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

SB 813 (McNerney)
Version: March 26, 2025
Hearing Date: April 29, 2025
Fiscal: Yes
Urgency: No
CK

SUBJECT

Multistakeholder regulatory organizations

DIGEST

This bill provides civil immunity for harms caused by an AI model or application if it is certified by a private "multistakeholder regulatory organization" (MRO) that is designated by the Attorney General (AG), 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.

This bill creates an immunity shield against personal injury and property damage caused by AI models and applications that are certified at the time of the injuries by a private entity designated by the AG, called an MRO. MRO applicants are to put forward plans for certifying AI developers and security vendors and the AG assesses their adequacy, the quality of the MRO's personnel, and its independence from the AI industry before designating them with these certification powers.

This bill is sponsored by Fathom, an entity with hopes of becoming an MRO. It is supported by 21 individuals. It is opposed by industry and advocacy groups, including the California Chamber of Commerce and the Consumer Attorneys of California.

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) Requires the California Department of Technology (CDT) to conduct a comprehensive inventory of all high-risk automated decision systems (ADS) that have been proposed for use, development, or procurement by, or are being used, developed, or procured by, any state agency. It defines the relevant terms:
    a) "Automated decision system" means a computational process derived from machine learning, statistical modeling, data analytics, or artificial intelligence that issues simplified output, including a score, classification, or recommendation, that is used to assist or replace human discretionary decisionmaking and materially impacts natural persons. "Automated decision system" does not include a spam email filter, firewall, antivirus software, identity and access management tools, calculator, database, dataset, or other compilation of data.
    b) "High-risk automated decision system" means an ADS that is used to assist or replace human discretionary decisions that have a legal or similarly significant effect, including decisions that materially impact access to, or approval for, housing or accommodations, education, employment, credit, health care, and criminal justice. (Gov. Code § 11546.45.5.)

This bill:

1) Requires the AG to designate one or more MROs pursuant hereto by determining whether an applicant MRO's plan ensures acceptable mitigation of risk from any MRO-certified AI models and applications by considering a series of factors:
    a) The applicant's personnel and the qualifications of those personnel.
    b) The quality of the applicant's plan with respect to ensuring that artificial intelligence model and application developers exercise heightened care and comply with best practice-based standards for the prevention of personal injury and property damage, considering factors including, but not limited to, both of the following:
        i. The viability and rigor of the applicant's evaluation methods, technologies, and administrative procedures.

ii. The adequacy of the applicant's plan to develop measurable standards for evaluating artificial intelligence developers' mitigation of risks.

c) The applicant's independence from the artificial intelligence industry.
d) Whether the applicant serves a particular existing or potential artificial intelligence industry segment.

2) Requires these plans to include the following elements:
   a) The applicant's approach to auditing of artificial intelligence models and artificial intelligence applications to verify that a developer has exercised heightened care and adhered to predeployment and postdeployment best practices and procedures to prevent personal injury or property damage.
   b) The applicant's approach to mitigating specific high-impact risks, including cybersecurity, chemical, biological, radiological, and nuclear (CBRN) threats, malign persuasion, and artificial intelligence model autonomy and exfiltration.
   c) An approach to ensuring disclosure by developers to the MRO of risks detected, incident reports, and risk mitigation efforts.
   d) An approach to specifying the scope and duration of certification of an artificial intelligence model or artificial intelligence application, including technical thresholds for updates requiring renewed certification.
   e) An approach to data collection for public reporting from audited developers and vendors that addresses specified elements.
   f) The applicant's intended use, if any, of security vendors to evaluate developers, models, or applications, including a method of certifying and training vendors to accurately evaluate an artificial intelligence model or developer exercising heightened care and complying with best practices.
   g) Implementation and enforcement of whistleblower protections among certified developers.
   h) Remediation of postcertification noncompliance.
   i) An approach to reporting of societal risks and benefits identified through auditing.
   j) An approach to interfacing effectively with federal and non-California state authorities.

3) Prohibits the AG from modifying these plans.

4) Provides that a designated MRO shall:
   a) Certify developers' and security vendors' exercise of heightened care and compliance with best practices for the prevention of personal injury and property damage.
   b) Implement the plan submitted.
   c) Decertify an artificial intelligence model or application that does not meet the requirements prescribed by the MRO.

   d) Submit a specified report to the Legislature and to the AG annually that addresses specified details.
   e) Retain for 10 years a document that is related to the MRO's activities hereunder.

5) Provides that the applicant is to *audit itself* to ensure independence from industry, including assessment of its own board composition, availability of resources, and funding sources.

6) Provides that an MRO designation expires after three years, and the MRO may apply for a new designation.

7) Authorizes the AG to revoke a designation if any of the following are true:
   a) The MRO's plan is materially misleading or inaccurate.
   b) The MRO systematically fails to adhere to its plan.
   c) A material change compromises the MRO's independence from the artificial intelligence industry.
   d) Evolution of technology renders the MRO's methods obsolete for ensuring acceptable levels of risk of personal injury and property damage.
   e) An artificial intelligence model or artificial intelligence application certified by the MRO causes a significant harm.

8) Defines the relevant terms, including:
   a) "MRO" means an entity designated as an MRO by the Attorney General pursuant hereto that performs the functions specified in the bill, including certification of developers' exercise of heightened care and compliance with standards based on best practices for the prevention of personal injury and property damage with respect to an artificial intelligence model or application.
   b) "Security vendor" means a third-party entity engaged by an MRO or developer to evaluate the safety and security of an artificial intelligence model or application through processes that include red teaming, risk detection, and risk mitigation.
   c) "AI model" means an engineered or machine-based system that can, for explicit or implicit objectives, infer from the input it receives how to generate outputs that can influence physical or virtual environments.
   d) "AI application" means a software program or system that uses artificial intelligence models to perform tasks that typically require human intelligence.
   e) "Developer" means a person who develops an artificial intelligence model or artificial intelligence application that is deployed in the state.

9) Provides that in a civil action asserting claims for personal injury or property damage caused by an artificial intelligence model or artificial intelligence

application, it shall be an affirmative defense to liability that the artificial intelligence model or artificial intelligence application in question was certified by an MRO at the time of the plaintiff's injuries. This does not apply to claims of intentional misconduct.

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]

TechEquity, an organization committed to ensuring technology's evolution benefits everyone equitably, has also laid out their straightforward AI Policy Principles:

---

[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. All internet citations are current as of April 19, 2025.

- People who are impacted by AI must have agency to shape the technology that dictates their access to critical needs like employment, housing, and healthcare.
- The burden of proof must lie with developers, vendors, and deployers to demonstrate that their tools do not create harm—and regulators, as well as private [individuals], should be empowered to hold them accountable.
- Concentrated power and information asymmetries must be addressed in order to effectively regulate the technology.

The need for thoughtful regulation and accountability is especially urgent with regard to the existential risks that many believe unfettered AI advancement poses. It may seem like ancient history, but, in response to these risks, the Future of Life Institute published an open letter in 2023, calling for a pause on giant AI experiments:

> Contemporary AI systems are now becoming human-competitive at general tasks, and we must ask ourselves: *Should* we let machines flood our information channels with propaganda and untruth? *Should* we automate away all the jobs, including the fulfilling ones? *Should* we develop nonhuman minds that might eventually outnumber, outsmart, obsolete and replace us? *Should* we risk loss of control of our civilization? Such decisions must not be delegated to unelected tech leaders. **Powerful AI systems should be developed only once we are confident that their effects will be positive and their risks will be manageable.** This confidence must be well justified and increase with the magnitude of a system's potential effects. OpenAI's recent statement regarding artificial general intelligence, states that "At some point, it may be important to get independent review before starting to train future systems, and for the most advanced efforts to agree to limit the rate of growth of compute used for creating new models." We agree. That point is now.
>
> Therefore, **we call on all AI labs to immediately pause for at least 6 months the training of AI systems more powerful than GPT-4.** This pause should be public and verifiable, and include all key actors. If such a pause cannot be enacted quickly, governments should step in and institute a moratorium.[2]

Signatories to the letter include Stuart Russell, Berkeley, Professor of Computer Science, director of the Center for Intelligent Systems, and co-author of the standard textbook "Artificial Intelligence: a Modern Approach"; Elon Musk, CEO of SpaceX, Tesla & X; and Steve Wozniak, Co-founder of Apple. Clearly no such pause has occurred.

---

[2] Future of Life Institute, *Pause Giant AI Experiments: An Open Letter* (2023) https://futureoflife.org/open-letter/pause-giant-ai-experiments/ [emphasis in original].

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.[3]

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.

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. Designating MROs to certify AI models and applications

This bill tasks the AG with designating MROs who are then qualified to certify AI models and applications. An MRO is defined circularly as an MRO designated as such that carries out the functions required in the bill. Essentially, any entity is eligible to

[3] 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.

apply to be an MRO. "Artificial intelligence application" means a software program or system that uses AI models to perform tasks that typically require human intelligence.

The AG must determine whether an applicant MRO's plan ensures acceptable mitigation of risk by considering various factors. These include assessing the qualifications of the MRO's personnel and how well adapted their plan is to ensure developers exercise heightened care and comply with best practices. The AG must also assess the applicant MRO's independence from the AI industry; however, there is no threshold set or specific factors to consider in making this assessment.

The AG is required to designate at least one MRO. According to the sponsor, Fathom, in its estimation, there are likely only a handful of entities that have the required expertise and other qualifications to successfully become designated as an MRO. Fathom identifies itself as one of those entities.

The required plan must contain specified elements. These include the applicant MRO's approach to: data collection and auditing, mitigating specific high-impact risks, ensuring proper disclosures from developers, and reporting of societal risks and benefits identified through auditing. The plan must also state the applicant's intended utilization of third-party security vendors and the method of certifying and training them to accurately evaluate an AI model or developer. The bill permits a plan to be "tailored to a particular artificial intelligence market segment" but does not limit the power of such an MRO to only certifying AI models and applications in that market segment.

As a stated check on ensuring these MROs are not simply serving the interests of the industry, the bill requires annual audits of the MRO's board composition, its availability of resources and funding sources, and the representation of civil society in its functions. However, the applicant will audit itself and report the findings to the AG. The bill prohibits the AG from modifying an MRO's plan, thus tying the hands of the AG to a certain extent.

MROs must reapply for this designation after three years. During these periods, the AG has the authority to revoke designations under certain circumstances, such as where the plan is misleading or inaccurate, the MRO fails to adhere to its plan, a certified AI model or application causes a significant harm, or where a change compromises the MRO's independence from the industry being certified. However, there is no specific provision for how an AG will or should monitor these factors, outside of receiving the MRO's internal audit and annual report. A number of groups have also raised concerns with both the resources of the AG to handle this massive new duty and the technical sophistication necessary to assess the technical aspects of an applicant MRO's plan.

Once designated, an MRO is required to certify developers' and security vendors' exercise of heightened care and compliance with best practices for the prevention of

personal injury and property damage. The MRO is required to implement its plan and decertify AI models or applications that do not meet the prescribed requirements.

There are also reporting and documentation retention requirements. MROs must submit an annual report to the AG and the Legislature that addresses, among other things, aggregated information on various factors, the adequacy of evaluation resources and mitigation measures, and developer and security vendor certifications.

A number of the procedures and definitions are arguably unclear. For instance, it is not clear whether the MROs are to certify AI developers or the models and applications they are developing, or both. Writing in opposition, the California Chamber of Commerce highlights some of these issues:

> While we appreciate the goal of promoting safety without harming innovation, we are concerned that this proposal currently raises a number of complex questions about how the proposed system would operate in practice, such as identifying the appropriate state entity for providing approval of an organization of this nature - where the focus should be on technical expertise, which in this case would not be the AG's office as currently identified in the bill. Additionally we have concerns over insufficient clarity and the lack of objective criteria (e.g., what constitutes "significant harm"? "systemic failure"? "heightened care"?; what is required for an "incident report"?; what are objective criteria that could be used in place of "acceptable mitigation of risk" or "viability and rigor"?); and concerns around the exclusion of industry participation in MROs and the potential omission of vital perspectives on implementing requirements at scale and across various use cases.

4. <u>A realignment of incentives</u>

Having independent third parties, experts in the field, evaluate whether or not a developer is complying with best practices and effectively mitigating risk is a strong policy aim. The "carrot" for getting certified could arguably be simply the trust that is established for deployers, consumers, and government procurement officers. However, this bill goes much farther and provides near total immunity against damages for personal injuries and property damages.

During the time an AI model or application is certified by an MRO, any party is immunized from liability for claims for personal injury or property damage if that model or application caused the damage, unless the defendant's misconduct is found to be intentional, an extremely high bar. This means a developer could widely distribute an AI application, and if it is certified, the developer, and any deployer or end user, are not responsible for any resulting injuries caused by the application, even if they are negligent.

For example, a developer that certifies an AI application intended for self-driving vehicles is sold to a company that includes them in their fleet. Despite noticing serious safety issues regarding the capability of the AI application, the company deploys the vehicles onto public streets leading to the death of a pedestrian. This bill prevents the family of the pedestrian from holding either the company or the developer liable. The bill provides only for the MRO to have their designation revoked. The only remaining party is potentially the AG, which designated the MRO that improperly certified the AI application/developer.

A similar hypothetical could apply to situations addressed by a recent bill before this Committee, SB 243 (Padilla, 2025). If a certified AI application used in a companion chatbot causes it to encourage users to commit suicide, the family of a victim would be restricted from bringing a claim, despite the cause of action provided for in SB 243.

Providing immunity from liability based on private entity certification of AI models and applications creates a dangerous gap in consumer protection that undermines safety incentives. Such an approach fundamentally misaligns these incentives in ways that can increase public risk. Rather than the risk of liability motivating actors to proactively seek to reduce the likelihood of harm to others, deployers, for instance, are incentivized not to examine the potential dangers of a certified model or application they deploy, so that they cannot be shown to have intentionally caused resulting harms.

Negligence law serves a crucial purpose in our legal system by incentivizing companies to take reasonable precautions to prevent foreseeable 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. This creates a direct financial incentive to prioritize safety throughout a product's lifecycle. This is the state of the law currently.

Private certification with immunity from liability removes this essential motivation. Once certified, developers would have little economic reason to continuously assess emerging risks or implement proactive safety measures beyond minimum certification requirements. This creates a "check-the-box" approach to safety rather than the constant vigilance required for rapidly evolving AI technologies, as discussed above. It is unclear why this technology should be afforded a different legal standard that does not hold developers and deployers accountable for acting without due care to avoid harm.

Furthermore, a law allowing private companies to certify AI safety could arguably create a race to the bottom. MROs would face an inherent conflict of interest. When MROs compete for business, they are incentivized to lower standards to appeal to potential clients. While the AG has the authority to designate MROs and revoke that designation, there are not robust oversight mechanisms beyond that, and the capacity of the AG to continuously monitor the monitors is unclear. The only repercussions

provided for in the bill, despite the level of misconduct, is a revocation of their designation.

No certification process can anticipate all potential risks of complex AI systems that operate in open-ended environments and evolve through ongoing learning. Rather than immunity, a balanced approach that maintains liability for failure to take proper care while establishing clear standards for responsible actors would better protect public safety while still fostering innovation in AI technologies. A relevant article in the International Journal of Law and Information Technology, while focused on the law in Europe, provides relevant analysis of the importance of liability regimes in ensuring proper incentives for AI safety:

> Responsible AI requires robust forward-looking governance, and at its core there must be questions of who should be liable if AI harms humans and under which circumstances. We posit that there can be no responsible AI without AI liability. There can also be no AI safety without AI liability, [i.e.] a clear and comprehensive liability framework for AI, one that would provide strong incentives to develop and deploy systems that are safe by giving victims easy ways to access compensation.[4]

Last year, 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. The bill passed the Legislature but was ultimately vetoed. SB 1047 would have created a floor for AI governance, albeit one that many thought too high. This bill takes the very opposite approach by outsourcing the critical role of government oversight and setting a ceiling, asking developers to do just enough to achieve and maintain certification and held harmless for the damages that follow. Given the potentially existential risks and the near ubiquity of AI deployment in every facet of our lives, arguably stronger incentives for constant vigilance and risk mitigation are necessary.

According to the author:

> California is a world leader in AI development, so it is incumbent on our state to ensure that the use of artificial intelligence is safe and beneficial. To do so, it is imperative that we establish strong yet workable standards — standards created by independent, third-party experts and academics who can nimbly adapt to evolving technology.

---

[4] Guido Noto La Diega & Leonardo Bezerra, *Can there be responsible AI without AI liability? Incentivizing generative AI safety through ex-post tort liability under the EU AI liability directive*, International Journal of Law and Information Technology, Volume 32, 2024, https://doi.org/10.1093/ijlit/eaae021.

> SB 813 is an innovative and pragmatic approach to ensuring that artificial intelligence is developed responsibly. With the public-private governance concept, we can both advance high-level standards to improve consumer awareness and safety, while also not constraining California developers with endless red tape.

5. Stakeholder positions

TechNet writes in opposition:

> Rather than establishing bespoke organizations and frameworks from scratch, we encourage California to build on the work of established, globally respected standards-setting bodies, such as the International Organization for Standardization (ISO) and the American National Standards Institute (ANSI). Agencies like the California Air Resources Board and the Department of Motor Vehicles have successfully worked through these channels, demonstrating their effectiveness in balancing stakeholder input with technical rigor.
>
> We are also concerned that industry perspectives may be underrepresented in the proposed MRO structure. Excluding key stakeholders could hinder the development of practical, scalable standards that reflect real-world implementation challenges and the deep subject-matter expertise within industry. Overly academic approaches to AI safety risk emphasizing outward-facing disclosures while overlooking critical internal processes such as risk testing, mitigation protocols, and product-specific safety guardrails.

A group of academics and civil society experts in AI, including some of the people quoted above, write in support of the bill:

> Advanced AI technology is ever-changing, which makes it incredibly difficult to envision the technology's nearly infinite future capabilities or to forecast exactly when those capabilities will come online. This dynamic complicates traditional government agencies' ability to regulate this important technology. However, the pace of innovation does not obviate the need for sensible guardrails. To the contrary, the pace of AI innovation proves that our society needs creative approaches to governance that allow the technology to flourish and ensure widespread adoption based on trust and legal and regulatory clarity.
>
> SB 813 is the first-of-its-kind AI governance framework that is both nimble and built upon proven regulatory models that will continue to spur innovation and incentivize AI platforms to comply with state-of-the-art

> requirements to identify, monitor, and mitigate known, foreseeable risks. By establishing a "third-way" governing model, independent experts will be able to devise strong safety standards that also promote innovation while still being accountable to government leaders. This legislation harnesses the benefits of AI while also curbing its potential excesses.

The Consumer Attorneys of California writes in opposition:

> The courts are already well-equipped to assess whether a defendant acted reasonably based on the facts of a given case, including whether the defendant complied with industry standards or reasonable safety practices. If an MRO generates meaningful and independent guardrails, courts can already determine whether this sets a standard of care or duty. If the MRO establishes meaningful standards for protections, then a company will be allowed to argue that its compliance with those standards fulfilled its duty of care.
>
> But if the MRO fails to set adequate standards, harmed California consumers should not be stripped of their rights through an unjustified affirmative defense. Courts regularly determine standards of care and whether a defendant's conduct meets those standards. Courts can and should be allowed to perform this function in this context. Granting immunity erases the very purpose of civil justice. Letting companies "get away with it" is not how public trust in AI—or any other technology—is built.

A coalition of groups, the California Initiative for Technology & Democracy (CITED), Children's Advocacy Institute, and Tech Equity explain their opposition:

> SB 813 shields AI developers from liability for tort injuries and property damages caused by their AI products for everything but their intentional acts. This would allow a developer responding to business pressures or a competitive marketplace to recklessly cut corners in design and training of an AI system and rush unsafe products to market.
>
> For example, if an AI drone armed with a weapon accidentally takes down a passenger airplane, under your bill the AI developer would be relieved of liability as long as the AI developer did not actually intend to shoot down the plane, provided that the AI system had been certified by a private regulator. In this scenario, the AI developer would still be shielded from liability even if the Attorney General, who is charged with designating the private multistakeholder regulatory organizations (MROs), had taken steps to remove the MRO's designation for failing to

comply with requirements and even if the MRO had begun to decertify the AI developer for failing to comply with its safety guidelines.

Designers, developers, manufacturers, and sellers make products safe in part to reduce their liability for any harm their products may cause. If they are immunized from liability, there is less business reason to ensure the safety of their products. Given the immense potential risks associated with AI, including scams, voter manipulation and disinformation, bias, discrimination, child sexual abuse material and non-consensual intimate imagery, not to mention additional catastrophic risks, we believe that AI developers should be especially incentivized to make their products safe and keep Californians protected from the foreseeable harms of AI technology.

Because of the harm the liability shield will cause to innocent victims, we strongly urge you to remove it from the bill. There are other ways to incentivize AI developers to agree to be regulated, but we think the best option would be simply for policy makers to appropriately require that they be regulated just like other products that have the potential for substantial harm.

Second, we have significant concerns with the bill's directive to privatize government regulatory authority over such a critical sector of our economy and our lives. The responsibilities of the Attorney General under the bill are too general and constrained, and they fail to provide sufficient legislative direction to the Attorney General for determining that "an applicant MRO's plan ensures acceptable mitigation of risk," the "adequacy of the applicant's plan to develop measurable standards for evaluating artificial intelligence developers' mitigation of risks," and the "applicant's independence from the artificial intelligence industry." We believe much more clear and enforceable definitions and directives would be needed before such decisions should be delegated by policy makers to private entities.

Also, as currently drafted the bill appears to require the Attorney General to designate at least one entity as an MRO (even if the Attorney General determines that no applicants actually meet the minimal requirements in the bill), and severely limits the Attorney General's discretion to revoke an MRO's designation, even when the MRO has been misleading or inaccurate (though not "materially") in its application, even if the MRO fails to adhere (though not "systemically") to its plans; and even if certified AI models causes multiple and ongoing harm (though not "significant" harm).

The requirements for MROs also cause concern. MROs are given significant and overbroad discretion in the certifying and decertifying of developers and security vendors. It is not clear whether there are any entities currently in the United States other than perhaps the sponsor of this legislation who might be able to comply with the current MRO qualifications in the measure.

6. Amendments

In response to the concerns outlined above, the author has agreed to a series of amendments that do the following:

- Clarify that MROs shall certify specific AI models and applications, not individuals or entities.
- Replace the affirmative defense provided by the bill to a rebuttable presumption as follows:
    - Amend Section 8898.4 to read:
      (a) In a civil action asserting claims for personal injury or property damage caused by an artificial intelligence model or artificial intelligence application against a developer of the model or application, there shall be a rebuttable presumption of due care on the part of the developer if the artificial intelligence model or artificial intelligence application in question was certified by an MRO at the time of the plaintiff's injuries.
      (b) The rebuttable presumption provided for in subdivision (a) may be overcome by the introduction of admissible evidence the court finds is contrary to the presumption.
- Include a requirement that the Attorney General promulgate regulations, with input from stakeholders, that provide:
    - Baseline requirements for plans required to be submitted pursuant to Section 8898.2.
    - Conflict of interest rules for MROs that include, but are not limited to, reporting requirements on boards of directors and donors funding the MRO to ensure adequate independence from the artificial intelligence industry and transparency on revenues streaming from certification services.
- Authority for the Attorney General to develop a fee structure for offsetting costs incurred by the Attorney General in relation to carrying out its duties pursuant to this chapter.

SUPPORT

Fathom (sponsor)
21 individuals

OPPOSITION

California Chamber of Commerce
California Initiative for Technology & Democracy
Chamber of Progress
Children's Advocacy Institute
Consumer Attorneys of California
Tech Equity Action
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 Senate Health 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 Senate Appropriations 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.

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 the Assembly Privacy and Consumer Protection Committee.

AB 1405 (Bauer-Kahan, 2025) establishes an enrollment process within the Government Operations Agency (GovOps) for AI auditors. Enrolled auditors could then perform "covered audits," which are audits mandated by the legislature or regulations. It establishes a central repository within GovOps through which one could find an auditor. AB 1405 is currently in the Assembly Appropriations Committee.

Prior Legislation:

SB 892 (Padilla, 2024) would have required CDT to develop and adopt regulations to create an ADS procurement standard, as specified, and prohibited a state agency from procuring ADS, entering into a contract for ADS, or any service that utilizes ADS, until CDT has adopted regulations creating an ADS procurement standard, as specified. SB 892 was vetoed by Governor Newsom, who stated in his veto message that aspects of the bill would disrupt ongoing work, "including existing information technology modernization efforts, which would lead to implementation delays and higher expenses for critical projects."

SB 1047 (Wiener, 2024) *See* Comment 3. 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.

AB 302 (Ward, Ch. 800, Stats. 2023) required CDT, on or before September 1, 2024, to conduct a comprehensive inventory of all high-risk ADS that have been proposed for use, development, or procurement by, or are being used, developed, or procured by, any state agency.

AB 331 (Bauer-Kahan, 2023) was substantially similar to AB 2930. AB 331 died in the Assembly Appropriations Committee.

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

****