May 7, 2024

TO: Members, Assembly Appropriations Committee

SUBJECT: AB 3204 (BAUER-KAHAN) DATA DIGESTERS REGISTRATIONS ACT
OPPOSE – AS AMENDED APRIL 18, 2024
SCHEDULED FOR HEARING – MAY 8, 2024

The undersigned organizations must respectfully **OPPOSE AB 3204 (Bauer-Kahan)** as amended April 18, 2024, as it is overbroad, burdensome, and unnecessary given the protections that already exist under the California Consumer Privacy Act (CCPA) that require entities training artificial intelligence (AI) to respect privacy rights of the consumers to whom that information belongs. We have always greatly respected and appreciated how hard the author works to not only pass strong laws but to do so in a way that strikes a balance among competing rights and interests and avoids unnecessary harm. Unfortunately, in the case of **AB 3204**, the mandated disclosures are not only impractical if not infeasible, but they can also be rather invasive. To the extent that the intent is to promote transparency and understand the degree to which consumers' personal information (PI) is being used to train AI systems, we believe that the approach taken by **AB 3204** is akin to taking a sledgehammer to crack a nut. We are unclear what the end goal of creating a central repository of these businesses is, when reasonable disclosures could accomplish the same purpose without incurring incredible costs to the state and to businesses.

As a general matter, this bill appears modeled off the data broker registry which enables consumers to effectuate their CCPA rights against data brokers. That registry was specifically implemented after the passage of the CCPA, to address a gap in consumer awareness as to the identity of data brokers that might be in possession of their PI. (AB 1202, Chau (Chapter 753, Statutes of 2019)). Creating a central repository was necessary for consumers to identify and initiate requests under the CCPA with third parties with which they do not directly interact. In direct contrast, **AB 3204** would now create a central repository of businesses that train AI using personal data for 1,000 or more individuals or households, despite there not being any circumstances comparable to the ones that necessitated the development of the data broker registry that might warrant this registry and despite the fact that it does not appear to effectuate or enhance any rights associated with the personal data used to train AI.

In creating this registry, **AB 3204** raises several questions. First, the scope of this bill is incredibly broad given the definitions (or lack thereof) referenced, starting with the definition of "data digesters." As introduced the bill would have required any *business* that *trains AI* using *PI* to register as "data digestors." In doing so, **AB 3204** may as well have required all CCPA-covered businesses to register, given the breadth of the CCPA's definitions of the terms "business" and "personal information" and lack of clarity around what is considered "train[ing] AI." Recent amendments seemingly attempt to narrow the definition by clarifying that data digesters are *covered entities* that design, code, or produce, or substantially modify, an AI system or service by training the system or service *on the personal data of 1,000 or more individuals or households*. Realistically, these amendments have no real narrowing effect and will still result in thousands of

businesses—not just large California based businesses[1] or large AI developers—having to register. Bear in mind, of course, that the utility of the data broker registry for consumers was questioned over the inclusion of several hundred businesses. Yet, despite being modeled upon the data broker registry, **AB 3204** fails to include any of the reasonable, but necessary exemptions that are included in either of those laws.

Second, companies that are required to register are not simply asked to provide identifying details about high-risk and low-risk AI alike; for each of those models, they are required to identify *each* category of "PI", as that term is defined under the CCPA, that the data digester uses to train AI, identified *by reference to each applicable subparagraph* within that definition of PI—of which there are twelve. They also must do the same when it comes to sensitive PI (SPI), which adds another half dozen categories.

Any registered business (not just those subject to CMIA) must then also identify each category of information related to consumers' receipt of "sensitive services", as defined under the Confidentiality of Medical Information Act (CMIA, defining sensitive services as those related to mental or behavioral health, reproductive health, gender affirming care, and a host of other services enumerated under nine other provisions of law), that the data digester uses to train AI, identified by reference to the specific category of sensitive service enumerated in the definition. While likely not the intent, this effectively forces hundreds of thousands of businesses to infringe on the privacy of Californians to provide disclosures with the level of detail demanded by **AB 3204**.

Further exacerbating all these issues, is the fact that "PI" under the CCPA, is any "information that identifies, relates to, describes, *is reasonably capable of being associated with, or could reasonably be linked, directly or indirectly, with* a particular consumer *or household*." While deidentified or aggregated consumer data are exempted from the definition, it still captures information that on its own may not be identifiable, but that when pieced together with other pieces of information, becomes identifiable. This is beyond burdensome. It is incredibly impractical, privacy invasive, and at times completely impossible (certainly, not without violations of many other privacy laws). Imagine businesses having to dedicate employees to determine if they could feasibly trace back each individual piece of information used to train AI to a particular individual, and then also review medical records of those individuals to identify if they were provided sensitive services – even if the business has nothing to do with health care. In this way, **AB 3204** fails to take into practical consideration more nuanced relationships between companies, their service providers, and consumers, and the reality that some businesses may use "PI," as understood under the CCPA, to train data, but do not have the necessary relationship with end users, that is needed to comply with the requirement that they identify "each category of information related to consumers' receipt of sensitive services."

Finally, in addition to imposing significant penalties, fines, fees, and expenses that are problematic particularly for smaller businesses, **AB 3204** bill fails to provide any protections or otherwise address copyright and data ownership issues, trade secrets or patents for the information that businesses are required to divulge and that will be made available by the Privacy Agency on a public website. Requiring this level of granular data about each category of PI and SPI used, will invariably force businesses to divulge trade secrets and other highly confidential or patented information, helping their competitors to their own detriment. Again, because this impacts not only businesses developing AI in California, but those doing business in California, should this bill become law, it is highly unlikely that businesses would risk rolling out certain patented tools and algorithms in California, altogether.

Ultimately, it is unclear to us the benefit of creating this registry and requiring such an incredibly cumbersome and privacy-invasive registration process. There are much more reasonable mechanisms that can also ensure that AI developers inform deployers and consumers about the types of AI used to train their consumer-facing AI products. Even if all the legal risks and practical issues of this data digester registry could be addressed, making it possible to comply with bill's disclosures, to what end? The situation addressed by **AB 3204** is not comparable to the data broker registry which was necessary to bridge a gap

[1] The bill's proposed definition of "covered entity" has no requirement that the entity be based in California or otherwise be doing business in California. (See Proposed Section 1798.321(c): "Covered entity" means an organization or enterprise, including, but not limited to, a proprietorship, partnership, firm, business trust, joint venture, syndicate, corporation, association, or nonprofit.)

in consumer awareness that effectively precluded them from effectuating their CCPA rights with certain businesses. In contrast, this registry is not a necessary element to transparency.

For all the aforementioned reasons, and because the bill would so clearly undermine innovation, we must **OPPOSE AB 3204 (Bauer-Kahan)**.

Sincerely,

Ronak Daylami
Policy Advocate
on behalf of

American Council of Life Insurers
Association of California Life and Health Insurance Companies
Association of National Advertisers
California Bankers Association
California Chamber of Commerce
California Credit Union League
California Retailers Association
Computer & Communications Industry Association
Insights Association
Internet Coalition
Los Angeles Area Chamber of Commerce
National Association of Mutual Insurance Companies
Personal Insurance Federation of California
Software & Information Industry Association
TechNet

cc: Legislative Affairs, Office of the Governor
Estefani Avila, Office of Assemblymember Bauer-Kahan
Consultant, Assembly Appropriations Committee
Joe Shinstock, Consultant, Assembly Republican Caucus

RD:ldl