FLOOR ALERT
AB 3211 (Wicks):
OPPOSE as Amended – 4/18/2024

TechNet and the following organizations must respectfully oppose AB 3211 (Wicks), which sets overly prescriptive and technological infeasible requirements on developers of artificial intelligence (AI), large online platforms, camera and recording device manufacturers to incorporate content provenance and watermarking technology into their products.

We agree with the intent to create greater trust in user generated content online by fostering the adoption of content provenance verifications and watermarks. However, this bill presents a multitude of issues and requires platforms to comply with technically infeasible and impossible standards.

In light of the immense volume of artificial intelligence (AI) and AI related bills in California and across the country, our member companies are still reviewing this bill and many others. We hope to be able to provide suggested amendments soon. In the meantime, there are many issues with this bill that we should highlight for consideration.

Federal Standard

While we understand the desire to regulate an emerging technology, this is an area that would benefit from Federal standards and regulation rather than a state by state approach. In President Biden's AI Executive Order, he tasked the Department of Commerce with "identifying the existing standards, tools, methods, and practices, as well as the potential development of further science-backed standards and techniques, for: (i) authenticating content and tracking its provenance; (ii) labeling synthetic content, such as using watermarking; (iii) detecting synthetic content" and more. We believe in allowing this federal process to advance in order to establish standards that are "science-backed" and can be consistently applied across the country is important.

In the meantime, disclosure provides a meaningful way to alert consumers when they are interacting with AI-generated content. We would prefer an approach that prioritizes disclosure in the short-term while watermarking technologies are developing and able to advance to be deployed in a consistent and cost-effective manner down the line. This approach would allow consumers and businesses to benefit from transparency and allow time for further innovation with regard to watermarking.

Prescriptive Requirements on Content Provenance and Watermarking are Technologically Premature

Many of our companies and platforms are at the forefront of developing content provenance and watermarking technology, which is still in its early stages. However, AB 3211 enacts incredibly prescriptive requirements for a technology that is still under development and rapidly evolving. For example, there isn't a program that can watermark text, making the bill's requirements to do so impossible to comply with. We believe references to text watermarking should be removed to reflect this reality.

Furthermore, content provenance and watermarking is still incredibly unreliable and in many cases easy to break. Researchers at the University of Maryland were able to break all the currently available watermarking methods. Some can be avoided by simple cropping, resizing, or screenshotting an image. More concerning, these researchers were able to insert fake watermarks and credentials into images, creating false positives. Provenance and watermarking tools tend to help good faith actors act virtuously, but they have limits on stopping bad actors. No provenance solution that's been created so far, including watermarking or metadata, stops bad actors from simply 'stripping' provenance elements and posting a fake piece of content as authentic.

In its standards for large online platforms, AB 3211 should more clearly delineate between 1st party and 3rd party content. 1st party content would be images, videos, or audio that is generated using a large online platform's generative AI tools and is then posted or distributed on that platform. In this instance, a platform can actually control the creation of a content provenance or watermark into the content. As mentioned, many of our companies are already working to incorporate this type of technology to increase transparency around AI-generated content. It is currently technically infeasible to accurately and reliably detect content that is created using a different platform's AI tools. As noted above, considering the current ease with which current watermarks can be broken, a legal requirement and mandate for 3rd party content isn't appropriate.

Privacy Concerns

In general, content credential requirements raise several privacy concerns. Content credentials allow users to see information about who created an image. This can give content creators more ownership over their work but also leaves traces of user data that can now be widely accessed.

Moreover, AB 3211 requires platforms to use "state-of-the-art techniques" to detect and label inauthentic content that is uploaded by users. As a threshold matter, "state-of-the-art techniques" is a term of art used several times throughout the bill and has no legal meaning. It should be struck in favor of a clear standard, especially considering the exorbitant penalties the bill would seek to levy for any violation ($1 million or 5% of global annual revenue, whichever is higher).

Section (e)(2) authorizes platforms to use privacy intrusive methods to detect inauthentic text content including spyware and user authentication. The bill allows platforms to consider a user's typing cadence, which is only possible with spyware that can measure and record that information. The bill also allows a platform to verify that users are matched to their unique device identifier such as a subscriber identity module (SIM) card or multifactor authentication (MFA). This seems to give platforms the authority to destroy users' anonymity in the pursuit of detecting and labeling inauthentic content. To be clear, our platforms are firmly against such invasive measures into our users' information. It is a highly concerning precedent that California would be setting.

Additionally, (f) requires a platform to create a verification process for users to apply a digital signature to content created by a human being. However, this would effectively require a user to identify and authenticate themselves in order to prove they were the content creator. As we've raised in several bills before the Legislature this year, there are not any reliable methods to identify and verify users that do not require the collection of more personal information, such as government IDs. The bill acknowledges uploading a government-issued identification and matching picture identification or verifying a user possesses a unique device with a SIM card and active phone number. At the same time though it requires the platform to verify users in a variety of options "that do not necessarily require disclosure of personal identifiable information". This is simply not possible.

Enforcement

Considering the prescriptive nature and technical infeasibility of some requirements and the technical impossibility of others, we believe the penalties for noncompliance of $1 million or 5% of global annual revenue are unjustifiable. As mentioned above, platforms cannot watermark text content. The technology to apply watermarks to audio, images, and videos are much further along but vary in their resistance to removal or inadvertent breaking.

To help illustrate how exorbitant a potential fine of 5% of global annual revenue is, consider other bills the Legislature is considering this year regarding the intentional creation of misleading and deceptive media to influence an election. AB 2839 (Pellerin) prohibits the intentional distribution of materially deceptive and digitally altered campaign material with the intent to influence an election. The bill would authorize a candidate or committee which had been targeted by the materially deceptive and digitally altered campaign material to seek general or special damages as well as injunctive or equitable relief against the creator of the material. A person *intentionally* trying to manipulate and influence an election with deceptive, AI-generated media is liable for general or special damages while a developer or large online platform doing their best to comply with the infeasible and impossible requirements of this bill could be charged 5% of their annual global revenue for their efforts.

Relatedly, AB 3211 creates an undefined new process that grants sweeping authority to the California Department of Technology to require companies to turn their data and information over to "independent researchers". There are no requirements outlined for vetting independent researchers, data and intellectual property protections, and it would be ripe for abuse. We believe this requirement should be struck.

AI-Generated Content Isn't Inherently Bad

AB 3211 seems to treat all AI-generated content as inherently bad or risky. By requiring such thorough and prescriptive requirements for content labeling, the bill makes a value judgment that consumers must be notified and aware of any content that was created by AI. This applies to all inauthentic content including purely artistic or satirical content. While we agree with the intent to provide more information to consumers, in some instances it could create disclosure or notification fatigue. If watermarks and content credentials become so routine and placed on all AI-generated content, users may start to ignore and disregard their presence. Rather than focusing on whether the content itself was AI-generated, synthetic, or inauthentic, we would advise focusing on the misuse of this technology.

VOTE "NO" ON AB 3211