March 26, 2024

The Honorable Tom Umberg
Chair, Senate Judiciary Committee
1021 O Street, Room 3240
Sacramento, CA 95814

Re: S.B. 1047 – Support if amended (As amended 3/26/24)

Dear Senator Umberg:

I write today on behalf of the Electronic Frontier Foundation, a San Francisco-based, non-profit organization that works to protect civil liberties in the digital age. EFF represents more than 30,000 active donors and members, including thousands of supporters in California. EFF would like to express a "support if amended" position on S.B. 1047, authored by Sen. Scott Wiener, which would establish a public cloud computing cluster, CalCompute, to promote competition and fairness, but also sets an abstract and confusing set of regulations for those developing AI systems.

Advances in artificial intelligence sophistication have raised concerns about whether developers are considering the potential harms of these powerful models, and whether they are creating them in an ethical and transparent way. This bill recognizes the potential power of such systems and aims to prevent catastrophic outcomes from the development of generative AI along with goals to address competition in the AI space and promote developer responsibility.

We do not believe the current bill can accomplish all of these goals, which are broad, unwieldy, and difficult to address in a single measure. The lack of clarity from the bill's definitions and structure could potentially create an ecosystem where very few people understand how to comply—paving the way for only those corporations with the most money to be able to function in this ecosystem. And focusing on long-term, catastrophic outcomes from artificial intelligence unfortunately pulls attention away from AI-enabled harms that are directly before us.

Several critical pieces of this bill are unclear or problematic. For example, the bill defines an "artificial intelligence safety incident" to include "A covered model autonomously engaging in a sustained sequence of unsafe behavior other than at the request of a user." Neither "sustained sequence" nor "unsafe behavior" is defined in the bill, however. One could also wonder about the exclusion of "unsafe behavior … at the request of a user."

Similarly, the bill directs developers to assert that their models cannot be used for "hazardous capacity." The bill lists a few scenarios that would qualify, and then adds, "Other threats to public safety and security that are of comparable severity to the harms described in paragraphs (A) to (C), inclusive." This leads to obvious questions about the definition that go largely unanswered: Who determines what threats are of comparable safety? By what metric? With whose input?

Asking developers to assess the risks and potential harms their systems can create is good practice and should be standard. However, creating criminal penalties for developers who find their "positive safety determinations" were wrong sets up a harsh penalty for failing to predict the future correctly, particularly when the goals are so hazy. While the bill has added a 30-day grace period for good faith errors someone may make in this process, those problems may not present themselves that quickly. Developers can take steps to ensure their models are safe, but they cannot control the actions of the people who use their models once they are out in the world.

Leading artificial intelligence scholars have noted that AI models are a tool that, like any tool, can be used for good or bad purposes.[1] Arvind Narayanan and Sayash Kapoor illustrate the difficulty of writing a "safe" model in the example below:

> Consider the concern that LLMs can help hackers generate and send phishing emails to a large number of potential victims. It's true — in our own small-scale tests, we've found that LLMs can generate persuasive phishing emails tailored to a particular individual based on publicly available information about them.
>
> But here's the problem: phishing emails are just regular emails! There is nothing intrinsically malicious about them. A phishing email might tell the recipient that there is an urgent deadline for a project they are working on, and that they need to click on a link or open an attachment to complete some action. What is malicious is the content of the webpage or the attachment. *But the model that's being asked to generate the phishing email is not given access to the content that is potentially malicious*. So the only way to make a model refuse to generate phishing emails is to make it refuse to generate emails. That would affect many non-malicious uses, such as marketing.

That is simply one example of how imprecise language in this bill, while well-intentioned, could do more harm than good. In addition to the overall ambiguity of the bill, we also have concerns about the privacy impacts of certain provisions, such as the "Know Your Customer" requirements for computing cluster operators (Section 22064). This section requires operators to collect personal information and "maintain for seven years and provide to the Frontier Model Division or the Attorney General, upon request, appropriate records of actions taken under this section, including policies and procedures put into effect." It also requires these operators to have a mechanism to shut down a model. The lack of clarity around the ways that the Frontier Model Division and the California Attorney General, as well as other law enforcement agencies, can request or demand this information or the shutdown of a model, opens the door to troublesome government control over models that may be used for perfectly legal or legitimate purposes.

EFF does, however, applaud with the sponsors' goals to promote competition and fairness, and does not oppose the creation of a public cloud computing cluster to act as a counterbalance to private enterprise.

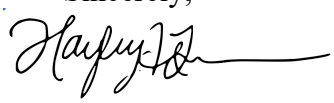
---

[1] Arvind Narayanan and Sayash Kapoor. *AI safety is not a model property,* https://www.aisnakeoil.com/p/ai-safety-is-not-a-model-property

A public alternative, created thoughtfully, could both allow California to set high standards for policies that reduce harm, protect privacy, and promote innovation, while also addressing the very present threat of an AI market where only large companies can afford to participate. To accomplish this, California must be willing to fund this project fully to create a public computing cluster that does not find itself beholden to the companies that it is trying to compete with, through contracts or other secretive agreements. Overall, we encourage the author and sponsors to continue talking to affected stakeholders and to consider building mechanisms into the bill itself to allow for ongoing input to guard against industry capture.

For the above reasons, EFF cannot support the bill as written but respectfully suggests it be amended to focus on the creation of CalCompute. This will allow all stakeholders to focus their attention on the lofty goal of building a public cloud computing cluster that follows the principles of transparency, competition, privacy, and access that are clearly important to the author and sponsors. We welcome further conversation and thank the sponsors for their open and ongoing discissions about the bill. Thank you.

Sincerely,



Hayley Tsukayama
Associate Director of Legislative Activism
Electronic Frontier Foundation
(415) 436-9333 x 161

cc: Honorable Members and Committee Staff, Senate Judiciary Committee