Meta

June 25, 2024

The Honorable Scott Wiener
California State Senate
1021 O Street
Suite 8620
Sacramento, CA 95814-4900

Dear Senator Wiener,

We are writing to express significant concerns about SB 1047, the “Safe and Secure Innovation for Frontier Artificial Intelligence Models Act.” Meta's mission is to build technologies that help people connect, find communities, and grow businesses. AI is key to advancing this mission and has been since our earliest days.

This bill sets out to advance an objective that Meta shares: the safe and responsible development of the most capable AI models, which the bill refers to as "the frontier." We have followed the bill's progress closely, and we appreciate your openness to stakeholder feedback, which we have provided at multiple points, as well as your good faith efforts to acknowledge and address concerns raised. We have also evaluated the concerns other stakeholders have raised, some of which we share.

Despite these conversations, we remain concerned that this bill fundamentally misunderstands how advanced AI systems are built and therefore would deter AI innovation in California at a time when we should be promoting it. The bill focuses solely on large foundation model developers, rather than addressing actions of the deployers who translate those models into the systems with which people actually interact (and therefore have the most control over how AI systems can be used), and the users themselves. The bill imposes liability on model developers for downstream harms regardless of whether those harms relate to how the model was used rather than how it was built, effectively making them liable for scenarios they are not best positioned to prevent.

The AI ecosystem is diverse, and this approach simply does not reflect that there are many different ways that generative AI systems are built and many different actors involved in their development. There is almost never one entity or person that builds an entire AI system.

Instead, these systems are built by multiple actors with differing functions: the model developer who puts a model onto the market; AI deployers who build models into systems to provide a service and control use of the model; and end-users who leverage those systems.

The bill's fundamental flaw is that it fails to take this full ecosystem into account and assign liability accordingly, placing disproportionate obligations on model developers for parts of the ecosystem over which they have no control. This is not to suggest that regulation should exempt developers altogether. The problems lie in the types of obligations imposed on developers, which overwhelmingly center on certifying specific outcomes (an unrealistic task) rather than requiring governance and accountability measures.

As a result, the bill will make the AI ecosystem less safe, jeopardize open-source models relied on by startups and small businesses, rely on standards that do not exist, and introduce regulatory fragmentation—compromising the leading roles of both California and the United States in AI innovation.

SB 1047 will make model developers less likely to open source advanced models, making the AI ecosystem less safe and hurting startups and small businesses

We appreciate your effort to modify SB 1047 to avoid foreclosing open-source development. But the amended bill is still unworkable in an open-source context because it would require the provider of an open source model to make risk certifications that anticipate every possible use or alteration to their model — something that is plainly impossible to do.

Even as it imposes unrealistic obligations on open source providers, the bill also actively discourages the release of open source AI because providers would have no way to open-source models without facing untenable legal liability.

This will narrow the market to just a handful of proprietary models—lessening choice for developers and consumers.

By disincentivizing open-source, SB 1047 makes AI less safe

Historical evidence from cybersecurity shows that open-source approaches to technology development provide heightened security benefits because they allow for community feedback and independent evaluation. Indeed, many technologies underlying today's Internet are considered secure because they are open source.

For AI, this external oversight improves both specific AI models and furthers the ecosystem's understanding of large model security and safety. Relying on the safety of a closed model will mean trusting the developer. In contrast, open models allow for evaluation by third parties, including independent researchers, for safety, security, fairness, helpfulness, and other

capabilities without an inherent power imbalance and information asymmetry between the model developer and the tester.

SB 1047 will hurt startups and small businesses

Startups and small businesses are often the most innovative and agile in their adoption of new technologies, and AI is no exception. By limiting open-source AI, SB 1047 discourages a key avenue for small businesses and startups to compete and participate in the innovation ecosystem—the very environment California has led the world in fostering. SB 1047 lessens their ability to use free, readily available and fine-tuned models to create new jobs, businesses, tools, and services that often go on to be used by other businesses, governments, and civil society groups.

Making AI models available openly gives small businesses, startups, entrepreneurs, and researchers access to tools developed at a scale they might not otherwise access. This opens up a world of opportunities for them to experiment, innovate in exciting ways, and ultimately benefit from economically and socially. To that end, we have seen a groundswell of enthusiasm for open source AI from small businesses. They recognize the transformative potential of AI for competition, innovation, and productivity, which is already contributing to U.S. economic growth. Having access to state-of-the-art AI creates opportunities for everyone, not just a small handful of big companies.

Already we've seen how open-source AI drives innovation through developers using our Llama open-source large language models. Developers are using Llama to generate data-secure training data for clinical tools, making it faster and easier for doctors to treat cancer patients. This kind of critical health innovation is happening in California today. Others are using Llama to help young mothers answer questions about early childhood care and education and to build AI systems to help farmers identify and mitigate invasive insects.

Disincentivizing an open-source ecosystem disproportionately hurts these developers, stifling innovation and limiting the potential of AI to drive economic growth and job creation. Startups and small businesses are critical to the California economy, and SB 1047 would undermine their ability to compete and succeed in the AI marketplace.

SB 1047 relies on nonexistent standards

The bill establishes a series of requirements developers must meet prior to training and deploying covered models. These requirements are not just risk assessments and testing already being implemented; instead they require developers to "identif[y] specific tests and test results that would be sufficient to provide reasonable assurance" that a covered model or derivatives do not pose unreasonable risk of causing a critical harm. But a standard for evaluating whether these tests and test results are sufficient simply does not exist.

As safety experts recognize, there are no existing benchmarks for the types of AI "safety testing" that the bill contemplates. Industry and governments are working together to determine which benchmarks and frameworks for testing and evaluating models can and should be used. That's why the US AI Safety Institute's recently published strategic plan emphasizes that there is currently:

> (1) A lack of commonly accepted definitions for AI safety, as well as AI safety capabilities and measurements of those capabilities, especially for frontier models and advanced AI agents and systems.
> (2) Underdeveloped testing, evaluation, validation, and verification (TEVV) methods and best practices to provide holistic assessments of risk – from model capabilities to human-AI interaction to system level and societal-level impacts.
> (3) An absence of scientifically-established risk mitigations across the lifecycle of AI design and deployment.

Rather than creating a mandate to meet standards that don't exist, California should foster the development of those standards through legislation creating incentives and safe harbors to advance this work.

SB 1047 is out of step with leading global efforts

Finally, we are concerned that SB 1047 pushes AI governance in a direction that is at odds with the global consensus on the issue. Specifically, although details vary, the global trend in AI frameworks has coalesced around requiring accountability mechanisms, such as transparency or simulated adversary testing (or "red-teaming"). SB 1047 takes a completely different and impracticable approach, requiring model developers to certify that their models will not lead to specific outcomes, regardless of actions by a third-party model deployer or user.

This approach would not only be impractical (because risk evaluation is, by definition, probabilistic) but also move California in the opposite direction of global approaches like:

- The recent commitments by US frontier AI companies, including Meta and others, to the Seoul Frontier AI Safety Commitments, which clearly set out steps that frontier model providers will take regarding safety assessments, including publishing safety frameworks;
- The Biden Administration's Executive Order, which focuses governance of "dual-use" foundation models on accountability and transparency practices (e.g., placing obligations on foundation model developers to disclose planned model training, cybersecurity protections, and results of red-teaming);
- The White House commitments for frontier AI models, for which Meta was one of the first signatories, which acknowledge that different AI actors have different responsibilities depending on their role in the development lifecycle and require

signatories to take appropriate steps to innovate responsibly—not certify they will dispositively prevent certain harms;
- Colorado's newly adopted regulation of high-risk use cases, which assigns governance requirements to model developers and deployers depending on their position in the AI value chain; and
- The EU AI Act's approach to General Purpose AI Models with "systematic risk" (models with high impact capabilities, e.g., models over 10^25 FLOPS), which anchors on accountability-based requirements, such as transparency and red-teaming – not a blanket assurance that, even if significantly modified, the model could never be used to cause downstream harms.

California's AI policy must be informed by, and work to drive alignment with, these global efforts rather than out of step with them.

In light of the concerns outlined here and by numerous other organizations, we believe this bill is not ready to move forward. We are committed to continuing to work with you and other lawmakers to find the right approach for California.

Sincerely,

Rob Sherman
VP, Deputy Chief Privacy Officer, Policy

cc:
The Honorable Ash Kalra, Chair of the California Assembly Committee on Judiciary
The Honorable Rebecca Bauer-Kahan, Chair of the California Assembly Committee on Privacy and Consumer Protection
The Honorable Buffy Wicks, Chair of the California Assembly Committee on Appropriations
Christine Aurre, Secretary of Legislative Affairs for the Honorable Governor Gavin Newsom
Liz Enea, Consultant, Assembly Republican Caucus