Congress of the United States
Washington, DC 20515

August 15, 2024

The Honorable Gavin Newsom
Governor of the State of California
1021 O Street, Suite 9000
Sacramento, CA 95814

Dear Governor Newsom,

It is somewhat unusual for us, as sitting Members of Congress, to provide views on state legislation. However, we have serious concerns about SB 1047, *the Safe and Secure Innovation for Frontier Artificial Intelligence Models Act,* and we felt compelled to make those concerns known to California state policymakers. Our views on this legislation have been formed from our respective experience serving on the relevant House Committees that have oversight of Artificial Intelligence (AI), as well as the House AI Taskforce. We have extensive experience with AI, including helping to write one of the first Federal AI laws: *The National Artificial Intelligence Initiative Act of 2020*. This legislation focused heavily on the sociotechnical risks associated with AI systems, including safety risks, and the activities that resulted from it have helped underpin our current understanding of these risks. Based our experience, we are concerned that SB 1047 creates unnecessary risks for California's economy with very little public safety benefit, and because of this, **if the bill were to pass the State Assembly in its current form, we would support you vetoing the measure**.

The methodologies for understanding and mitigating safety and security risks related to AI technologies are still in their infancy. The current state of the technical solutions that would underpin implementation of SB 1047, including standards, benchmarks, and evaluations, is significantly underdeveloped. The bill requires firms to adhere to voluntary guidance issued by industry and the National Institute of Standards and Technology, which does not yet exist. For example, even though we do not yet have the standardized evaluations necessary for a developer to confirm with confidence that an AI system could cause a "critical harm," the bill bases its liability provisions upon such hypothetical guidance. The current industry best practice for detecting safety risks in an AI system is called "red teaming." Unfortunately, as of the publishing of this letter, there exists no standard guidance for conducting red teaming. Furthermore, by itself

this form of evaluation is inadequate to discover the types of risks that SB 1047 contemplates because developers must know all the risks and variables in advance of testing.[1] Similarly, the bill hinges liability on arbitrary compute thresholds that will almost certainly be obsolete by the time the law would go into effect. Such premature requirements based on underdeveloped science call into question from the outset the efficacy of the bill in achieving its goals of protecting public safety. We should not seek to cement our current understanding of AI safety science into law but should instead provide amble flexibility, agility, and public consultation to allow the law to grow as our understanding grows. In its current form, SB 1047 falls short.

To justify the severe burdens this bill proposes on AI developers and deployers, SB 1047 should ensure its restrictions are proportionate to real-world risks and harms. Unfortunately, SB 1047 is skewed toward addressing extreme misuse scenarios and hypothetical existential risks while largely ignoring demonstrable AI risks like misinformation, discrimination, nonconsensual deepfakes, environmental impacts, and workforce displacement. There is little scientific evidence of harm of "mass casualties or harmful weapons created" from advanced models.[2] The proponents of SB 1047 have cited the risk of AI models creating chemical, biological, radiological, or nuclear weapons (CBRN). Their focus on CBRN threats illustrates the problems with SB 1047. Take, for instance, the threat of AI models creating nuclear weapons. Obviously, AI models cannot create the industrial infrastructure necessary to produce the fissile material for a nuclear weapon. The production or acquisition of this fissile material is the primary impediment to the creation of such a weapon, and that is the reason we have strict controls on national and international trafficking of fissile materials and enrichment equipment. The technical design of nuclear weapons has been known since at least 1945 and can almost certainly be acquired on dark corners of the internet. Similarly, synthetic biology is an area where AI could democratize harmful information,[3] but the clearest solution is to harden protections against misuse where biological agents are synthesized[4]—rather than the software used to simulate them.

Regulations aimed at CBRN risks should focus on restricting the physical tools needed to create these physical threats, rather than restricting access to knowledge that is already available through non-AI means. Fortunately, we already have robust controls in place for many of these items, federal agencies are contemplating additional rules and restrictions to combat issues created by AI, and Congress is currently debating additional controls in emerging fields like gene editing.

---

[1] Sorelle Friedler et al., "AI Red-Teaming Is Not a One-Stop Solution to AI Harms: Recommendations for Using Red-Teaming for AI Accountability," Data and Society, October 25, 2023, https://datasociety.net/library/ai-red-teaming-is-not-a-one-stop-solution-to-ai-harms-recommendations-for-using-red-teaming-for-ai-accountability/.

[2] Sayash Kapoor et al., "On the Societal Impact of Open Foundation Models," Stanford, accessed on arXiv, February 27, 2024, https://arxiv.org/pdf/2403.07918.

[3] Christopher Mouton, Caleb Lucas, and Ella Guest, "The Operational Risks of AI in Large-Scale Biological Attacks," RAND, October 16, 2023, https://www.rand.org/pubs/research_reports/RRA2977-1.html.

[4] Forrest Crawford et al., "Securing Commercial Nucleic Acid Synthesis," RAND, 2024, https://www.rand.org/content/dam/rand/pubs/research_reports/RRA3300/RRA3329-1/RAND_RRA3329-1.pdf.

To be clear, we firmly support AI governance to guard against demonstrable risks to public safety. There is more we can do to advance safety science and develop the evidence that would underpin future action.[5] Understanding, measuring, and monitoring the risks inherent in AI systems as they evolve will be key—especially the marginal risk of AI systems causing mass casualty events as contemplated by SB 1047.[6] There is also a pressing need to develop the scientific tools, tests, and standards to enable to effective evaluation and assurance of AI systems. We are currently working with my colleagues across the aisle and in the Senate to support federal activities to detect CBRN issues in AI systems and mature the state of evaluations for AI systems. More work also needs to be done to develop the AI governance workforce and reduce fragmentation in the AI community on the nature of AI risks and regulatory response.[7] Further, we support some provisions of SB 1047, like CalCompute research and the whistleblower protection provisions.

We also do not oppose action by the California legislature to address demonstrated harms by AI technologies. There are more than 30 AI-related bills still under consideration in the California State House and Senate that are targeted at real world harms with large societal impact. Several bills each look at various harms that arise from the proliferation of synthetic content—from nonconsensual pornography to synthetic images in election advertisements. For example, AB 2355, offered by Assemblymember Carrillo, would require disclosure of AI used in election advertisements.[8] Similarly, AB 1856, offered by Assemblymember Ta, would make it a crime for a person who is 18 years or older to intentionally distribute synthetic images of an intimate body part of an individual.[9] While we have not thoroughly reviewed the underlying legislation and cannot yet endorse the individual proposals, the evidence of harm created by synthetic content in these contexts is not speculative.[10,11,12] As such, these bills have a firmer evidentiary basis than SB 1047.

---

[5] Arvind Narayana and Sayash Kapoor, "AI existential risk probabilities are too unreliable to inform policy," AI Snake Oil, July 26, 2024, https://www.aisnakeoil.com/p/ai-existential-risk-probabilities.

[6] Sayash Kapoor et al., "On the Societal Impact of Open Foundation Models."

[7] "Understanding AI Safety," Open Letter on SB-1047, accessed August 13, 2024, https://docs.google.com/forms/d/e/1FAIpQLScKAHaHKQl8g1RADNwLTSzx_GgAfg88hPxFwjN1gpZdIxb19w/viewform.

[8] AB 2355, Political Reform Act of 1974: political advertisements: artificial intelligence (2024). https://legiscan.com/CA/text/AB2355/id/2925425.

[9] AB 1856, Disorderly conduct: distribution of intimate images (2024). https://legiscan.com/CA/text/AB1856/id/2999125.

[10] Natasha Singer, "Teen Girls Confront an Epidemic of Deepfake Nudes in Schools," New York Times, April 8, 2024, https://www.nytimes.com/2024/04/08/technology/deepfake-ai-nudes-westfield-high-school.html.

[11] Heather Chen and Kathleen Magramo, "Finance worker pays out $25 million after video call with deepfake 'chief financial officer'," CNN, February 4, 2024, https://www.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk/index.html.

[12] Alex Seitz-Wald, "Political consultant who admitted deepfaking Biden's voice is indicted, fined $6 million," NBC News, May 22, 2024, https://www.nbcnews.com/politics/politics-news/steve-kramer-admitted-deepfaking-bidens-voice-new-hampshire-primary-rcna153626.

In addition to its misplaced emphasis on hypothetical risks, we are also concerned that SB 1047 could have unintended consequences from its treatment of open-source models. Currently, some advanced models are released as open source and made widely available. This openness allows smaller, lesser-resourced companies and organizations, including universities, to develop on top of them, stimulating innovation and having large economic impact. This bill would reduce this practice by holding the original developer of a model liable for a party misusing their technology downstream. Not only is it unreasonable to expect developers to completely control what end users do with their products, but it is difficult if not impossible to certify certain outcomes without undermining the rights of end users, including their privacy rights. As such, the natural response from developers will be to stop releasing fully open AI models and instead implement limited release models, like APIs.[13] Further, the bill would require developers to add kill switches to their AI systems, which would allow them to turn off the AI system at any time. Proponents of SB 1047 say this requirement is only placed on an AI model if it is in the developer's possession, but by applying liability to the downstream use of model derivatives, the bill would effectively force developers to maintain possession. These types of polices that force closed development, while they may sound good in theory, would decimate the ecosystems that spring up around AI models.[14] As the renowned Stanford AI researcher Fei-Fei Li wrote, "If developers are concerned that the programs they download and build on will be deleted, they will be much more hesitant to write code and collaborate."[15]

It may be the case that the risks posed by open sourcing models with potentially dangerous capabilities justify this precaution. But current evidence suggests otherwise. After seeking comment from the community and looking at the risks, the National Telecommunications and Information Administration released a report last month saying government should not restrict access to open-source models with widely available model weights at this time, but instead should actively monitor the ecosystem should risks evolve.[16] Further, the openness and transparency that come from open source and open science can lead to better AI governance models.[17] Given that most of the discoveries that led us to this moment were achieved through

---

[13] Irene Solaiman, "The Gradient of Generative AI Release: Methods and Considerations," Computers and Society, accessed on Arxiv, February 5, 2023, https://arxiv.org/abs/2302.04844.

[14] Rishi Bommasani et al., "Considerations for Governing Open Foundation Models," Stanford, December 2023, https://hai.stanford.edu/sites/default/files/2023-12/Governing-Open-Foundation-Models.pdf;

[15] Fei-Fei Li, "'The Godmother of AI' says California's well-intended AI bill will harm the U.S. ecosystem," Fortune, August 6, 2024, https://fortune.com/2024/08/06/godmother-of-ai-says-californias-ai-bill-will-harm-us-ecosystem-tech-politics.

[16] "Dual-Use Foundation Models with Widely Available Model Weights," National Telecommunications and Information Administration, July 2024, https://www.ntia.gov/sites/default/files/publications/ntia-ai-open-model-report.pdf.

[17] Kavin Bankston, Jennifer Hodges, et al., "Openness and Transparency in AI Provide Significant Benefits for Society," Letter to Department of Commerce, Center for Democracy and Technology, Mozilla, and 23 organizations, March 25, 2024, https://cdt.org/wp-content/uploads/2024/03/Civil-Society-Letter-on-Openness-for-NTIA-Process-March-25-2024.pdf.

openness[18], SB 1047 could have a pernicious impact on U.S. competitiveness and governance in AI, especially in California.

In short, we are very concerned about the effect this legislation could have on the innovation economy of California without any clear benefit for the public or a sound evidentiary basis. High tech innovation is the economic engine that drives California's prosperity. This is particularly the case in the Bay Area. There is a real risk that companies will decide to incorporate in other jurisdictions or simply not release models in California. This is not entirely speculative. As an example, Meta recently decided not to release advanced multimodal AI systems in Europe due to their precautionary rules.[19]

While we are confident of the good intentions of the bill's proponents, we are equally confident that this bill would not be good for our state, for the start-up community, for scientific development, or even for protection against possible harm associated with AI development. Because of those reasons, if the bill were to pass the legislature in its current form, we would support you vetoing the measure.

If you have any questions, please contact Alan McQuinn of the House Science Committee's Minority staff at (202) 225-6375.

Sincerely,

Zoe Lofgren
Member of Congress

Anna Eshoo
Member of Congress

Ro Khanna
Member of Congress

Scott Peters
Member of Congress

[18] For example, see Jakob Uszkoreit, "Transformer: A Novel Neural Network Architecture for Language Understanding," Google Research, 2017, https://research.google/blog/transformer-a-novel-neural-network-architecture-for-language-understanding/.

[19] Jess Eatherbed, "Meta won't release its multimodal Llama AI model in the EU," *The Verge*, July 18, 2024, https://www.theverge.com/2024/7/18/24201041/meta-multimodal-llama-ai-model-launch-eu-regulations.

Tony Cárdenas

Member of Congress

Ami Bera

Member of Congress

Nanette Barragán

Member of Congress

J. Luis Correa

Member of Congress