California Senate Bill 813: A Novel Approach to Artificial Intelligence Governance

Google Gemini 2.0 Flash Deep Research prompted by Kris Carlson

You can get a shorter overview of the proposed bill by scrolling 60% down to "IV. Translation into Plain English: Key Provisions of SB 813."

I. Executive Summary

California Senate Bill 813 proposes a distinct method for governing the rapidly advancing field of artificial intelligence (AI). The bill seeks to establish a process wherein the state's Attorney General (AG) would designate private entities as Multistakeholder Regulatory Organizations (MROs). These MROs, if designated, would play a pivotal role in the AI ecosystem by voluntarily certifying the safety of AI models and applications. This certification process aims to ensure that AI developers adhere to heightened standards of care and best practices, particularly concerning the prevention of personal injury and property damage. A key aspect of this legislative proposal is the introduction of potential legal benefits for AI developers who obtain MRO certification, including an affirmative defense and a rebuttable presumption of reasonable care in civil actions related to their certified AI products. This initiative appears to be a unique attempt to blend the agility and expertise of the private sector with the oversight and authority of a government entity in the complex and evolving landscape of AI governance.

II. Introduction: The Evolving Landscape of AI Governance

The proliferation of artificial intelligence across numerous sectors of society marks a significant technological shift, bringing with it both unprecedented opportunities and potential risks.[1] As AI models and applications become increasingly sophisticated and integrated into daily life, novel safety and security concerns emerge.[1] Automated systems, for instance, can make critical errors in high-stakes scenarios, and AI-powered tools can be exploited for malicious purposes.[1] Recognizing the profound impact of AI, there is a growing global effort to establish governance frameworks that promote its safe, private, and ethical use.[2] These frameworks are aimed at improving transparency, accountability, and the fair use of AI technologies.[3] They provide a structured approach to addressing concerns such as bias, privacy infringement, and misuse, while simultaneously fostering innovation and building public trust.[4]

California has historically been at the forefront of regulating emerging technologies, and it continues this trend with a proactive approach to AI governance.[6] The state has already enacted legislation such as the California AI Transparency Act, which mandates certain disclosures for generative AI systems. Furthermore, a series of new AI-related laws came into effect in California in 2025, addressing issues ranging from deepfake technology to data privacy and the use of AI in healthcare.[6] Senate Bill 813 represents a novel addition to this evolving landscape, proposing a multistakeholder regulatory approach to AI safety and innovation.

III. Deconstructing Senate Bill 813: A Framework for Multistakeholder AI Regulation

A. Establishment of Multistakeholder Regulatory Organizations (MROs)

At the heart of SB 813 is the concept of Multistakeholder Regulatory Organizations (MROs). The bill envisions these as private entities that would be designated by the Attorney General. The

Legislature explicitly states its findings that an MRO tasked with defining standards based on best practices and certifying adherence to them represents an agile, public-private model. This model is designed to promote innovation, ensure the security of AI platforms, reduce regulatory uncertainty, and build societal trust in AI. The designation as an MRO would be for a renewable period of three years, allowing for periodic review and reassessment. This approach reflects a belief that by proactively setting clear standards and offering legal and economic incentives, compliance can become a competitive advantage, accelerating responsible growth in the AI-driven economy.

B. The Role of the Attorney General

The Attorney General plays a crucial oversight role in the framework proposed by SB 813. The AG is empowered to designate one or more MROs based on whether an applicant's plan ensures acceptable mitigation of risk from MRO-certified AI models and applications. In making this determination, the AG is required to consider various factors, including the applicant's personnel, the quality of their risk mitigation plan, their independence from the AI industry, and whether they serve a specific AI industry segment. Furthermore, the bill mandates that the Attorney General adopt regulations, with input from stakeholders, to establish minimum requirements for these risk mitigation plans and to set conflict of interest rules for MROs. These regulations would also include reporting requirements for MRO boards of directors and donors to ensure adequate independence and transparency regarding revenues from certification services. The AG is also authorized to establish a fee structure for applicants and designated MROs to offset the reasonable costs incurred in carrying out their duties under the bill. Importantly, the Attorney General retains the authority to revoke an MRO's designation if certain conditions are met. These conditions include a materially misleading or inaccurate plan, systematic failure to adhere to the plan, a material change compromising the MRO's independence, technological evolution rendering the MRO's methods obsolete, or if an AI model or application certified by the MRO causes significant harm.

C. Requirements for MRO Designation

Private entities seeking designation as an MRO must meet specific criteria to demonstrate their capability to effectively regulate AI safety. A primary requirement is the submission of a comprehensive plan that ensures acceptable risk mitigation from MRO-certified AI models and applications. The Attorney General will assess the applicant's personnel and their qualifications. A critical aspect is the quality of the applicant's plan in ensuring that AI developers exercise heightened care and comply with best practice-based standards to prevent personal injury and property damage. This assessment will consider the viability and rigor of the applicant's evaluation methods, technologies, and administrative procedures, as well as the adequacy of the plan to develop measurable standards for evaluating developers' risk mitigation efforts. The applicant's independence from the AI industry is another key consideration, along with whether the applicant serves a particular existing or potential AI industry segment.

The plan submitted by an applicant must contain several essential elements. This includes the applicant's approach to auditing AI models and applications to verify that developers have exercised heightened care and adhered to best practices both before and after deployment to prevent harm. The plan must also detail the applicant's strategy for mitigating specific high-impact

risks, such as cybersecurity threats, chemical, biological, radiological, and nuclear threats, malign persuasion, and AI model autonomy and exfiltration. Furthermore, the plan needs to outline how developers will be required to disclose detected risks, incident reports, and risk mitigation efforts to the MRO. The scope and duration of certification for AI models and applications, including the technical thresholds for updates requiring renewed certification, must also be specified. A crucial component is the approach to data collection for public reporting from audited developers and vendors, addressing the aggregation and tracking of evaluation data, categories of metadata, and measures to protect trade secrets and mitigate antitrust risks from information sharing. If the applicant intends to use security vendors, their plan must include a method for certifying and training these vendors to accurately evaluate AI models and developers. The plan must also detail the implementation and enforcement of whistleblower protections among certified developers, the remediation of post-certification noncompliance, an approach to reporting societal risks and benefits identified through auditing, and a strategy for effectively interfacing with federal and non-California state authorities. The bill acknowledges that these plans can be tailored to a particular AI market segment. To further ensure independence from the AI industry, applicants are required to annually audit their board composition, resource availability, funding sources, and representation of civil society representatives in evaluation and reporting functions, and report these findings to the Attorney General. It is important to note that the Attorney General is explicitly prohibited from modifying a plan submitted for designation.

D. Responsibilities and Functions of MROs

Once designated by the Attorney General, MROs are tasked with several key responsibilities related to ensuring AI safety. Their primary function is to certify that developers and security vendors exercise heightened care and comply with best practices for preventing personal injury and property damage. This includes certifying qualified AI models and applications that meet the requirements prescribed by the MRO. MROs are also responsible for implementing the plan they submitted for designation. They have the authority to decertify an AI model or application that fails to meet their prescribed requirements. A significant responsibility is the submission of an annual report to the Legislature and the Attorney General. This report must address aggregated information on the capabilities of AI models, the observed and potential societal risks and benefits associated with these capabilities, the adequacy of existing evaluation resources and mitigation measures, developer and security vendor certifications, aggregated results of certification assessments, remedial measures prescribed by the MRO and the compliance of developers and vendors, and any additional risks identified beyond personal injury or property damage. Finally, MROs are required to retain all documents related to their activities under this chapter for a period of ten years.

E. Legal Implications of MRO Certification

SB 813 introduces significant legal implications for AI models and applications certified by an MRO. In a civil action asserting claims for personal injury or property damage caused by an AI model or application against a developer, the bill provides an affirmative defense to liability if the AI in question was certified by an MRO at the time of the plaintiff's injuries. Furthermore, the bill creates a rebuttable presumption that the developer exercised reasonable care if the AI model or application was MRO-certified at the time of the injury. This presumption can be overcome by the introduction of admissible evidence contrary to it. However, this affirmative defense and

rebuttable presumption do not apply to claims of intentional misconduct by the defendant. This legal framework aims to incentivize developers to seek MRO certification by offering a degree of protection from liability, thereby encouraging the development and deployment of safer AI technologies.

F. Definitions of Key Terms

To ensure clarity and precision, SB 813 provides definitions for several key terms. An "artificial intelligence application" is defined as a software program or system that uses AI models to perform tasks typically requiring human intelligence. An "artificial intelligence model" refers to an engineered or machine-based system that can infer from input how to generate outputs that can influence physical or virtual environments. A "developer" is defined as a person who develops an AI model or application that is deployed in the state. The term "multistakeholder regulatory organization (MRO)" denotes an entity designated as such by the Attorney General, performing functions like certification to ensure developers exercise heightened care and comply with best practices to prevent personal injury and property damage. A "plan" refers to the plan submitted by an applicant seeking MRO designation. Finally, a "security vendor" is a third-party entity engaged by an MRO or developer to evaluate the safety and security of an AI model or application through processes like red teaming and risk mitigation.

IV. Translation into Plain English: Key Provisions of SB 813

In simpler terms, California Senate Bill 813 proposes creating a system where private organizations can become officially recognized "AI safety inspectors" by the state's Attorney General. These organizations, called Multistakeholder Regulatory Organizations (MROs), would develop their own rules and standards for AI safety, focusing on preventing harm to people and property. AI developers could then choose to have their AI systems certified by these MROs if they meet the standards. To become an MRO, a private group would need to convince the Attorney General that they have a strong plan to ensure AI safety, have qualified people, are independent from the AI industry, and can effectively check if AI developers are following best safety practices. They would also need to detail how they will handle potential high-risk scenarios like cybersecurity breaches or misuse of AI.

Once approved, these MROs would act as watchdogs, regularly checking if certified AI systems and their developers are following the rules. If an AI system causes harm despite being certified, or if the MRO's safety methods become outdated, the Attorney General can revoke their "inspector" status. A major incentive for AI developers to get certified is that if their certified AI causes injury or damage, they might have a stronger legal defense in court, suggesting they took reasonable steps to ensure safety. However, this protection wouldn't apply if the harm was intentional. This whole system aims to encourage the AI industry to prioritize safety by offering them a way to demonstrate their commitment and potentially reduce their legal risks, all under the watchful eye of the state government.

V. Novel Governance Proposals in SB 813: A Unique Approach?

The approach to AI governance proposed by SB 813 presents several novel elements when compared to other existing and proposed frameworks.

Government-led regulation: Traditional government regulation typically involves state agencies directly setting and enforcing rules.[3] SB 813, however, takes a different path by empowering private entities to develop and enforce safety standards, albeit under the oversight and designation of the Attorney General. While the AG sets minimum requirements and conflict-of-interest rules, the MROs have significant autonomy in defining the specific standards and evaluation methods. This contrasts with a purely top-down regulatory approach.

Industry self-regulation: Pure industry self-regulation relies on voluntary commitments and standards developed by AI companies themselves.[8] While many companies are proactively implementing ethical AI principles [9], concerns exist about the sufficiency of these voluntary measures to prevent harm.[9] SB 813 bridges this gap by establishing a framework where private organizations are empowered to regulate, but with a formal designation and oversight mechanism from the government, adding a layer of public accountability that is often absent in pure self-regulation.

Existing multistakeholder initiatives: The concept of multistakeholder governance, involving collaboration between government, industry, civil society, and academia, is not entirely new in technology policy, particularly in areas like internet governance.[10] These initiatives often focus on collaborative problem-solving and the development of shared principles.[11] SB 813 formalizes this concept by creating designated MROs with specific legal powers related to certification and risk assessment. The direct involvement of the Attorney General in designating and overseeing these organizations, coupled with the legal incentives tied to their certifications, distinguishes this approach from more general multistakeholder dialogues and frameworks.

Private certification bodies: Various organizations offer AI certifications.[16] These certifications typically aim to validate the skills and knowledge of AI professionals or assess the ethical considerations of AI systems. However, none of these existing private certifications appear to carry the same government-backed legal implications as the MRO certification proposed in SB 813, which includes an affirmative defense and a rebuttable presumption in civil liability cases.

The unique elements of SB 813's approach include the specific legal incentives tied to certification by a private organization that is formally designated by a government entity. The detailed requirements for MROs and their risk mitigation plans, which cover a broad spectrum of high-impact risks, also stand out. Furthermore, the formal oversight role of the Attorney General, with the power to both designate and revoke MRO status, provides a level of government accountability that is not typically found in purely private or self-regulatory models. This combination of private sector agility, government oversight, and legal incentives appears to be a novel approach to AI governance.

VI. Comparison with Existing AI Governance Frameworks

The table below compares SB 813's MRO model with other prominent AI governance frameworks:

Framework	Approach	Enforcement Mechanism	Focus	Novelty

SB 813 (MRO)	Private regulation with government oversight	Attorney General designation and revocation, legal incentives	Safety, risk mitigation, innovation, trust	Legal incentives for private certification, broad risk focus, AG oversight
EU AI Act	Risk-based regulation	Government fines and prohibitions	Ethical use, safety, fundamental rights	Risk-based categorization and prohibitions, emphasis on high-risk systems
NIST AI Risk Management Framework	Voluntary framework	Voluntary adoption	Trustworthy and responsible AI development and use	Comprehensive guidance on identifying, evaluating, and managing AI risks
California AI Transparency Act	Transparency mandate	Civil actions by Attorney General, city attorneys, county counsel	Transparency of generative AI content and training data	Mandates disclosure of AI-generated content and summaries of training data
Colorado's AI Law	Anti-bias focus	Government fines	Preventing algorithmic discrimination in high-risk AI	Focus on developers' duty to avoid algorithmic discrimination
Texas Responsible AI Governance Act	Sector-specific rules	AI Council oversight and fines	Financial services, healthcare, criminal justice	Sector-specific rules for high-risk AI, establishment of an AI Council for oversight and enforcement

VII. Potential Impacts and Considerations

The MRO framework proposed by SB 813 holds the potential for several positive impacts on AI governance. Its agility and adaptability could allow for quicker responses to the rapid advancements in AI technology compared to traditional legislative processes. By leveraging the specialized expertise of the private sector, the framework may lead to more informed and effective safety standards tailored to specific AI applications. The establishment of clear standards and the potential for legal protection could reduce regulatory uncertainty for businesses, fostering innovation in the AI industry. The certification process and the required transparency measures could contribute to building greater public trust in AI technologies. Furthermore, the bill explicitly aims to create a framework that can effectively interface with federal and international AI governance efforts, potentially leading to greater harmonization across different jurisdictions.[28]

However, several challenges and considerations need to be addressed for the successful implementation of this framework. Ensuring the genuine independence of MROs from the AI industry and preventing regulatory capture will be crucial]. There is a potential risk of inconsistent standards or varying levels of rigor across different MROs, which could lead to confusion and undermine the overall effectiveness of the system. The voluntary nature of MRO certification raises questions about the extent to which developers will adopt it, especially if the perceived benefits do not outweigh the costs and complexities involved. The delegation of regulatory authority to private entities could also face legal challenges. The cost and complexity for businesses to obtain and maintain MRO certification need to be carefully considered to avoid hindering smaller companies and startups. Finally, given the dynamic nature of AI, the MRO framework will require ongoing evaluation and adaptation to remain relevant and effective in the face of future technological advancements. AI developers may view MRO certification not just as a compliance requirement but as a strategic asset that offers both legal protection and a potential reputational advantage.[28]

VIII. Conclusion

California Senate Bill 813 proposes a novel approach to artificial intelligence governance through the establishment of Multistakeholder Regulatory Organizations. This framework seeks to combine the strengths of private sector expertise and agility with government oversight and legal incentives to promote the safe and responsible development and deployment of AI technologies. By empowering the Attorney General to designate private entities as MROs, the bill aims to create a dynamic and adaptable regulatory mechanism that can evolve alongside the rapidly changing landscape of AI. The legal benefits tied to MRO certification, including an affirmative defense and a rebuttable presumption of reasonable care in civil actions, offer a significant incentive for AI developers to prioritize safety and adhere to best practices. While the MRO model presents several potential benefits, including reduced regulatory uncertainty and increased public trust, it also raises important considerations regarding the independence and consistency of these private regulatory bodies. The success of this unique approach will likely depend on the careful implementation of the Attorney General's oversight role and the willingness of the AI industry to embrace voluntary certification as a means of ensuring both safety and legal security. Ultimately, SB 813 represents a significant and potentially groundbreaking step in the ongoing effort to navigate the complex challenges and opportunities presented by artificial intelligence.