but can be mitigated by allowing wide latitude for contracts to waive and allocate liability between model developers, users, and various intermediaries—with the notable exception of third-party harm, where the absence of contractual privity does not allow for efficient contracting. In fact, allocation of responsibility by contract goes hand-in-hand with the transparency and

information-sharing recommendations highlighted in the report—full information allows for efficient contracting. Risk of excessive litigation also varies by context, being least worrisome where the trigger for liability is clear and rare (as is the case with liability for extreme risks) and most worrisome where the trigger is more common and occurs in a context where injuries are common even when the standard of care is followed (e.g., in the context of healthcare). There may be a case for limiting liability in contexts where false positives are likely to abound, but liability is a promising, innovation-compatible tool in some of the contexts at the center of this report..

A strong summary of the potential use and limitations of liability for AI risk would note that:

- Liability is a promising policy tool, especially in light of uncertainty about the nature and severity of AI risks
- Liability has the advantages of being proportional, flexible, and fact-bound
- The case for liability is particularly strong for extreme risks (where uncertainty and under-developed best practices make prescriptive approaches more difficult, and liability allows for adequate flexibility)
- The case for liability is particularly strong for third-party harms (where efficient contracting is not possible)
- Transparency requirements and third-party evaluations enhance the ability of private parties to contract efficiently and allocate liability costs via indemnification
- Liability is not uniformly bad for innovation. It can be innovation-compatible or promoting in certain contexts.

**Comment 3: Creating safe harbors that protect AI companies from liability is a risky strategy, given the uncertainty about both the magnitude of risks posed by AI and the effectiveness of various risk mitigation strategies. The report should note this.**

In recent months, several commentators have called for preemption of state tort law or the creation of safe harbors in return for compliance with some of the suggestions made in this report. While we believe that the policy tools outlined in the report are important, it would be a valuable clarification for the report to state that these requirements alone do not merit the removal of background tort law protections.

Under existing negligence law, companies can, of course, argue that their compliance with many of the best practices outlined in this report is *evidence* of reasonable care. But, as outlined above, tort law creates additional and necessary incentives that cannot be provided through reporting and evaluation alone.

As we see it, tort law is compatible with—not at odds with or replaceable by—the evidence-generating, information-rich suggestions of this report. In an ecosystem with greater transparency and better evaluations, parties will be able to even more efficiently distribute liability via contract, enhancing its benefits and more precisely distributing its costs to those best positioned to address them.

It also merits noting that creating safe harbors based on compliance with relatively light-touch measures like transparency and third-party verification would be an unusual step historically, and would greatly reduce AI companies' incentives to take risk-mitigation measures that are not expressly required.

Because tort law is enhanced by the suggested policies of this report and addresses the key dilemma (uncertainty) that this report seeks to address, we recommend that the report clarify the risk posed by broad, general liability safe harbors.

**Comment 4: The lesson of climate governance is that transparency alone is inadequate to produce good outcomes. When confronting social externalities, policies that directly compel the responsible parties to internalize the costs and risks that they generate are often the most efficient solutions. In the climate context, the best way to do this is with an ex ante carbon price. Given the structural features of AI risk, ex post liability plays an analogous role in AI governance.**

Section 2.4 references lessons from climate change governance. “The case of fossil fuel companies offers key lessons: Third-party risk assessment could have realigned incentives to reward energy companies innovating responsibly while simultaneously protecting consumers.” In our view, this overstates the potential of transparency measures like third-party risk assessment alone and undervalues policies that compel fossil fuel companies and their consumers to internalize the costs generated by fossil fuel combustion. After all, the science on climate change has been reasonably clear for decades now and that alone has been far from sufficient to align the incentives of fossil fuel companies with social welfare. The core policy challenge of climate change is that fossil fuel combustion generates global negative externalities in the form of heat-trapping effects of greenhouse gas emissions. Absent policies, like carbon pricing, to compel fossil fuel companies and their consumers to internalize the costs generated by fossil fuel combustion, mere transparency about climate impacts is an inadequate response.

Third-party risk assessments and other transparency measures alone are similarly unlikely to be sufficient in the AI risk context. Transparency and third-party evaluation are best thought of as tools that help prepare us for further action (be it through generating better quality evidence on which to regulate, enabling more efficient contracting to allocate risk, or enabling efficient litigation once harms occur). But without that further action, they forego much of their potential value. Aligning the incentives of AI companies will require holding them financially accountable for the risks that they generate, and Liability is the best accountability tool we have for AI risk and plays a structurally similar role to carbon pricing for climate risk mitigation.

We propose amending the report language to read, “The case of fossil fuel companies offers key lessons: Third-party risk assessment could have helped build the case for policies, like carbon pricing, that would have realigned incentives to reward energy companies innovating responsibly while simultaneously protecting consumers.”

Section 2.4 further states, “The costs of action to reduce greenhouse gas emissions, meanwhile, were estimated [by the Stern Review] at only 1% of global GDP each year. This is a useful lesson for AI policy: Leveraging evidence-based projections, even under uncertainty, can reduce long-term economic and security costs.”

But this example only further evidences the fact that cost internalization mechanisms, in addition to transparency mechanisms, are key to risk reduction. The Stern Review's cost estimates were based on the assumption that governments would implement the most cost-effective policies, like economy-wide carbon pricing, to reduce greenhouse gas emissions. Actual climate policies implemented around the world have tended to be substantially less cost-effective. This is not because carbon pricing is more costly or less effective than Stern assumes but because policymakers have been reluctant to implement it aggressively, despite broad global acceptance of the basic science of climate change.

This lesson is highly relevant to AI governance inasmuch as the closest analog to carbon pricing is liability, which directly compels AI companies to internalize the risks generated by their systems, just as a carbon price compels fossil fuel companies to internalize the costs associated with their incremental contribution to climate change. An AI risk tax is impractical since it is not feasible to measure AI risk ex ante. But, unlike with climate change, it will likely generally be feasible to attribute AI harms to particular AI systems and to hold the companies that trained and deployed them accountable.

## Supporting documents

For more on the analogy between AI liability and carbon pricing and an elaboration of a proposed liability framework that accounts for uninsurable risks, see Gabriel Weil, Tort Law as a Tool for Mitigating Catastrophic Risk from Artificial Intelligence, [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4694006](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4694006).

This proposal is also summarized in this magazine article: Gabriel Weil, [Your AI Breaks It? You Buy It: AI developers should pay for what they screw up](#), Noema Mag (2024)

For more on the case for prioritizing liability as an AI governance tool, see Gabriel Weil, Instrument Choice in AI Governance: Liability and its Alternatives, Google Docs, <https://docs.google.com/document/d/1ivtgFLDQqG05U2vM1211wNtTDxNCjZr1-2NWf6tT5cU/edit?tab=t.0>.

The core arguments are also laid out in this Lawfare piece: Gabriel Weil, [Tort Law Should Be the Centerpiece of AI Governance](#), Lawfare (2024).