September 6, 2024

The Honorable Gavin Newsom
Governor, State of California
State Capitol
Sacramento, CA 95814

SUBJECT: AB 1008 (BAUER-KAHAN) CALIFORNIA CONSUMER PRIVACY ACT OF 2018: PERSONAL INFORMATION
REQUEST FOR VETO

Dear Governor Newsom:

The California Chamber of Commerce and the undersigned respectfully urge you to **VETO AB 1008 (Bauer-Kahan)**. While earlier iterations of **AB 1008** were even more problematic, altering the definition of what is considered publicly available information and therefore exempted from the definition of "personal information" under the California Consumer Privacy Act (CCPA), the final bill is still significantly problematic. It is not only inaccurate in its representation of the formats in which personal information can exist, but it is misleading both to consumers and businesses, as well. While we offered amendments to the author to correct the error and ensure clarity is maintained in the law, the amendments were inexplicably rejected. As a result, in a bill that seeks to provide clarity, what we are left with is unnecessary confusion where none previously existed before.

AB 1008 seeks clarity that is clearly not needed

As reflected in the final floor analyses, the purpose of **AB 1008** is to clarify that personal information can exist in various formats, regardless of how it is transmitted or stored. Such clarity, however, is neither needed, nor helpful in this instance due to the inaccuracies reflected in **AB 1008**.

As our organization has publicly stated on numerous occasions in recent years, the CCPA is a comprehensive, industry neutral, technology neutral law that already provides strong consumer privacy protections around the collection, use, and disclosure of all Californians' personal information. Terms like collection and personal information (PI) capture far more than they ever did before under other laws: it does not matter if PI is collected actively or passively or if the PI can only be indirectly linked to an individual as opposed to directly linked. The law applies to traditional, brick and mortar businesses, as well as technology companies and companies that are exclusively online. And the law applies whether it is being collected using traditional modes of pen and paper, printed copies, or electronic devices and digital documents, or otherwise. The CCPA was intentionally drafted to be incredibly broad.

In other words, **there has been zero dispute**: the format it appears in, is transmitted by, or is stored in, by and large, does not alter the nature of the protection afforded to information under the CCPA. It can be a copy of a furniture delivery slip stored on an electronic file on a desktop, just as easily it can be stored in a physical one in a filing cabinet. It can be records kept a shoebox, a database, or the cloud. It can be information in an online account or reflected on a paper bill. It can be a call center voice recording, an image, or fingerprint. It can get transmitted to the business by fax, email, mail, by hand, or on foot; without the use of technology just as easily as it can with the use of technology. These examples are not provided facetiously. It's to illustrate that it is precisely what makes the law so protective and strong, and at the same time so complex and tricky, making certain bill proposals more problematic than they might have appeared.

And that is why we are unclear what **AB 1008** is trying to solve for, when there has clearly never been any confusion on this matter. Absent a specific exemption, unless it is publicly available information, deidentified

information, or aggregate consumer information– all of which have specific meanings under the CCPA – personal information is personal information. But certainly, if the goal is to codify clarity, it needs to be done with accuracy, particularly on a topic as complex as AI. Treating "AI systems capable of outputting personal information" as a "format" in which personal information can be stored only adds confusion for businesses and consumers as to their CCPA responsibilities and rights.

AB 1008's inclusion of "AI systems that output personal information" as a format that personal data can take is inaccurate and misleading

Our primary concern here is that in seeking clarity, the bill would in fact create uncertainty and confusion where none previously existed. As in print, **AB 1008** lists different "formats" in which personal information "can exist". These include physical formats such as paper documents or video tapes, digital formats including text, audio or video files, and "abstract digital information" such as compressed or encrypted files, metadata, or "*artificial intelligence systems that are capable of outputting personal information*." Putting aside the fact that the bill fails to even define "AI systems", which is itself a significant problem (whereas every other AI bill that has gone through this Legislature has at least seen the same definition of AI put into them), the issue here is this final clause.

Seeking to treat "AI systems capable of outputting PI" as a "format" in which PI can be stored is simply not grounded in science. To list it as an example of "abstract digital formats" suggests there is PI in the AI system beyond that embodied in text, images, video. This risks functionally expanding the law to include trained model weights as a format in which PI can exist. Model weights are just numbers and themselves do not contain any PI. To say or even suggest otherwise is simply not accurate and will create significant risk and uncertainty for model developers. While we had hoped that the author would have accepted our suggested amendment in the Senate, at this point, creating, if not perpetuating, such an inaccuracy needs to be seriously reconsidered.

Other inaccuracies reflected in public analyses require clarification and in fact calls into question what the CCPA currently does and does not apply to

The Senate Floor analysis referenced a paper that the author cited in support of **AB 1008**, which lists Google researchers and others, among the authors. The problem with that citation and conclusion drawn from it is twofold.

First, the description provided implies that the paper supports the underlying premise of **AB 1008** and the technical foundations upon which it attempts to classify all AI systems as a "format" that PI "can exist in." The research paper, which relates to LLMs and compression ("Language Modeling Is Compression,"), is more complex than the title might suggest. It explores the potential of LLMs that have already been trained to function as highly effective compressors on various forms of media by comparing it to other lossless compression techniques. But the paper does not conclude that all AI systems are forms of compression of their training data as the author statement in the Senate Floor Analysis suggests[1]. Not only does the research not support **AB 1008**, but it also reinforces that LLMs, which are only one form of AI system, are prediction engines. It is those prediction capabilities that make fully-trained LLMs effective for compression of other content such as spreadsheets as noted in the Floor Analysis.

Second, after incorrectly conflating text-based genAI systems as a form of data compression, similar to zip files, the author concluded that "[...] a business could conceivably use a language model (like ChatGPT or LLAMA) to compress personal information and transfer it to a buyer. This business should be subject to CCPA, just as if they had sent a compressed spreadsheet containing the same information." (Senate Floor analysis, page 5.) In a bill seeking to provided clarity, perhaps the only thing worse than creating confusion where none previously existing by codifying inaccurate statements, is suggesting that existing law does not

[1] Insofar as the author seems to have drawn that conclusion, it bears repeating: **models do not store personal data, even if they were trained on it**. A recent report issued by the Hamburg Commissioner for Data protection and freedom of information (the state Data Protection Authority for the German state of Hamburg, in charge of enforcing the EU's GDPR) specifically examined that question, and whether LLMs, functioning as a component of an AI system, stores personal data when the output may contain information relating to natural persons: it does not. "*When training data contains personal data, it undergoes a transformation during machine learning process, converting it into abstract mathematical representations. This abstraction process results in the loss of concrete characteristics and references to specific individuals. Instead, the model captures general patterns and correlations derived from the training data as a whole*."

apply/protect certain information when it does. Consumers get misled to believe that there is something nefarious currently happening, and businesses are left confused as to their obligations and whether they actually only have to apply the CCPA to PI in certain contexts over others.

Finally, we note that to the extent there are concerns around how genAI systems are trained[2], such concerns are more directly addressed by way of transparency measures found in a separate bill passed by the Legislature, AB 2013 (Irwin). In contrast, **AB 1008**, does not help resolve any issues or add clarity to the law—it merely creates new ones.

Ultimately, regulations should build public confidence and not create doubts in the status quo where none should exist. Thus, because we believe that we should not be codifying legal fiction about how AI systems operate and creating confusion where none previously existed, we request you **VETO AB 1008 (Bauer-Kahan)**.

Sincerely,

Ronak Daylami
Policy Advocate
on behalf of

American Council of Life Insurers
Association of California Life and Health Insurance Companies
Bay Area Council
California Chamber of Commerce
Computer and Communications Industry Association
Insights Association
Software & Information Industry Association
TechNet

[2] Part of the impetus for the bill, per the author's statement in the Assembly floor analysis, specifically pertains to concerns relating to the training of genAI: "Today, advanced GenAI systems are frequently trained using data obtained through the untargeted and automated scraping of internet websites. Once trained, these systems are capable of accurately reproducing their training data, including Californians' personal information."