August 5, 2024

TO: Members, Assembly Appropriations Committee

SUBJECT: SB 1047 (WIENER) SAFE AND SECURE INNOVATION FOR FRONTIER ARTIFICIAL INTELLIGENCE MODELS ACT
OPPOSE – AS AMENDED JULY 3, 2024
SCHEDULED FOR HEARING – AUGUST 7, 2024

The undersigned organizations must respectfully **OPPOSE SB 1047 (Wiener)** as amended July 3, 2024, which would enact the Safe and Secure Innovation for Frontier Artificial Intelligence Models Act to require a frontier AI developer, before initially training a covered model, to comply with certain obligations, implement full shutdown capabilities, and implement a specified safety and security protocol. The developer is further prohibited from using a covered model commercially or publicly, or making a covered model or a covered derivative model, available for commercial or public use, if there is an unreasonable risk of that the model or derivative can cause or enable a critical harm, as defined. While we share the goal of ensuring the safe and responsible development of AI, we believe that it is an issue that is appropriately being addressed at the federal level and are concerned that **SB 1047** will add more confusion to the already-fragmenting AI regulatory landscape in the U.S.

In addition to creating inconsistencies with federal regulations, the bill demands compliance with impractical, if not technically infeasible, requirements for which developers will be subject to exclusive and excessively harsh penalties, including potential criminal liability. We are concerned that the bill regulates AI technology as opposed to its high-risk applications, creates significant regulatory uncertainty, and therefore high compliance costs, and poses significant liability risks to developers for failing to foresee and block any harmful use of their models by others – all of which inevitably discourages economic and technological innovation. And while recent amendments vastly streamline the bill, add clarity to various terms and clarifying various obligations, including those requiring the implementation of full shutdown capabilities, on the one hand, and on the other, they expand what is considered a covered model or a derivative covered model take important steps in responding to the open-source community, we remain concerned about the impact of the bill on AI research and development in California and the impact on startups. While we are continuing to analyze the latest set of amendments on June 20, 2024, overall, the bill still makes AI business too risky in California, particularly given that the significant liability issues under SB 1047 were not addressed in any of the amendments to date.

This, unfortunately, does not better protect Californians. Instead, by hamstringing businesses from developing the very AI technologies that could protect them from dangerous models developed in territories beyond California's control, it risks only making them *more* vulnerable.

Again, we recognize that the June 20th amendments have made many changes, some of which impact the concerns below. Given the sheer volume of AI bills continuing through the Legislature, we unfortunately are still reviewing the impact of those changes to provide updated comments.

SB 1047 creates significant regulatory uncertainty by mandating compliance with novel requirements that rely on standards that are overbroad, vague, and impractical, if not infeasible

While **SB 1047** is often interpreted to simply require a risk assessment of models to avoid critical harms, doing so would dramatically misunderstand what the bill does in practice and miscalculate the impact it will have on Californians and the economy—even in the streamlined version. At its core, **SB 1047** regulates the development of AI, seeking to keep frontier AI developers from innovating AI models that will result in any kind of foreseeable harm—even harms that would not manifest from the model itself. In doing so, the bill requires developers to comply with incredibly vague, broad, impractical, if not impossible, standards when developing "covered models" and determining whether they can provide reasonable assurance that a covered model does not have a hazardous capability or come close to one, creating significant regulatory uncertainty.

Amendments adopted on June 5, 2024, have made a few positive, but limited, changes. These include updated definitions to key terms such as "hazardous capability", "covered model", among other things. As further outlined below, such changes, are unfortunately not sufficient to alleviate concerns in any meaningful way for most developers, particularly given that they will still face aggressive penalties for failing to make the determination that a covered model qualifies for a limited duty exemption.

Covered models:

As amended, **SB 1047** applies to AI models that: (1) meet a size threshold (was trained using a computing power greater than 10^26 integer or floating-point operations), and (2) the cost of that quantity of computing power would exceed $100 million, as specified. On its face, this definition is more concrete than the previous version of the bill, which instead scoped in models "that perform similarly" to that computing power. That said, the new $100 million threshold is not difficult to meet, particularly as the bill does not clearly state that it is the actual cost *to the developer* (in developing a model with such computing power) that is relevant, but rather the "cost of the quantity of computing power…***if*** calculated using average market prices…". (*See* Proposed Section 22602(f).)

In any case, by equating model size/cost to risk, the definition of "covered models" remains simultaneously overly broad and too narrow as smaller and/or less performant models can present much greater risks than large/higher performant ones. As a result, **SB 1047** both fails to adequately address the very real risks posed by small but malicious models and imposes significant costs on innovating performant but responsible ones.

Hazardous capabilities

The ambiguity around what is and is not a "covered model" aside, we are concerned that the regulatory regime envisioned by **SB 1047** sets unrealistic expectations that developers can provide reasonable assurance and certify that a model, prior to fine-tuning, does not have a "hazardous capability" and will not come close to one, even if someone removes all the protections that a developer adds to a model. (Proposed Section 22602(o).)

Recent amendments have narrowed the term "hazardous capability" in some respects, but only minimally so. For example, while the definition previously captured "other threats to public safety and security that are of comparable severity to" other listed harms, the amendments now define it to include "other *grave* threats…of comparable severity". Yet the bill still provides no additional clarity as to what is meant by "comparable severity"), leaving the definition of "hazardous capability" incredibly broad and vague.

Moreover, as amended, a covered model has a hazardous capability if it has the capability to be used to enable certain identifiable harms in a way that would be significantly more difficult to cause without access to "***a*** covered model that does not qualify for a limited duty exemption". This is problematic on two fronts: first, the amendment presumes that a developer will know whether each user of the model will have access to any other covered model that has not qualified for a limited duty exemption. Second, the identified harms include not only the creation or use of a chemical, biological, radiological, or nuclear weapon that results in mass casualties, or that cause $500 million of damage through cyberattacks of critical infrastructure, but also "other grave threats…that are of comparable severity" to those harms.

Even more problematically, a covered model is said to have a hazardous capability, "even if the hazardous capability *would not manifest* *but for* fine tuning and posttraining modification performed by third-party experts intending to demonstrate those abilities" – meaning, if third parties essentially jailbreak the model. (Proposed Section 22602(n)(2).) A developer cannot reasonably be held responsible for the future acts of others over which the developer has no control. In doing so, the bill is almost certain to undermine open-source development.

Limited duty exemption

As amended, a covered model qualifies for a "limited duty exemption" if a developer can "provide reasonable assurance" that the model does not have a "hazardous capability" and will not come close to possessing a hazardous capability when accounting for a reasonable margin of safety and the possibility of posttraining modifications[1]. Amendments further clarify that "reasonable assurance" does not require full certainty or practical certainty. Such changes certainly improve upon the previous requirement that a developer "reasonably *exclude the possibility* that a covered model has a hazardous capability or come close to one." That said, SB 1047 also still leaves it entirely ambiguous as to what is and is not considered sufficiently "close" to possessing a hazardous capability, or what would be considered within a "reasonable margin for safety," for purposes of determining if a model qualifies for a "limited duty exemption."

Furthermore, to make a determination as to whether a covered model qualifies for a limited duty exemption, **SB 1047** requires that a developer incorporate "all applicable covered guidance," without ever stating what those might be. However, industry and others are still trying to ascertain how to define what constitutes a highly-capable, foundational model and it is therefore unclear what will qualify as "industry best practices" for the purpose of incorporating all applicable covered guidance.

Even assuming that a developer could accurately ascertain what each of these standards require in order to make a determination as to whether a model qualifies for a limited duty exemption, **SB 1047** still makes it impossible for developers to actually determine if they can provide reasonable assurance that a covered model does not have hazardous capabilities and therefore qualifies for limited duty exemption because it requires developers to make the determination *before* they initiate training of the covered model. (*See* Proposed Section 22603). Because a developer needs to test the model by training it in a controlled environment to make determination that a model qualifies for the exemption, and yet cannot train a model until such a determination is made, **SB 1047** effectively places developers in a perpetual catch-22 and illogically prevents them from training frontier models altogether.

The unavoidable result of such issues is regulatory uncertainty that will only discourage economic and technological innovation. It would make far more sense to let the NIST (the National Institute of Standards and Technology) complete its work first, after which safety and security protocols tied to those safety standards could be better evaluated and considered.

SB 1047 focuses exclusively on developer liability, deters open-source development, imposes questionable requirements on operators of "computing clusters" and imputes harsh penalties

There are a host of other issues and unintended consequences remaining in the bill:

- ***SB 1047*** *fails to account for the AI value chain, impeding open source*. The bill almost exclusively focuses on developer liability, failing to account for the AI value chain. Under **SB 1047,** developers must build "full shutdown" capabilities into their models and may be held liable for downstream uses over which they have no control, impeding their ability to open-source their models. Ultimately, liability should rest with the user who intended to do harm, as opposed to automatically defaulting to the developer who could not foresee, let alone block, any and all conceivable uses of a model that might do harm. While recent amendments seemingly seek to narrow what is meant by "full shutdown" capabilities, the exclusions are unnecessarily difficult to interpret as drafted (full shutdown "does not mean the cessation of operation of a covered model to which access was granted pursuant to a license that was not created by the licensor…") and altogether insufficient.

[1] As opposed to the prior iteration of the bill where a limited duty exemption required that the developer be able to *reasonably exclude the possibility* that the covered model has a hazardous capability *or* may come close to one.

First**,** to the extent that the author's intent is to clarify that open-source models that have been distributed are no longer deemed to be in the control, custody, or possession of the developer, the bill should simply state exactly that (i.e., "Distributed open-source artificial intelligence models shall not be deemed to be in the custody, control, or possession of the developer of such open-source artificial intelligence model.")

Second, this does nothing to address the concerns with mandating a kill switch by operators of computing clusters, as described further below.

- ***SB 1047*** *sets unreasonable safety incident reporting requirements that deter open-source development*. Developers are required to report each AI safety incident "in the most expedient time possible and without unreasonable delay" and in no event later than 72 hours after learning of the incident, or learning facts that would lead to the reasonable belief that a safety incident occurred. Vagueness aside of such standards aside, the definition of "AI safety incident" covers a range of circumstances that are incompatible with open source because it would require monitoring of all downstream uses and applications.

- ***SB 1047*** *imposes intrusive, if not unreasonable, requirements on operators of "computing clusters"*. Under the bill, there are a host of requirements that apply to any company that "operates a computing cluster" – presumably, data centers or cloud computing companies that provide cloud compute for frontier model training. As drafted, however, it is unclear as to what the bill means by "operate", given that several entities could technically be seen operating a computer cluster: the owner of the cluster, the owner of the software operating the cluster, or the owner of the instance operating the cluster.

  Moreover, the bill not only forces operators of computing clusters to collect personally identifiable data from their prospective customers, but it expects them to predict if a prospective customer "intends to utilize the computing cluster to deploy a covered model," and requires that they implement a kill switch to enact a full shutdown the event of an emergency. The recent White House Executive Order on AI directs federal agencies to determine when and how frontier models may pose national security implications, including developing "know your customer" expectations and safety practices. **SB 1047** creates similar but different regulatory standards for these models. Absent alignment, there could be catastrophic implications for the technology industry in California and the US's leadership in cloud computing.

- ***SB 1047*** *establishes a new regulatory body with an ambiguous and ambitious purview*. The new "Frontier Model Division" within the Department of Technology would be responsible for a sweeping array of AI-related regulation, including developing novel safety tests and benchmarks, which could very well result in greater inconsistencies with federal rules. Conformity with national and international standards, such as NIST and ISO, should hold authority over those determined by the proposed Frontier Model Division. For example, best practices around red teaming and testing of these covered models are actively being determined by these organizations. Furthermore, additional details and assurances are needed regarding how information and disclosures provided to the Frontier Model Division would be transmitted and stored with the utmost security. Requiring developers and deployers to maintain documentation internally rather than California retaining sensitive, proprietary information on file, would be much more secure. Without clear, realistic requirements, and extraordinary protection of sensitive customer data and proprietary information, developers of frontier AI models are likely to move their training activities and other operations outside of California.

- ***SB 1047*** *imputes excessively harsh penalties, including potentially criminal liability and model deletion*. For instance, developers are required to submit certification to the new Frontier Model Division under penalty of perjury specifying the basis for their determination that a covered model qualifies for a limited duty exemption, yet the certainty required for that assessment is impracticable if not impossible to obtain. Potential civil penalties include model deletion (in the face of imminent risk or threat to public safety) and "an amount not exceeding 10 percent of the cost of the quantity of computing power used to train the covered model to be calculated using average market prices of cloud compute at the time of training for a first violation and in an amount not exceeding 30 percent of that value for any subsequent violation." Considering the significant resources to train covered models, this sum could amount to many millions.

Ultimately, certain problems demand federal solutions: SB 1047's inconsistencies will only further fracture the AI regulatory landscape and undermine federal efforts

We cannot overemphasize the importance of ensuring consistency in the AI regulatory landscape, nationally, and the need to follow federal guidance on certain issues that transcend national borders. Relevant to this bill, in October 2023, the White House issued an Executive Order (EO)[2] that requires companies that are developing any foundation model that poses a serious risk to national security, national economic security, or national public health and safety to notify the federal government when training the model and share the results of all red-team safety tests to ensure that AI systems are safe, security and trustworthy before companies make them public.

While we appreciate that in some respects, **SB 1047** appears in line with the goals of the federal government and the White House's EO, the National Institute of Standards and Technology (NIST) is already working with other agencies at the federal level to establish testing and safety guidelines for large models. If enacted, **SB 1047** would likely result in confusion about the correct standards to apply and place additional burdens on AI developers without commensurate gains in safety, especially as it fails to align with regulations nationally and introduces novel concepts and standards including around the assessment of what is a "hazardous capability". Indeed, given the definition of "covered models" under this bill which also scopes in any fine-tuning by downstream customers and users, **SB 1047** is more far-reaching than anything seen to date in voluntary commitments, federal guidance, or the laws of any other countries.

Ultimately, enacting a patchwork of inconsistent AI regulations that go into as much detail as **SB 1047**, will further fracture the U.S. regulatory landscape. As a result, instead of enhancing AI safety, this bill is bound to undermine sensible federal efforts that are already underway and hamper AI innovation in California unnecessarily, encouraging developers to move into other states. Again, this is a conversation that should be had and *is* being had at the national level and there is no need to replicate or duplicate those efforts, particularly in such an inconsistent manner. To the extent that a goal of **SB 1047** might be to set the prevailing standards and practices that the rest of the nation will follow, the lack of clarity and specificity in key definitions outlined above, will only discourage any widespread adoption.

Amendments to SB 1047 have generally failed to address our concerns

While we appreciate the demonstrated willingness of the author to amend this bill to address concerns, this bill has undergone no less than six iterations of amendments, none of which have managed to substantially address concerns raised to date.

For example, one of the more major sets of amendments shifted **SB 1047** away from requiring developers to make a "*positive safety determination*" with respect to a frontier model prior to initiating training of the model, in favor of setting rules for models that qualify for "limited duty exemptions". Such a change in terminology was largely a distinction without a difference: until a determination is made that the model qualifies for a "limited duty exemption", the developer must still comply with the exact same requirements as a developer who could not make a positive safety determination. These include, for example, implementing the capability to promptly enact a full shutdown of the covered model until the developer can make the necessary determination.

That said, there have been many changes made **to SB 1047** since its introduction that have been more substantive in nature. In large part, however, those changes equally failed to address our stated concerns – and in some cases, exacerbated them, as was the case with amendments that provided for punitive damages and those that expanded the already-sweeping array of AI related regulations required by the new Frontier Model Division. In other cases, the amendments added new concerns altogether. Among other things, those amendments included the following:

- The Attorney General is no longer required to commence a civil action when it has "reasonable cause to believe" that a violation has occurred. Instead, the Attorney General is given the discretion to do so, upon finding that a violation has occurred. At the same time, however, the bill was amended to expressly authorize punitive damages to be awarded, in addition to other monetary damages and the possibility of an order for the full shutdown of the model as well as other preventative relief that includes deletion of a model and the weights utilized in that model.

[2] FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence | The White House.

- Before initiating training of a frontier model that is not the subject of a limited exemption, and until the model becomes the subject of a limited duty exemption, a developer must now also ensure that their safety and security protocol describes in detail how their testing procedure address the possibility that a covered model can be used to make posttraining modifications or create another covered model in a manner that may generate hazardous capabilities.
- Developers now must provide a reasonable internal process through which an employee can anonymously disclose information to the developer if the employee believes in good faith that the information indicates the developer is out of compliance or has made false or materially misleading statements related to its safety and security protocol. This process includes, at minimum, monthly updates to the employee regarding the status of their disclosure and actions taken in response to the disclosure – presumably in perpetuity, even if the specific issue has been fully addressed, as the bill does not provide any guidance on when those mandated monthly updates can end.
- The new Frontier Model Division must issue guidance on or before July 1, 2026 regarding both the information relevant to determining whether an AI model is a covered model, and the technical thresholds and benchmarks relevant to determining if a covered model is subject to a limited duty exemption, taking into account the quantity of computing power used to train covered models that have been identified as having hazardous capabilities and "similar thresholds" used in federal law or regulation for the management of hazardous capabilities. Such guidance is to be updated at least every 24 months after initiate publication.

Unfortunately, none of these changes mitigate concerns we have raised.

Again, we applaud the intent of this bill but are concerned that its execution will have counterproductive impacts, not only chilling AI innovation, but also preventing AI's beneficial uses. During an incredibly challenging budget year, this bill will result in significant costs to the State in the realm of tens of millions of dollars. In addition to the cost of standing up the new Frontier Model Division and CalCompute, there is also the bigger picture of the incredible potential for future tax revenue that the AI ecosystem can bring to California – meaning, not simply from AI companies, but also from all the industries and businesses looking to leverage AI to increase their efficiency and profitability. Enacting legislation that regulates the development of technology itself, instead of the implementation and uses of it, will be seen as creating a hostile environment for innovation and drive investment to other tech hubs both inside and outside the U.S., with far reaching implications for state revenues. As such, we must unfortunately **OPPOSE SB 1047 (Wiener)**.

Sincerely,

Ronak Daylami
Policy Advocate
on behalf of

Association of National Advertisers (ANA), Christopher Oswald
California Chamber of Commerce, Ronak Daylami
California Land Title Association, Anthony Helton
California Manufacturers and Technology Association (CMTA), Robert Spiegel
Civil Justice Association of California (CJAC), Jaime R. Huff
Computer and Communications Industry Association (CCIA), Naomi Padron
Insights Association, Howard Fienberg
Silicon Valley Leadership Group, Peter Leroe-Muñoz
Software and Information Industry Association (SIIA), Anton van Seventer
TechNet, Dylan Hoffman

cc: Legislative Affairs, Office of the Governor
Severiano Christian, Office of Senator Wiener
Consultant, Assembly Appropriations Committee
Liz Enea, Consultant, Assembly Republican Caucus

RD:ldl