April 23, 2024

TO: Members, Assembly Privacy and Consumer Protection Committee

SUBJECT: AB 2013 (IRWIN) ARTIFICIAL INTELLIGENCE: TRAINING DATA TRANSPARENCY
OPPOSE UNLESS AMENDED – AS AMENDED APRIL 22, 2024
SCHEDULED FOR HEARING – APRIL 30, 2024

The undersigned organizations must respectfully **OPPOSE UNLESS AMENDED AB 2013 (Irwin)** as amended April 22, 2024, which requires a developer of any artificial intelligence (AI) system or service to post certain documentation on their websites starting on January 1, 2026, regarding data used to train their system or service. We are in no way opposed to transparency, as evidenced by the fact that many companies already provide disclosures voluntarily about how their AI models work. Indeed, robust transparency is a critical component in fostering and maintaining the trust and confidence of our customers and clientele, which are necessary for the widespread uptake of these technologies. At the same time, disclosure mandates must be sufficiently clear to enable compliance and not force companies divulge proprietary information or otherwise confidential business information. Furthermore, *targeted* transparency should be the goal, both to prevent consumer fatigue (as we have seen occur in the context of other well-intentioned disclosure laws), and to avoid unnecessary burdens from being placed on businesses that are subject to multiple regulatory regimes and requirements.

Unfortunately, as currently drafted, we have significant concerns with the approach taken in **AB 2013**, and specifically around overburdensome mandates, the technical feasibility of the bill's transparency measures (including its assignment of responsibilities and the unique challenges presented for different types of developers in meeting the standards set in this bill), insufficient clarity around key terms, and potential exposure to liability. Moreover, we are heavily concerned about **AB 2013's** failure to provide protections for trade secrets and intellectual property, though we do not believe that is the intended outcome of this bill. While it may not be obvious on its face, the expertise and judgment as well as selection of data and datasets is part of what differentiates providers, thereby causing significant concern among developers as to the potential of this bill to undermine their intellectual property and harm competition. And lastly, we question whether the disclosure of training data will result in any substantial benefit when it comes to determining an AI model's performance for a particular use case. Stated another way, simply because a model has been trained on certain data does not mean it will perform as needed in a specific use case.

Regrettably, due to the volume of AI legislation that we must provide feedback on, we are still working on finalizing the precise amendments that would be necessary not only to alleviate our concerns but, ideally, to enable us to support **AB 2013**. These include amendments that address onerous requirements and definitional issues, clarify that developers are not required to disclose any proprietary or otherwise confidential information, and ensure that developers are only required to provide the mandated information to the extent that is practicable and that the bill applies only to high-risk AI, as further highlighted, below. You may note that most of the issues we have highlighted are focused on the introduced version of the bill. We are continuing to evaluate the impact of the recent amendments and will provide additional feedback for your consideration, as soon as possible.

AB 2013 should clearly delineate what is and is not considered "training" and narrow the scope of AI systems or services subject to these transparency measures to high-risk AI systems

One of the major issues we identified with the introduced version of this bill relates to the need to define "training", and to define it narrowly, in order to provide utmost clarity to developers as to their obligations and to ensure that the transparency measures mandated are manageable.

We note that recent amendments have defined "training" to include testing, validating, or fine tuning an AI system or service. We are concerned that amendments expanded the scope of the bill even further and effectively captures all data, regardless of risk level. The bill should be narrowed to scope in only high-risk AI systems. Mandating disclosures for low-risk AI unnecessarily burdens businesses for little to no benefit to the public. A system or service is not a "high-risk" AI system or service, for example, if it is only intended to either perform a narrow procedural task or detect decision-making patterns or deviations from prior decision-making patterns but not meant to replace or influence the previously completed human assessment without proper human review.

For disclosures to be meaningful and not overly burdensome, amendments are also needed to narrow various definitions, starting with the bill's definition of "artificial intelligence system or service". **AB 2013**'s current definition of AI system or service is over broad, arguably capturing regression-based models and even the most rudimentary prediction models or machine-based systems that generate content and make decisions using solely linear functions. Such issues can be addressed by recognizing that the system or service must be capable of "operating with varying levels of autonomy," in line with the OECD[1] AI definition.

AB 2013's definition of "developer" is both overbroad and vague and should include guardrails to address compliance challenges

AB 2013's definition of "developer" is overly broad and fails to clearly indicate if "use by a third party" includes publicly available uses, business-to business uses, or both—or if it even applies to internal uses of a business's subsidiary. (*See* Proposed Section 3110 (b), stating that "developer means …a corporation that designs, codes, or produces an [AI] system or service, or substantially modifies an artificial intelligence system or service for use by a third party for free or for a fee.") For example, as drafted, it is unclear if a corporation would be considered a "developer" subject to **AB 2013** by virtue of allowing a separate, wholly owned subsidiary to use internal AI systems or models that the corporation developed.

Additionally, insofar as a "developer" includes not only those entities that design, code or produce an AI system or service, but also those that "substantially modif[y]" an AI system or service, compliance issues are bound to arise on a couple fronts. First, it is unclear what exactly constitutes "substantially modif[ied]" for these purposes. Second, the lack of clarity as to the obligations of the primary developer, as compared to the secondary developer that substantially modified the system or service, and vice versa, are likely to cause to compliance challenges. Both the primary developer and the secondary developer may lack insight into information that is needed for compliance depending on their obligations under the bill. At minimum, **AB 2013** should include feasibility guardrails to ensure that developers must comply with requirements only to the extent practicable.

We are significantly concerned that recent amendments requiring developers to disclose whether a dataset furthers the intended purpose of the system or service have only exacerbated such problems. While this might make sense for developers that have a limited set of use cases intended for their systems, an open-source developer would not know all the uses of its software – by design. It is misguided to impose this type of requirement on developers broadly, as not all developers have insight into all uses of their models or systems. This obligation amplifies the need for feasibility guardrails.

The application of **AB 2013** to each individual data set, and the detailed disclosures that are necessary for each data set, are burdensome at best and infeasible at worst. There are a lot of different data points within any given data set, making this a significant resource investment. Furthermore, compliance with this requirement can also pose a breach of confidentiality for client work.
Overall, the bill's prescriptive transparency obligations depart from emerging norms in both U.S. and global frameworks around how to promote transparency balanced with competitiveness and feasibility considerations. An obligation to provide a *summary* with information on the data used to train AI models, including categories of personal information used, would establish a more adaptable framework in line with the pace of the development of the technology and reduce concerns around the impact on proprietary or confidential information. To the extent that the goal here is not merely about transparency but more so

[1] Organization for Economic Co-operation and Development. OECD is an intergovernmental organization with 38 member countries, including the United States, founded in 1961 to stimulate economic progress and world trade.

about accountability, another less risky approach might be to require businesses to retain this information for reporting or auditing purposes where there have been detrimental outcomes, as opposed to publishing this information on public websites.

AB 2013 should not apply to AI systems and services that were in use prior to the bill's effective date

Currently, **AB 2013** requires developers to comply by January 1, 2026. Given the extraordinarily broad definition of AI systems or services, and incredibly broad disclosure requirements, this would be a near impossible task for some businesses, especially if the bill were to apply retrospectively, to systems and services that are already in use, or to open-source models. Consider, for example, the requirement to identify the dates that each dataset was first and last used during the development of the system or service. This is not common practice currently and could be incredibly burdensome if not infeasible to trace backward. Nor is it relevant to whether or not a system will effectively and safely perform a particular task.

That said, for compliance to be possible, the bill should at minimum be amended to clearly apply only to AI that is developed from the effective date of the bill, forward. Even then, we believe it is important to also limit this requirement to when the data is first used, as it could become incredibly burdensome to have to constantly update the "last used" date. Alternatively, businesses will require far greater time and flexibility in meeting the bill's mandates, if not totally different mandates for systems and services in existence prior to January 1, 2026, because a developer may not have, or be privy to, all the information necessitated for compliance under **AB 2013**.

AB 2013 should expressly preclude any private right of action

While we greatly appreciate the bill does not expressly include a new private right of action or statutory damages, we hope that language can be added ensuring that the bill shall not be construed provide any basis for a private right of action. This is particularly important where documentation of the data used to train the AI system or service is open-ended (*see* Proposed Section 3111, which states that a developer … shall post on the developer's internet website documentation regarding the data used to train the artificial intelligence system or service, *including, but not limited to*, all of the following …").

Again, we are not opposed to creating a standard around what information is required to be made available by developers of high-risk artificial intelligence systems. However, because the existing definitions and requirements are overly broad and incredibly onerous, if not unattainable, we must **OPPOSE UNLESS AMENDED AB 2013 (Irwin)** in its current form.

Sincerely,

Ronak Daylami
Policy Advocate
California Chamber of Commerce
on behalf of

California Bankers Association
California Chamber of Commerce
California Land Title Association
Insights Association
National Association of Mutual Insurance Companies
Personal Insurance Federation of California
Software & Information Industry Association
TechNet

cc: Legislative Affairs, Office of the Governor
Brandon Bjerke, Office of Assemblymember Irwin
Consultant, Assembly Privacy and Consumer Protection Committee
Liz Enea, Consultant, Assembly Republican Caucus

RD:ldl