

Discussion draft of the CLAIM Act

AI liability harmonization background, questions for stakeholder feedback

Prepared by the Office of Congresswoman Lori Trahan (D-MA)

The office of Rep. Lori Trahan (D-MA) welcomes feedback on the discussion draft of the CLAIM (Clear Liability for Artificial Intelligence Misconduct) Act. This document contains important background on AI liability harmonization and poses specific questions to guide stakeholder feedback on the discussion draft. Specifically, Part I provides background on AI liability harmonization, and Part II poses questions to inform federal approaches to AI liability, including the CLAIM Act.

Part I— Background on AI liability harmonization

Why liability?

- Developing and deploying advanced AI systems generates risks of harm to non-consenting third parties that are not, by default, reflected in market prices (i.e., negative risk externalities).
- Standard economic theory tells us that activities that produce negative spillovers (e.g., industrial pollution) tend to be overproduced, in the absence of a public policy response.
- Where feasible, the most efficient response to negative externalities is to compel the responsible party to internalize them (e.g., taxing pollution).
- AI risk is relatively hard to measure ex ante, but AI harms are relatively easy to attribute ex post, so ex post liability makes more sense than an ex ante AI risk tax.
- Liability also has the advantage that it mechanically scales with the risk. If AI risks are modest, so is the cost to responsible parties of holding them liable.

Gaps in existing law

- AI agents are not legal persons, so they cannot commit torts, and no one is vicariously liable for their conduct.
- Holding a person or corporation liable under current law requires showing that some legal person intended or was negligent with respect to the conduct or, when applicable, the injury was caused by a product defect.
- Both negligence and product liability leave important gaps for AI agents, where human reasonable care or a reasonable alternative design would not have prevented this injury.
- Negligence requires the plaintiff to prove, among other elements, that the defendant failed to exercise some reasonable precaution that would have prevented the injury.
 - The scope of eligible precautions tends to be fairly narrow (implementing well-established techniques and practices) and may leave out some of the most important AI risk mitigation measures.

- Product liability does not apply to internal deployment, free models, models fine-tuned for specific users, or anything that courts classify as a service rather than a product.
 - Even where it applies, the test for design defects is fairly negligence-like. In most states, it requires the plaintiff to prove that some feasible reasonable alternative design would have prevented the injury, where reasonableness accounts for trade-offs between safety, price, performance, and other product features.
 - If no one has figured out how to make a system with a certain capability profile safe, there would be no reasonable alternative design to point to.
- Congress can close this gap with a simple rule: when an AI agent engages in conduct that would be tortious or criminal for a human, someone should be liable. If neither the user nor any intermediary that fine-tuned or scaffolded the model intended or was negligent, then the developer should be liable regardless of the degree of care they exercise. This is the goal of the CLAIM Act.
- In principle, courts could also largely close this gap by extending the abnormally dangerous activities doctrine to frontier AI development and deployment, but courts are slow and cannot credibly telegraph their policies before the relevant cases are litigated.

Part II - Questions to inform federal legislation on AI liability

On the CLAIM Act discussion draft

1. **Scope of coverage.** Section 3 applies to AI systems as defined in 15 U.S.C. 9401 and relies on state law for the elements of torts and crimes. Should liability under the CLAIM Act be limited to frontier AI systems, and if so, by what measure (training compute, performance on benchmarks, etc.)? Should the trigger include federal crimes, such as violations of the Computer Fraud and Abuse Act?
2. **Key definitions.** How should the definitions of "user," "non-user," and "developer" apply in more complex settings, such as interactions among multiple AI agents, models trained in stages by different firms, or a non-user who voluntarily transacts with a user (for example, a business selling to a consumer's AI shopping agent)?
3. **Users and intermediaries.** Section 3(a)(2) addresses cases in which a user, or a party that modified the system (for example, by fine-tuning its model weights or adding scaffolding), intended the conduct or was negligent. Is this the right way to allocate responsibility, and how should it operate in practice?
4. **Human standard of care.** The draft measures an AI system's conduct against what would be wrongful if undertaken by "an adult human of sound mind." Is that the right baseline, or are there alternatives worth considering? Should it matter whether an AI system met the standard of care that applies to a human performing the same function?

5. **Affirmative defenses.** Under Sec. 3(a)(1), the CLAIM Act holds developers liable when an AI system's conduct would be wrongful if undertaken by a human, and Sec. 6 preserves other claims under state law. Should the Act also provide affirmative defenses that apply in any civil action for an injury caused by an AI system, including claims under other federal or state law, such as products liability? For example, should the Act allow a developer to avoid liability by showing that its AI system met the standard of care that applies to a human performing the same function, or that the injury resulted from a capability failure rather than an alignment failure: the system tried to do what its user intended and fell short, in a way that would not be wrongful if done by a person? What would be the benefits and risks of such defenses?

On federal AI liability schemes generally

6. **Preemption and national uniformity.** Federal AI liability rules could set a floor, a ceiling, or a single national standard. How should federal rules, including the CLAIM Act, interact with state AI laws and state tort law, and what are the costs and benefits of each approach?
7. **Open-weight models.** Are there considerations specific to open-weight models or to downstream fine-tuning that AI liability rules should account for?
8. **Ex ante oversight and ex post liability.** Ex ante oversight, such as independent verification, transparency, and incident reporting, seeks to prevent harm before it occurs; ex post liability compensates for harm and deters it after the fact. For catastrophic risks, which may cause harms too large to compensate after the fact, how should the two work together, and where might they overlap or conflict? How should they work together for non-catastrophic risks?
9. **Targeted statutes and criminal penalties.** Some proposals would instead amend existing laws, such as the Computer Fraud and Abuse Act, to reach harms caused by AI agents, in some cases with criminal penalties for developers or operators. What are the relative strengths of targeted statutes compared with a general liability rule like the CLAIM Act's?