March 26, 2024

TO: Members, Senate Judiciary Committee

SUBJECT: SB 1047 (WIENER) SAFE AND SECURE INNOVATION FOR FRONTIER ARTIFICIAL INTELLIGENCE SYSTEMS ACT
OPPOSE – AS AMENDED MARCH 20, 2024

The undersigned organizations must respectfully **OPPOSE SB 1047 (Wiener)** as amended March 20, 2024, which would enact the Safe and Secure Innovation for Frontier Artificial Intelligence Models Act to require frontier AI developers to make a positive safety determination before initiating training of a covered model, among other things. While we share your 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 various vague and impractical, if not technically infeasible, requirements for which developers will be subject to 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. 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.

SB 1047 mandates compliance with novel requirements based on standards that are often overbroad, vague, and impractical, if not simply infeasible

At its core, **SB 1047** seeks to regulate 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".

For example, **SB 1047** applies to AI models that either: (1) meet a size threshold (a computing power greater than 10^26 integer or floating-point operations in 2024), or (2) that perform similarly. What the latter category of covered models looks like, however, is not entirely clear. (*See* Proposed Section 22602(f).) The bill merely states that they are models "trained using a quantity of computing power *sufficiently large that it could be reasonably expected to have similar or greater performance* as an AI model trained using a quantity of computing power greater than 10^26 integer or floating-point operations in 2024 as *assessed using benchmarks commonly used to quantify the general performance of state-of-the-art foundation models*". There is little to no certainty as to what this translates to in practice and, in any case, such thresholds will become obsolete within a year, requiring the law to change yet again. Moreover, by equating model size to risk, the definition of "covered models" is 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, the bill 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.

The ambiguity around what is and is not a "covered model" aside, we are concerned that the regulatory regime envisioned in the bill sets unrealistic expectations that developers can certify that a model, prior to fine-tuning, is safe from any *possibility* that it could be hazardous, even if someone removes all the protections that a developer adds to a model. Specifically, to make a "positive safety determination" under this bill, a developer must be able to "reasonably exclude the possibility that a covered model has a *hazardous capability* or *may come close* to possessing a hazardous capability when accounting for a *reasonable margin for safety* and the possibility of posttraining modifications." (Proposed Section 22602(s).)

First, **SB 1047's** definition of "hazardous capability" is so broad that it in fact captures not only covered models that have the capability to be used to enable certain harms (e.g. the creation or use of a chemical, biological, radiological, or nuclear weapon, or other threats of "comparable severity") in a way that would be significantly more difficult to cause without access to the covered models, but also those that have such capabilities, "even if the hazardous capability *would not manifest <u>but for</u>* 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).) The overbreadth of the definition aside, what is considered sufficiently *close* to possessing a hazardous capability to prevent the finding of a positive safety determination, however, is even more unclear. Also unclear is what would be considered a "reasonable margin for safety", or an unreasonable one.

Second, assuming a developer could accurately ascertain what is and is not considered a "hazardous capability", **SB 1047** still makes it impossible for developers to actually make any positive safety determinations ruling out those hazardous capabilities by requiring developers to make the positive safety determination *<u>before</u>* they initiate training of the covered model. (Proposed Section 22603). Because a developer needs to test the model by training it in a controlled environment to make a positive safety determination 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.

Finally, **SB 1047** also requires that a developer "incorporate all covered guidance" before making a positive safety determination. However, industry and others are still trying to ascertain how to define what constitutes a high-capable, foundational model and it is unclear what will qualify as "industry best practices" or "standards setting organizations" for the purpose of incorporating all covered guidance. Such regulatory uncertainty will inevitably discourage economic and technological innovation. It would make far more sense to let the U.S. AI Safety Institute to complete its work first, after which safety and security protocols tied to those safety standards could be considered.

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

There are a host of other issues and unintended consequences that warrant further consideration:

- ***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.
- ***SB 1047*** *sets unreasonable safety incident reporting requirements that are not only vague but deter open-source development*. Developers are required to report each AI safety incident upon learning of it, or learning facts that would lead to the reasonable belief that a safety incident occurred. However, what is considered an "AI safety incident" is vague. Among other things, it includes a covered model "autonomously engaging in a sustained sequence of unsafe behavior other than at the request of a user" but fails to define what is considered "unsafe", leaving developers to guess if they must report an incident. At the same time, "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.

- ***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.
- ***SB 1047*** *imputes excessively harsh penalties, including potentially criminal liability and model deletion*. For instance, developers are required to submit certification of positive safety determinations to the new Frontier Model Division under penalty of perjury, 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, excluding labor cost, to develop the covered model for a first violation and in an amount not exceeding 30 percent of the cost, excluding labor cost, to develop the covered model 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)[1] 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.
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 as much as its harmful ones. As such, for all the aforementioned reasons, we must **OPPOSE SB 1047 (Wiener)**.

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

Sincerely,

Ronak Daylami
Policy Advocate
on behalf of

Association of National Advertisers (ANA), Christopher Oswald
California Chamber of Commerce, Ronak Daylami
California Manufactures 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
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, Senate Judiciary Committee
Morgan Branch, Senate Republican Caucus

RD:ldl