SENATE JUDICIARY COMMITTEE
Senator Thomas Umberg, Chair
2025-2026 Regular Session

AB 316 (Krell)
Version: April 28, 2025
Hearing Date: June 24, 2025
Fiscal: No
Urgency: No
CK

SUBJECT

Artificial intelligence: defenses

DIGEST

This bill prohibits a defendant from asserting the defense that AI autonomously caused harm to a plaintiff, as provided.

EXECUTIVE SUMMARY

As artificial intelligence models and applications become more sophisticated and integrated into our daily lives, they introduce new safety and security risks. Automated systems can make critical errors in high-stakes situations like self-driving vehicles, medical diagnostics, or home security systems when they encounter edge cases or adversarial inputs. AI-powered chatbots, phishing, identity theft, and deepfakes create novel threats to personal security and assets. Additionally, over-reliance on AI systems without adequate human oversight in critical infrastructure or emergency response could lead to cascading failures during unusual circumstances. While these technologies offer tremendous benefits, ensuring the highest level of due care on the part of AI developers and deployers is of paramount importance. Generally, individuals and entities are not only liable for their willful acts but also for injuries caused by their lack of ordinary care in managing their property or person. However, there are concerns that existing legal frameworks may be challenged in addressing the unique risks and complexities of AI technologies.

This bill addresses the issue by making clear that a defendant who developed, modified, or used AI that is alleged to have caused a harm to a plaintiff, cannot assert as a defense that the AI autonomously caused the harm to the plaintiff. This bill is sponsored by the Children's Advocacy Institute and the Organization for Social Media Safety. It is supported by a number of organizations, including the California Federation of Labor Unions and the California Initiative for Technology and Democracy (CITED). The bill is opposed by Technet and the California Chamber of Commerce.

PROPOSED CHANGES TO THE LAW

Existing law:

1) Provides that every person is responsible, not only for the result of their willful acts, but also for an injury occasioned to another by the person's want of ordinary care or skill in the management of their property or person, except so far as the latter has, willfully or by want of ordinary care, brought the injury upon themselves. (Civ. Code § 1714(a).)

2) Defines "artificial intelligence" as an engineered or machine-based system that varies in its level of autonomy and that can, for explicit or implicit objectives, infer from the input it receives how to generate outputs that can influence physical or virtual environments. (Gov. Code § 11546.45.5.)

This bill provides that in an action against a defendant who developed, modified, or used AI that is alleged to have caused a harm to the plaintiff, it shall not be a defense, and the defendant may not assert, that the artificial intelligence autonomously caused the harm to the plaintiff.

COMMENTS

1. The risks presented by AI models and applications

With recent dramatic advances in the capabilities of AI systems, the need for frameworks for accountability and responsible development have become ever more urgent.

In January of 2017, AI researchers, economists, legal scholars, ethicists, and philosophers met in Asilomar, California, to discuss principles for managing the responsible development of AI. The collaboration resulted in the Asilomar Principles. Aspirational rather than prescriptive, these 23 principles were intended to initiate and frame a dialogue by providing direction and guidance for policymakers, researchers, and developers. The Legislature subsequently adopted ACR 215 (Kiley, Ch. 206, Stats. 2018), which added the State of California to that list by endorsing the Asilomar Principles as guiding values for the development of artificial intelligence and related public policy. One key admonition from these principles is to "recognize that [AI's] risks are potentially catastrophic or existential."

As directed by the National AI Initiative Act of 2020, the National Institute of Standards and Technology (NIST) developed the AI Risk Management Framework to assist entities designing, developing, deploying, and using AI systems to help manage the many risks of AI and promote trustworthy and responsible development and use of AI

systems. That framework highlights the serious risks at play and the uniquely challenging nature of addressing them in this context:

> Artificial intelligence (AI) technologies have significant potential to transform society and people's lives – from commerce and health to transportation and cybersecurity to the environment and our planet. AI technologies can drive inclusive economic growth and support scientific advancements that improve the conditions of our world. AI technologies, however, also pose risks that can negatively impact individuals, groups, organizations, communities, society, the environment, and the planet. Like risks for other types of technology, AI risks can emerge in a variety of ways and can be characterized as long- or short-term, high or low-probability, systemic or localized, and high- or low-impact.
>
> While there are myriad standards and best practices to help organizations mitigate the risks of traditional software or information-based systems, the risks posed by AI systems are in many ways unique. AI systems, for example, may be trained on data that can change over time, sometimes significantly and unexpectedly, affecting system functionality and trustworthiness in ways that are hard to understand. AI systems and the contexts in which they are deployed are frequently complex, making it difficult to detect and respond to failures when they occur. AI systems are inherently socio-technical in nature, meaning they are influenced by societal dynamics and human behavior. AI risks – and benefits – can emerge from the interplay of technical aspects combined with societal factors related to how a system is used, its interactions with other AI systems, who operates it, and the social context in which it is deployed.
>
> These risks make AI a uniquely challenging technology to deploy and utilize both for organizations and within society. [. . .]
>
> AI risk management is a key component of responsible development and use of AI systems. Responsible AI practices can help align the decisions about AI system design, development, and uses with intended aim and values. Core concepts in responsible AI emphasize human centricity, social responsibility, and sustainability. AI risk management can drive responsible uses and practices by prompting organizations and their internal teams who design, develop, and deploy AI to think more critically about context and potential or unexpected negative and positive impacts. Understanding and managing the risks of AI systems will help to enhance trustworthiness, and in turn, cultivate public trust.

This highlights how the risks posed by AI are inherently complex and ever-changing. Constant adaptions and nimble responses to addressing potential risks is of critical importance.

More recently the Biden Administration published its Blueprint for an AI Bill of Rights, which is a set of five principles and associated practices to help guide the design, use, and deployment of AI to protect the rights of the American public. One key piece focuses on the safety of these systems: "*Safe and Effective Systems*: You should be protected from unsafe or ineffective systems. Automated systems should be developed with consultation from diverse communities, stakeholders, and domain experts to identify concerns, risks, and potential impacts of the system."[1]

While the future is unclear, the need to respond to these potential harms now is evident. The Center for New American Security puts a fine point on it:

> While there is significant uncertainty in how the future of AI develops, current trends point to a future of vastly more powerful AI systems than today's state of the art. The most advanced systems at AI's frontier will be limited initially to a small number of actors but may rapidly proliferate. Policymakers should begin to put in place today a regulatory framework to prepare for this future. Building an anticipatory regulatory framework is essential because of the disconnect in speeds between AI progress and the policymaking process, the difficulty in predicting the capabilities of new AI systems for specific tasks, and the speed with which AI models proliferate today, absent regulation. Waiting to regulate frontier AI systems until concrete harms materialize will almost certainly result in regulation being too late.[2]

2. Civil liability and immunity

As a general rule, California law provides that persons are responsible, not only for the result of their willful acts, but also for an injury occasioned to another by their want of ordinary care or skill in the management of their property or person, except so far as the latter has, willfully or by want of ordinary care, brought the injury upon themselves. (Civ. Code § 1714(a).) Liability has the primary effect of ensuring that some measure of recourse exists for those persons injured by the negligent or willful acts of others; the risk of that liability has the primary effect of ensuring parties act reasonably to avoid harm to those to whom they owe a duty.

---

[1] *Blueprint For An AI Bill Of Rights* (October 2022) Office of Science and Technology Policy, https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf [as of Jan. 22, 2025].

[2] Paul Scharre, *Future-Proofing Frontier AI Regulation* (March 2024) Center for New American Security, https://s3.us-east-1.amazonaws.com/files.cnas.org/documents/CNAS-Report_AI-Trends_FinalC.pdf. This, and all further, internet citations are current as of June 14, 2025.

Conversely, immunity from liability disincentivizes careful planning and acting on the part of individuals and entities. When one enjoys immunity from civil liability, they are relieved of the responsibility to act with due regard and an appropriate level of care in the conduct of their activities. Immunity provisions are also disfavored because they, by their nature, preclude parties from recovering when they are injured, and force injured parties to absorb losses for which they are not responsible. Liability acts not only to allow a victim to be made whole, but to encourage appropriate compliance with legal requirements.

3. Definitively prohibiting the autonomous AI defense

Negligence law serves a crucial purpose in our legal system by incentivizing individuals and companies to take reasonable precautions to prevent harm. When organizations face potential liability for negligent design, testing, or deployment, they are motivated to invest in robust safety measures, thorough testing protocols, and ongoing risk monitoring. When a company is held responsible for harms it causes through its systems or products, it prompts a proactive approach to avoid causing those harms. This creates a direct financial incentive to prioritize safety, especially when utilizing new technologies. This is the state of the law currently.

This bill makes clear that it shall not be a defense, and a defendant shall not assert, that AI developed, modified, or used by the defendant autonomously caused alleged harm to a plaintiff. This ensures that AI development and deployment is done with due care despite the novel nature of the technology and its inherent complexities. Ultimately, this preserves the principle that humans are responsible for the harms they cause, regardless of the sophistication or autonomy of the tools they use.

While there are no examples of defendants successfully utilizing such defenses, this bill proactively rules out this avenue of deflecting blame when someone suffers AI-related injuries. One example of where this defense has been attempted was recently widely reported on:

> In 2022, Air Canada's chatbot promised a discount that wasn't available to passenger Jake Moffatt, who was assured that he could book a full-fare flight for his grandmother's funeral and then apply for a bereavement fare after the fact.
>
> According to a civil-resolutions tribunal decision last Wednesday, when Moffatt applied for the discount, the airline said the chatbot had been wrong – the request needed to be submitted before the flight – and it wouldn't offer the discount. Instead, the airline said the chatbot was a "separate legal entity that is responsible for its own actions". Air Canada argued that Moffatt should have gone to the link provided by the chatbot, where he would have seen the correct policy.

> The British Columbia Civil Resolution Tribunal rejected that argument, ruling that Air Canada had to pay Moffatt $812.02 (£642.64) in damages and tribunal fees. "It should be obvious to Air Canada that it is responsible for all the information on its website," read tribunal member Christopher Rivers' written response. "It makes no difference whether the information comes from a static page or a chatbot."[3]

A recent article highlights the importance of addressing this issue and heading off any legal doctrine that allows AI itself to be blamed:

> On the question of any eventual push to juridical personhood for AIs, it does seem that a cautionary note is needed. Certainly, it can be strongly argued in terms both of equity and deterrence, that there is, in the present state of things, a need to avoid ideas or developments that may see those persons and corporate entities that have profited from the development of AIs ultimately avoid liability for the consequences of their creation, particularly where those consequences are harmful.[4]

According to the author:

> The California AI industry is rapidly growing, both from an economic and technological standpoint. AI has seen extraordinary advancements in its applications, complexity, and autonomy, to the point where AI is replacing human intelligence in certain tasks. As AI becomes more complex, it is increasingly involved in daily interactions and significant decision-making. While this has the potential to bring positive changes to various industries and facets of life, this also means that AI related harm can be much more significant. These harms are already manifesting and will only worsen as the AI race becomes more competitive. Specifically, AI being deployed through social media has been shown to be particularly harmful to youth.
>
> This bill ensures that companies benefiting from the use of AI are also responsible for the harms AI may cause. By eliminating a potential AI defense theory, this bill encourages careful vetting of AI products before they are used and ensures that there is a legal entity held to account if AI is shown to violate the law.

---

[3] Maria Yagoda, *Airline held liable for its chatbot giving passenger bad advice - what this means for travelers* (February 23, 2024) BBC, https://www.bbc.com/travel/article/20240222-air-canada-chatbot-misinformation-what-travellers-should-know.

[4] Michael Duffy, *Rise of the 'Machine Defendant'? A Cautionary Analysis and Conceptualisation of Civil and Criminal Liability Approaches to the Actions of Robots and Artificial Intelligence* (January 1, 2023). 49(2) Monash University Law Review 1-42, http://dx.doi.org/10.2139/ssrn.5032505.

4. Stakeholder positions

The Organization for Social Media Safety, a sponsor of the bill, writes:

> Given both the alarming speed at which AI-based tools are being deployed and the clear, convincing proof that these tools can cause severe harm, especially to children, we must ensure that our standard liability framework functions as expected to protect consumers. This established jurisprudence has been instrumental in ensuring that California's marketplace has an outstanding safety record, preventing deaths and injuries for millions of consumers while reliably fostering innovation.
>
> We cannot afford to wait decades for litigation to unfold while Big Social advances novel legal theories arguing that autonomously operating AI, rather than the companies themselves, should bear responsibility for the harm caused. At a minimum, this ambiguity must be clarified now.

Technet and the California Chamber of Commerce write in joint opposition:

> We understand the intent to prevent defendants from attempting to absolve themselves from liability by claiming that an artificial intelligence acted autonomously. Our concern is that the bill could be interpreted to prevent a defendant from presenting any evidence related to an AI or automated system, which may be relevant to causation, foreseeability of harm, and the comparative fault of other parties.

It should be noted that this bill simply prevents defendants from asserting that AI caused the harm on its own. Plaintiffs are still required to establish all the elements of their causes of action, and defendants are still permitted to present relevant evidence to rebut that evidence.

Writing in support, CITED argues:

> Artificial intelligence offers unprecedented opportunities for advances in untold fields. But this technological marvel can also be used to cause harm, both large and small. Whether it is defamatory deepfakes designed to fraudulently impact voting, chatbots that mimic human conversations and encourage children to harm themselves, price collusion among competitors designed to maximize profits at the expense of consumers, or the creation of child sexual abuse material, AI can be developed or deployed in ways that cause grave harm.
>
> Those who reap the benefits of AI must also accept the risks and responsibilities. AB 316 (Krell) takes a small step in that direction by

making clear that those who cause harm through the use of AI that they have created or deployed should not be able to avoid legal responsibility by claiming that the AI itself -- autonomously -- caused the harm.

As drafted, this bill does not in any way limit the plaintiff's burden of proof or make it easier for them to prove their case. While we believe that California ought to lead the nation in establishing an AI liability regime so Californians who are injured by AI can appropriately seek redress in the courts, as they can with all other products and all other industries, this bill does not do that. It simply says that no one should be able to shirk responsibility for the harm they cause through a consumer's use of their AI product with a specious legal argument about that product having its own autonomy.

SUPPORT

Children's Advocacy Institute (sponsor)
Organization for Social Media Safety (sponsor)
3strands Global Foundation
California Conference Board of the Amalgamated Transit Union
California Conference of Machinists
California Federation of Labor Unions, AFL-CIO
California Initiative for Technology & Democracy, a Project of California Common CAUSE
California Nurses Association
California Teamsters Public Affairs Council
Consumer Attorneys of California
Consumer Federation of California
Economic Security California Action
Engineers and Scientists of California, IFPTE Local 20, AFL-CIO
Oakland Privacy
TechEquity Action
The Center for AI and Digital Policy
UFCW - Western States Council
UNITE Here International Union, AFL-CIO
Utility Workers Union of America

OPPOSITION

California Chamber of Commerce
Technet

RELATED LEGISLATION

Pending Legislation:

SB 243 (Padilla, 2025) requires operators of "companion chatbot platforms" that allow users to engage with chatbots to take reasonable steps to prevent their chatbots from engaging in specified conduct, including offering unpredictable rewards and encouraging increased engagement. Operators must periodically remind users that the chatbot is not human and implement protocols for addressing suicidal ideation expressed by users, as well as conduct annual audits. SB 243 is currently in the Assembly Privacy and Consumer Protection Committee.

SB 420 (Padilla, 2025) regulates the use of high-risk automated decision systems (ADS). This includes requirements on developers and deployers to perform impact assessments on their systems. The bill establishes the right of individuals to know when an ADS has been used, details about the systems, and an opportunity to appeal ADS decisions, where technically feasible. SB 420 is currently in the Assembly Privacy and Consumer Protection Committee.

SB 468 (Becker, 2025) imposes a duty on a business that deploys a high-risk artificial intelligence system, or high-risk ADS, that processes personal information to protect that information and requires such a deployer to maintain a comprehensive information security program that meets specified requirements. SB 468 is currently in the Senate Appropriations Committee.

SB 813 (McNerney, 2025) provides a rebuttable presumption against liability for harms caused by an AI model or application if it is certified by a private multistakeholder regulatory organization that is designated by the Attorney General, as provided. SB 813 is currently in the Senate Appropriations Committee.

AB 1018 (Bauer-Kahan, 2025) requires a developer of a covered ADS to take certain actions, including conduct performance evaluations of the ADS, submit to third-party audits, and provide deployers to whom the developer transfers the covered ADS with certain information, including the results of those performance evaluations. It requires a deployer of a covered ADS to take certain actions, including provide certain disclosures to a subject of a consequential decision made or facilitated by the covered ADS, provide the subject an opportunity to opt out of the use of the covered ADS, provide the subject with an opportunity to correct erroneous personal information used by the ADS, and to appeal the outcome of the consequential decision, and submit the covered ADS to third-party audits, as prescribed. AB 1018 is currently in this Committee.

Prior Legislation:

SB 1047 (Wiener, 2024) would have, among other things, required developers of powerful AI models and those providing the computing power to train such models to put appropriate safeguards and policies into place to prevent critical harms. It would have established a state entity to oversee the development of these models. SB 1047 was vetoed by Governor Newsom. In his veto message, he stated:

> SB 1047 magnified the conversation about threats that could emerge from the deployment of AI. Key to the debate is whether the threshold for regulation should be based on the cost and number of computations needed to develop an AI model, or whether we should evaluate the system's actual risks regardless of these factors. This global discussion is occurring as the capabilities of AI continue to scale at an impressive pace. At the same time, the strategies and solutions for addressing the risk of catastrophic harm are rapidly evolving.
>
> By focusing only on the most expensive and large-scale models, SB 1047 establishes a regulatory framework that could give the public a false sense of security about controlling this fast-moving technology. Smaller, specialized models may emerge as equally or even more dangerous than the models targeted by SB 1047 - at the potential expense of curtailing the very innovation that fuels advancement in favor of the public good.

AB 2885 (Bauer-Kahan & Umberg, Ch. 843, Stats. 2024) established a uniform definition for "artificial intelligence" in California's code, which is used in this bill.

AB 2930 (Bauer-Kahan, 2024) would have regulated the use of ADS in order to prevent "algorithmic discrimination." This includes requirements on developers and deployers that make and use these tools to make "consequential decisions" to perform impact assessments on ADS. It would have established the right of individuals to know when an ADS is being used, the right to opt out of its use, and an explanation of how it is used. AB 2930 died without a vote on the Senate Floor.

ACR 215 (Kiley, Ch. 206, Stats. 2018) *See* Comment 1.

PRIOR VOTES:

Assembly Floor (Ayes 70, Noes 1)
Assembly Privacy and Consumer Protection Committee (Ayes 11, Noes 2)
Assembly Judiciary Committee (Ayes 12, Noes 0)
****