WINSTON
& STRAWN
LLP

ARTICLE

Navigating California’s SB 1047: Implications for AI Regulation and Industry Impact

SEPTEMBER 20, 2024

This article was originally published in The Recorder. Reprinted with permission. Any opinions in this article are not those of Winston & Strawn or its clients. The opinions in this article are the authors’ opinions only.

This article aims to provide a comprehensive examination of California’s currently pending SB 1047, summarizing its key provisions in plain language, exploring the perspectives of both proponents and opponents, and assessing the practical implications of the bill.

As artificial intelligence (AI) continues its rapid advancement, governments worldwide are increasingly focusing on regulating the transformative technology. In the United States, the Biden administration’s Executive Order on AI has addressed some of AI’s opportunities and challenges, while the EU adopted its comprehensive AI Act earlier this year, and China has introduced its own regulatory framework. Amid this global trend, on Aug. 28, 2024, California’s Senate passed SB 1047—a bill designed to govern foundational AI models and their derivatives.

The bill focuses on enhancing transparency, accountability, and fairness in AI development, emphasizing safety protocols, whistleblower protections, third-party audits, and the capability to fully shut down AI systems (i.e., “AI kill switches”) to proactively mitigate potential risks posed by advanced AI models.

The legislation now awaits Gov. Gavin Newsom’s decision by Sept. 30. Given California’s pivotal role as a leader in technology and regulation (e.g., privacy), all eyes are on Newsom to see whether the state will continue to set precedents in AI governance or heed concerns that the bill’s adoption may be premature.

SUMMARY OF KEY PROVISIONS OF SB 1047

- **Definition of Covered Model and Derivatives (22602 (e) and (f))**

The bill applies to developers of “covered models,” which are AI models meeting specific computational power and cost thresholds (Sec. 22602(e)). A covered model is defined as:

Before Jan. 1, 2027:

- An AI model trained using more than 1026 integer or floating-point operations, with training costs exceeding $100 million, calculated using average market prices at the start of training (Sec. 22602(e)(1)(A)(i)).

- An AI model fine-tuned from a covered model using at least 3×1025 operations with fine-tuning costs exceeding $10 million, using the same calculation (Sec. 22602(e)(1)(A)(ii)).

On and After Jan. 1, 2027:

- The Government Operations Agency (GOA) will determine new compute thresholds. If the GOA doesn't set new thresholds by this date, the pre-2027 definitions remain in effect until updated.

Commentary*: The dollar amount of training costs is to be "reasonably assessed by the developer," which may raise transparency concerns with respect to developer's evaluations of their own models or derivatives. Additionally, thresholds are subject to change based on GOA regulations, introducing potential uncertainty for developers planning long-term projects.*

Pre-training Requirements (22603 (a))

Before beginning to initially train a covered model, developers must:

- **Implement Cybersecurity Protections (Sec. 22603(a)(1)):**
    - Establish reasonable administrative, technical, and physical safeguards to prevent unauthorized access, misuse, or unsafe post-training modifications, considering risks from advanced persistent threats.
- **Enable Full Shutdown Capability (Sec. 22603(a)(2)):**
    - Develop the ability to promptly shut down the covered model and its derivatives, accounting for potential disruptions to critical infrastructure.
- **Establish a Safety and Security Protocol (Sec. 22603(a)(3)):**
    - **Documentation:** Create a detailed, written protocol outlining procedures to manage risks across the model's lifecycle.
    - **Compliance Measures:** Specify objective compliance requirements and testing procedures to evaluate risks of causing critical harms.
    - **Safeguard Implementation:** Describe how safeguards will be applied and under what conditions a full shutdown would occur.
- **Designate Responsible Personnel (Sec. 22603(a)(4)):**
    - Assign senior staff to ensure compliance and monitor implementation of safety protocols.
- **Retain Documentation (Sec. 22603(a)(5)):**
    - Keep unredacted copies of the safety protocol and any updates for at least five years after the model is no longer in use.
- **Annual Review (Sec. 22603(a)(6)):**
    - Review and, if necessary, update the safety and security protocol annually to reflect changes in the model's capabilities and industry best practices.
- **Public Disclosure (Sec. 22603(a)(7)):**
    - Publish a redacted version of the safety and security protocol, with redactions only for public safety, trade secrets, or confidential information.
    - Provide unredacted versions to the Attorney General upon request.

Commentary: *The requirement to publish redacted safety protocols seems to strike a balance between transparency and the protection of sensitive confidential information. However, the extent to which the "confidential* *information" exception (Sec. 22603(a)(7)(ii)(III)) may be relied upon may potentially limit the effectiveness of this provision.*

Pre-deployment Obligations (22603(b))

Before using a covered model (or derivative) for purposes beyond training or evaluation, or making it available for commercial or public use, developers must:

- **Risk Assessment (Sec. 22603(b)(1)):**
  - Assess whether the model could reasonably cause or materially enable critical harms.
- **Record-Keeping (Sec. 22603(b)(2)):**
  - Document specific tests and results with sufficient detail for replication, retaining records for at least five years after the model's use ends.
- **Implement Safeguards (Sec. 22603(b)(3)):**
  - Apply appropriate measures to prevent the model and its derivatives from causing or materially enabling critical harms.
- **Ensure Traceability (Sec. 22603(b)(4)):**
  - Ensure that the actions of the model and any resulting critical harms can be accurately and reliably attributed to it.

Restrictions on Use and Availability (22603 (c) and (d))

- Developers are prohibited from using or releasing a covered model (or derivatives) if there is an unreasonable risk that it will cause or materially enable critical harms.
- Developers must annually reevaluate procedures, policies, protections, capabilities, and safeguards to ensure ongoing compliance and risk mitigation.

Third Party Audits (22603(e))

Starting Jan. 1, 2026, developers must annually retain independent third-party auditors to assess compliance with the bill's requirements:

Audit Standards (Sec. 22603(e)(2)):

  - Conduct audits consistent with best practices and regulations issued by the Government Operations Agency.
- **Audit Reports (Sec. 22603(e)(4-5)):**
  - Produce detailed reports assessing compliance, internal controls, and any instances of noncompliance.
  - Retain unredacted copies for at least five years.
- **Disclosure Obligations (Sec. 22603(e)(6)):**
  - Publish redacted versions of the auditor's report, redacting only necessary information.
  - Provide unredacted reports to the Attorney General upon request, exempt from public records disclosure.

ADVOCATES AND CRITICS OF SB 1047

Advocates

Advocates of SB 1047 argue that the bill is necessary to proactively address the potential catastrophic harms that large, unregulated AI models could pose, emphasizing that without proper safeguards, AI systems might enable disasters such as cyberattacks on critical infrastructure or the development of bioweapons.

Advocates also argue that the bill's mandate for developers to test and implement safety protocols is vital for responsible development and deployment of AI, preventing future harms, and promoting equity, as vulnerable groups could be disproportionately affected by AI misuse.

Additionally, advocates point out that the bill includes provisions for whistleblower protections and third-party audits, enhancing transparency and accountability within the AI industry, which they argue the industry has trended away from in recent years, and contend that SB 1047 might catalyze broader adoption of robust AI safety standards, ultimately benefiting society at large.

Advocates see the bill not only as a protective measure for society, but also as an opportunity to influence global policies on AI safety. They argue that without such proactive regulation, industries and professionals could face unprecedented challenges related to the misuse of AI.

Critics

Critics of SB 1047 express significant concerns that the bill may stifle innovation, harm the state's economy, and impose undue burdens on AI developers without delivering substantial public safety benefits.

They also argue that the legislation is premature, given that methodologies for understanding and mitigating AI safety and security risks, particularly as written in the bill, are still underdeveloped. For example, the technical solutions and standards required for the bill's implementation, such as "industry best practices" as one of the pre-training requirements mentioned in Sec. 22603(a)(6), are still in their infancy, making compliance challenging and potentially ineffective.

Some critics suggest that issues of this scale and complexity are best addressed at the federal level rather than through state legislation to avoid a fragmented regulatory landscape that could complicate compliance for AI companies operating across multiple states. They contend that a national framework would not only provide clearer, more uniform standards, but also prevent states from enacting conflicting AI laws, which could also dampen innovation in a burgeoning industry.

Critics also assert that the bill focuses on extreme hypothetical risks, like AI creating weapons of mass destruction, while neglecting more immediate and demonstrable issues such as misinformation, discrimination, non-consensual deepfakes, environmental impacts, and workforce displacement, arguing that there is little scientific evidence supporting the likelihood of catastrophic harms as envisioned by the bill under current capabilities.

Further, critics warn that the stringent requirements might prompt a form of regulatory arbitrage where companies relocate to jurisdictions with more favorable regulatory environments, resulting in economic losses and a potential brain drain for the state.

Critics assert that they would more readily support a more measured approach that fosters innovation, while addressing more demonstrable risks through targeted legislation.

PRACTICAL IMPLICATIONS OF SB 1047

SB 1047 as a Template for Future Legislation

California has historically been a trailblazer in technology regulation, with laws like the California Consumer Privacy Act (CCPA) serving as models for other states, and it is possible that SB 1047 could continue this trend, inspiring similar state-level AI regulations around the country.

At the federal level, the prospect of SB 1047 serving as a template for federal regulation remains unlikely, given that neither the CCPA, in its six years, nor the GDPR, in its eight years, have led to a unified federal data privacy law. However, SB 1047 will likely, at the very least, reignite discussions about the need for comprehensive AI legislation, given that there is currently no unified federal approach. Discussion surrounding a federal right of publicity, for example, has already seen a renewed interest in the age of related state-level legislation such as Tennessee's ELVIS Act.

On the other hand, the EU's recently-adopted AI Act contrasts SB 1047's framework with a broader approach, categorizing AI systems based on a full spectrum of tiered risk levels, ranging from unacceptable risk, to high risk, to limited risk, all the way to minimal or no risk. Thus, it is also possible that future regulation may draw from both SB 1047 and the EU's AI Act, resulting in a hybrid approach that combines targeted oversight with comprehensive, risk-based categorization.

Impact of the Third-Party Audit Requirement

SB 1047 mandates annual independent third-party audits for developers of covered AI models (Sec. 22603(e)), which could lead to significant compliance costs, particularly for companies that would need to dedicate resources which they may not currently have in order to prepare for and undergo these audits.

Given that, should SB 1047 be adopted, the third-party audit requirement would come into effect in just over a year (Jan. 1, 2026), AI businesses would quickly be exposing themselves to a new degree of public relations challenges, and face even higher levels of scrutiny. While the bill allows for the redaction of confidential information and information that could threaten public safety, the disclosure of audit findings may still lead to a loss of trust among customers, investors, and partners. This may result in a compounding effect, whereby documented non-compliance could provoke further scrutiny, or even lawsuits by parties claiming harm caused by the company's AI models.

The need for specialized auditors with expertise in AI safety and compliance is also likely to create market demand for individuals with a deep understanding of AI and risk assessment, potentially spurring the creation of specialized training programs, certifications, or professional services focused on AI auditing.

The third party's audit must also be conducted "consistent with best practices" per Sec. 22603(e)(1), suggesting the establishment of objective criteria and standardized methodologies for auditing AI systems. Moreover, the audits must align with regulations that the GOA is tasked with issuing per Sec. 22603(e)(2), which would ultimately define binding auditing requirements.

Impact on Small and Large AI Companies Alike

While SB 1047 targets developers of AI models exceeding specific computational thresholds and costs, smaller companies may still be affected, as the cumulative expenses associated with AI development can add up quickly. The high costs of GPUs and other infrastructure necessary for training advanced AI models, along with the cost of salaries for AI experts and data scientists who are already in high demand, serve to only further escalate expenses associated with training and fine-tuning models.

Smaller companies may also struggle to compete with larger, more established organizations that have more resources to absorb compliance costs, which could also lead to a consolidation in the AI industry where only well-funded organizations could afford to develop advanced AI models within California.

Conversely, the bill could spur innovation in developing cost-effective compliance solutions. A novel branch of companies might emerge to create new tools, platforms, or services that help AI developers meet regulatory requirements more efficiently. This could open up new market opportunities and contribute to the growth of a compliance-focused tech sector within California and beyond.

International Corporations Facing Patchwork of Laws

Navigating a fragmented regulatory landscape demands substantial administrative resources, typically requiring localized compliance strategies to account for jurisdiction-specific regulatory nuances and continuous monitoring of regulatory changes, the burden of which will disproportionately fall on multinational companies that do business in such varied jurisdictions.

It's possible that the difficulties associated with patchwork regulation will prompt international companies to advocate for greater standardization. In addition, and as we have already seen from some Gen AI providers, they may limit the deployment or features of AI models in certain jurisdictions, leading to uneven access to AI technologies across geographic locations.

Additionally, as is often the case with specific regional legislation that has significant impact, the complexity of complying with SB 1047 is likely to generate increased demand for specialized legal services. Firms with capabilities to assist companies in navigating this complicated legal landscape, will be in higher demand, further driving the need for experts with both technical and legal expertise.

CONCLUSION

California's SB 1047 embodies a pivotal moment in the era of AI governance, standing squarely at the intersection of innovation and regulation, and encapsulates the complex challenges of governing emerging technologies with profound global impact.

In a conversation with a leading tech CEO on Sept. 17, 2024, Newsom expressed concerns about the bill, highlighting its potential to impede California's thriving AI industry, and he underscored the need for regulations that address current, demonstrable risks without stifling innovation. His remarks suggest uncertainty about the bill's future, though some skeptics believe his comments were tailored to the audience at the recent San Francisco tech conference and should be viewed with caution.

Whether or not the bill becomes law, it has already played a significant role in catalyzing broader engagement surrounding the regulation of AI, and has brought forth a plethora of stakeholders bringing compelling arguments on both sides of the issue.

Ultimately, the fate of the bill now rests in the hands of Governor Newsom, who will determine whether this bill merely contributes to ongoing discussions about future regulation, or becomes a groundbreaking law with international influence.

Reprinted with permission from the September 20, 2024 edition of The Recorder © 2024 ALM Media Properties, LLC. All rights reserved. Further duplication without permission is prohibited, contact 877-257-3382 or reprints@alm.com.

10 Min Read

Related Topics

Artificial Intelligence (AI)

Related Capabilities

Artificial Intelligence (AI)

Related Professionals

Bobby Malhotra

Carson Swope

© 2024 Winston & Strawn LLP.