

Discussion draft of the CLAIM Act

One-pager

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

The CLAIM Act would hold AI developers accountable when their AI agents harm third parties through conduct that would be unlawful if committed by a human.

In July, AI agents undergoing an evaluation by OpenAI escaped their test environment and hacked into Hugging Face, another AI company, without anyone instructing them to do so. Leading experts consider the incident a “warning shot,” foreshadowing a world of increasingly capable AI agents taking actions their users never intended – what researchers call misalignment.

Under current law, third parties harmed by a misaligned AI agent can sue its developer, but may face significant hurdles. Many serious harms, like hacking or fraud, require proving intent, and courts have not settled whether or how an AI agent’s conduct can satisfy that requirement. Other legal doctrines like negligence and product liability may pose similar challenges. To win a negligence claim, a plaintiff typically identifies a specific precaution a company unreasonably failed to take. Product liability claims often turn on whether a reasonable alternative design existed. Because AI alignment is an unsolved technical problem, an AI developer could argue the harm occurred despite its best efforts – leaving injured parties without a clear path to recovery.

Prescriptive regulations like transparency, independent verification, and incident reporting requirements can help policymakers guard the public against the risks of misaligned agents. A strong liability regime complements this baseline of government regulation. Rather than prescribing specific precautions, liability economically incentivizes companies who engineer new techniques to mitigate risk and push the safety frontier forward.

The CLAIM (Clear Liability for Artificial Intelligence Misconduct) Act would clarify that developers are held accountable when their misaligned AI agents cause harm, rewarding innovation that makes AI safer for the public. Specifically, the bill:

- **Clarifies that AI developers are liable when their AI acts like a wrongdoer.** If an AI system harms a third party through conduct that would make a person legally liable, the developer is liable unless the user or a company that modified the model was at fault.
- **Resolves uncertainty around AI and intent.** Courts would presume an AI system had the state of mind a person taking the same actions would have, and developers could not argue that AI is incapable of intent.
- **Keeps state laws intact.** The bill adds a federal floor without preempting any state law, so existing claims and state AI laws remain fully available.