[US] Californian AI safety bill moving forward despite challenges

IRIS 2024-7:1/3

Amélie Lacourt
European Audiovisual Observatory

In California, the Safe and Secure Innovation for Frontier Artificial Intelligence Models Act (AI bill), introduced in February 2024, is causing quite a stir among Silicon Valley tech giants, including Meta and Alphabet.

The bill is designed to reduce the risks posed by AI and impose safety regulations on artificial intelligence companies. In particular, it requires these companies to test their systems and add safety measures to prevent them from being potentially manipulated to wipe out the state's electric grid or help build chemical weapons.

The bill only applies to advanced or "frontier" AI models, which are systems that have cost over $100 million to train, an amount which has so far not been reached. Against criticism, Democrat State Senator Scott Wiener, who wrote the bill, emphasised that it is not about smaller AI models but that it is about "incredibly large and powerful models that, as far as we know, do not exist today but will exist in the near future."

The Bill covers some of the following points:

- The capability for an AI model to promptly enact a full shutdown, also referred to as a "kill switch". Full shutdown should be understood as the cessation of operation of either (1) the training of a covered model, (2) a covered model, or (3) all covered model derivatives controlled by a developer.

- The implementation of a written and separate safety and security protocol.

- The creation of the Frontier Model Division within the Government Operations Agency, which developers of a covered model would have to submit a certification of compliance with the bill's provisions, under penalty of perjury. Developers would also have to report each artificial intelligence safety incident affecting the covered model or any covered model derivative controlled by the developer to the new Division.

- The creation of the Board of Frontier Models within the Government Operations Agency, independent of the Department of Technology.

- Reasonable assurance that the developer will not produce a covered model or covered model derivative that poses an unreasonable risk of causing or enabling critical harm. Critical harms include: the creation or use of a chemical, biological, radiological, or nuclear weapon in a manner that results in mass casualties, or at least USD 500 000 000 of damage resulting from cyberattacks, or from an AI model that acts with limited human oversight, intervention, or supervision and results in death, great bodily injury, property damage, or property loss, and would, if committed by a human, constitute a crime specified in the Penal Code that requires intent, recklessness, or gross negligence, or the solicitation or aiding and abetting of such a crime.

A growing coalition of tech companies argue the requirements would discourage companies from developing large AI systems or keeping their technology open-source. Rob Sherman, Meta vice president and deputy chief privacy officer, wrote in a letter sent to lawmakers that “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”.

The text was passed by the Senate and ordered to the Assembly in May 2024. It was then voted to pass as amended and re-referred to the Committee on Appropriations on 2 July. It should be voted by the General Assembly in August. The bill, if passed, could have significant implications for the AI industry in California.

SB-1047 Safe and Secure Innovation for Frontier Artificial Intelligence Models Act

https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202320240SB1047