Date of Hearing: June 18, 2024

ASSEMBLY COMMITTEE ON PRIVACY AND CONSUMER PROTECTION
Rebecca Bauer-Kahan, Chair
SB 942 (Becker) – As Amended May 16, 2024

AS PROPOSED TO BE AMENDED

SENATE VOTE: 32-1

SUBJECT: California AI Transparency Act

As generative artificial intelligence (GenAI) technologies become more accessible, online content that appears real – but is actually false – threatens to flood social media and other large online platforms. The unmitigated spread of synthetic content threatens to harm individual Californians in numerous ways, such as through the proliferation of nonconsensual deepfake pornography, scams, and the distribution of targeted election disinformation.

The issue of GenAI-produced content can be tacked through a three-step plan. First, all artificial content must be labeled as such. Second, all "real" content – such as video and audio recordings – must be labeled as such. Third, large online platforms must be required to prominently display these labels, allowing users to distinguish between artificial and real content.

SB 942 seeks to enact step 1 of this plan. It would require GenAI providers to embed latent, machine-readable disclosures into content produced by their systems. These disclosures would include information about the origin of the content, and would be designed to be difficult to remove. GenAI providers would also be required to develop and make freely available software that identifies content generated by their systems. Finally, the bill requires GenAI providers who license their systems to contractually require licensees to maintain the systems' disclosure capabilities.

This bill is author-sponsored and supported by PERK Advocacy and Transparency Coalition.ai. It is opposed by a coalition of industry associations including Technet, Netchoice, Computer and Communications Industry Association, and the California Chamber of Commerce.

Proposed Committee amendments revise and clarify the bill. The full text of the amended bill is included at the end of the analysis. If the bill passes this Committee, it will next be heard by the Assembly Judiciary Committee.

SUMMARY: Requires the developers of GenAI systems to both include provenance disclosures in the content their systems produce, and make tools available to identify GenAI-content produced by their systems. Specifically, **this bill**:

1) Defines artificial intelligence (AI) to mean an engineered or machine-based system that varies in its level of autonomy and that can, for explicit or implicit objectives, infer from the input it receives how to generate outputs that can influence physical or virtual environments.

2) Defines GenAI to mean artificial intelligence that can generate derived synthetic content, including text, images, video, and audio, that emulates the structure and characteristics of the system's training data.

3) Defines "covered provider" to mean a person that creates, codes, or otherwise produces a GenAI system that has over 1,000,000 monthly visitors or users and is publicly accessible within the geographic boundaries of the state.

4) Defines "provenance data" to mean data that is embedded into digital content, or that is included in the digital content's metadata, for the purpose of verifying the digital content's authenticity, origin, or history of modification.

5) Defines "personal provenance data" to mean provenance data that contains either personal information or unique device, system, or service information that is reasonably capable of being associated with a particular user.

6) Defines "system provenance data" to mean provenance data that is not reasonably capable of being associated with a particular user and that contains either information regarding the type of device, system, or service that was used to generate a piece of digital content, or information that provides proof of content authenticity.

7) Requires a covered provider to create and make freely available an AI detection tool that can be used to assess whether digital content was created or altered by the covered provider's GenAI system, and that has various specified characteristics.

8) Prohibits a covered provider from collecting or retaining personal information from users of the covered provider's AI detection tool, or retaining personal provenance data from user-uploaded content, except as specified.

9) Requires a covered provider to offer a user of a GenAI tool the option to include a manifest disclosure in generated content, with various specified characteristics.

10) Requires a covered provider include a latent disclosure in generated content, with various specified characteristics.

11) Requires that if a covered provider licenses their GenAI system to a third party, the covered provider shall require by contract that the licensee maintain the system's capability to include a latent disclosure.

12) Requires that if a covered provider knows a licensee has removed the capability of a licensed GenAI system to include a latent disclosure, the covered provider shall revoke the licensee's contract.

13) Requires that a licensee cease using a licensed GenAI system after its license has been revoked.

14) Provides that a covered provider that violates the chapter be liable for a civil penalty of $5000 to be collected in a civil action filed by the Attorney General only.

15) Provides that the Attorney General, a district attorney, a county counsel, or a city attorney may bring a civil action for injunctive relief and reasonable attorney's fees and costs against a third party licensee who violates this chapter.

EXISTING LAW:

1) Provides, pursuant to the California Constitution, that all people are by nature free and independent and have inalienable rights. Among these are the fundamental right to privacy. (Cal. Const. art. I, § 1.)

2) States that the "right to privacy is a personal and fundamental right protected by Section 1 of Article I of the Constitution of California and by the United States Constitution and that all individuals have a right of privacy in information pertaining to them." Further states these findings of the Legislature:

   a) The right to privacy is being threatened by the indiscriminate collection, maintenance, and dissemination of personal information and the lack of effective laws and legal remedies.

   b) The increasing use of computers and other sophisticated information technology has greatly magnified the potential risk to individual privacy that can occur from the maintenance of personal information.

   c) In order to protect the privacy of individuals, it is necessary that the maintenance and dissemination of personal information be subject to strict limits. (Civ. Code § 1798.1.)

3) Defines "deepfake" to mean audio or visual content that has been generated or manipulated by artificial intelligence (AI) which would falsely appear to be authentic or truthful and which features depictions of people appearing to say or do things they did not say or do without their consent. Requires the Secretary of Government Operations to evaluate the impact of the proliferation of deepfakes on the state. (Gov. Code § 11547.5.)

4) Defines "personal information" to mean 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. States that personal information includes, but is not limited to, the following if it identifies, relates to, describes, is reasonably capable of being associated with, or could be reasonably linked, directly or indirectly, with a particular consumer or household (Civ. Code § 1798.140(v).):

   a) Identifiers such as a real name, alias, postal address, unique personal identifier, online identifier, Internet Protocol address, email address, account name, social security number, driver's license number, passport number, or other similar identifiers.

   b) Any personal information described in Section 1798.80(e).

   c) Characteristics of protected classifications under California or federal law.

   d) Commercial information, including records of personal property, products or services purchased, obtained, or considered, or other purchasing or consuming histories or tendencies.

e) Biometric information.

f) Internet or other electronic network activity information, including, but not limited to, browsing history, search history, and information regarding a consumer's interaction with an internet website application, or advertisement.

g) Geolocation data.

h) Audio, electronic, visual, thermal, olfactory, or similar information.

i) Professional or employment-related information.

j) Education information, defined as information that is not publicly available personally identifiable information as defined in the Family Educational Rights and Privacy Act. (20 U.S.C. Sec. 1232g; 34 C.F.R. Part 99).

k) Inferences drawn from any of the information identified in this subdivision to create a profile about a consumer reflecting the consumer's preferences, characteristics, psychological trends, predispositions, behavior, attitudes, intelligence, abilities, and aptitudes.

l) Sensitive personal information.

FISCAL EFFECT: As currently in print this bill is keyed fiscal.

COMMENTS:

1) **AI and GenAI.** The development of GenAI is creating exciting opportunities to grow California's economy and improve the lives of its residents. GenAI can generate compelling text, images and audio in an instant – but with novel technologies come novel safety concerns.

What is AI? In brief, AI is the mimicking of human intelligence by artificial systems such as computers. AI uses algorithms – sets of rules – to transform inputs into outputs. Inputs and outputs can be anything a computer can process: numbers, text, audio, video, or movement. AI is not fundamentally different from other computer functions; its novelty lies in its application. Unlike normal computer functions, AI is able to accomplish tasks that are normally performed by humans.

AI that are trained on small, specific datasets in order to make recommendations and predictions are sometimes referred to as "predictive AI." This differentiates them from GenAI, which are trained on massive datasets in order to produce detailed text and images. When Netflix suggests a TV show to a viewer, the recommendation is produced by predictive AI that has been trained on the viewing habits of Netflix users. When ChatGPT generates text in clear, concise paragraphs, it uses GenAI that has been trained on the written contents of the internet.

GenAI tools can be released in open-source or closed-source formats by their creators. Open-source tools are publically available; researchers and developers can access their code and parameters. This accessibility increases transparency, but it has downsides: when a tool's code and parameters can be easily accessed, they can be easily altered, and open-source tools have the potential to be used for nefarious purposes such as generating deepfake pornography and targeted propaganda. By comparison, closed-source tools are opaque with respect to their

security features. It is harder for bad actors to generate illicit materials using these tools. But unlike open-source tools, closed-source tools are not subject to collective oversight because their inner workings cannot be examined by independent experts.

2) **Deepfakes and disinformation.** Image manipulation and video doctoring have existed for nearly as long as photography and recording equipment, but they have historically required great effort and talent. In the past few years the rapid development of GenAI has drastically reduced those barriers to entry, allowing a vast quantity of convincing – but ultimately fake – content to be generated in an instant. The creation of imagery, video, and audio by GenAI has the potential to change the world by automating repetitive tasks and fostering creativity. When employed by bad actors, however, these capabilities have the potential to destroy lives and destabilize societies.

Nonconsensual pornography. GenAI has been used to create pornography since its inception. This content is inevitably nonconsensual, and as GenAI technology improves, these products will become harder to distinguish from reality. Women are the primary victims of these efforts; in the run-up to the 2024 Super Bowl, a series of images involving Taylor Swift began to appear on the social media platform X (formerly Twitter). These images were removed over the following days, but the damage had been done:

> "We are too little, too late at this point, but we can still try to mitigate the disaster that's emerging," says Mary Anne Franks, a professor at George Washington University Law School and president of the Cyber Civil Rights Initiative. Women are "canaries in the coal mine" when it comes to the abuse of artificial intelligence, she adds. "It's not just going to be the 14-year-old girl or Taylor Swift. It's going to be politicians. It's going to be world leaders. It's going to be elections."[1]

The harms of nonconsensual AI-powered pornography are already being felt in California:

> A third school in Southern California has been hit with allegations of digitally manipulated images of students circulating around campus . . . "Sixteen eighth-grade students were identified as being victimized, as well as five egregiously involved eighth-grade students," Superintendent Michael Bregy wrote. While Bregy acknowledged that children "are still learning and growing, and mistakes are part of the process," he affirmed disciplinary measures had been taken and noted that the incident was swiftly contained. The district vowed to hold accountable any other students "found to be creating, disseminating, or in possession of AI-generated images of this nature."[2]

Scams. GenAI-powered speech and video is driving a new era in scamming. These AI tools are often trained on publicly available data – the more data a target has online, the easier it is to develop a passable imitation of them or their loved ones. This is especially true of wealthy clients, whose public appearances, including speeches, are often widely available on the

---

[1] Brian Contreras, "Tougher AI Policies Could Protect Taylor Swift—And Everyone Else—From Deepfakes," Feb. 8. 2024, www.scientificamerican.com/article/tougher-ai-policies-could-protect-taylor-swift-and-everyone-else-from-deepfakes/.

[2] Mackenzie Tatananni, "'Inappropriate images' circulate at yet another California high school, as officials grapple with how to protect teens from AI porn created by classmates," *Daily Mail,* Apr. 11, 2024, https://www.dailymail.co.uk/news/article-13295475/Inappropriate-images-California-Fairfax-High-School-AI-deepfake.html

internet.[3] For example, a complicated scam utilizing both deepfake video and false audio was recently performed in Hong Kong. A multinational company lost $25.6 million after employees were fooled by deepfake technology, with one incident involving a digitally recreated version of its chief financial officer ordering money transfers in a video conference call. Everyone present on the video call, except the victim, was a fake representation of real people. The scammers applied deepfake technology to turn publicly available video and other footage into convincing versions of the meeting's participants.[4]

Political propaganda and disinformation. Deepfake technology is being used around the world to spread disinformation and propaganda. 2024 is a major election year in democracies around the globe: at least 64 countries will hold elections, representing close to 49% of the world's population.[5] It is also likely to be the first of many election years in which AI plays a pivotal role, as the technology becomes more widely available and easier to use. This has already been observed in Slovakia, where deepfake audio influenced an election in 2023:

> Days before a pivotal election in Slovakia to determine who would lead the country, a damning audio recording spread online in which one of the top candidates seemingly boasted about how he'd rigged the election. And if that wasn't bad enough, his voice could be heard on another recording talking about raising the cost of beer. The recordings immediately went viral on social media, and the candidate, who is pro-NATO and aligned with Western interests, was defeated in September by an opponent who supported closer ties to Moscow and Russian President Vladimir Putin.[6]

Similar deepfakes have now been deployed in the United States in advance of the 2024 presidential election. In late January, between 5000 and 20,000 New Hampshire residents received AI-generated phone calls impersonating President Biden that told them not to vote in the state's primary. The call told voters: "It's important that you save your vote for the November election." Concern about this call has led at least 14 states to introduce legislation targeting AI-powered disinformation.[7] It is still unclear how many people might not have voted based on these calls.

Deepfakes are not only being deployed by third parties; they can be used by the candidates themselves, either to improve their own self-images or to detract from their opponents. In mid-2023, former Republican presidential candidate Governor Ron DeSantis used AI to add fighter

---

[3] Emily Flitter and Stacy Cowley, "Voice Deepfakes Are Coming for Your Bank Balance", *New York Times*, Aug. 30, 2023, www.nytimes.com/2023/08/30/business/voice-deepfakes-bank-scams.html.
[4] Harvey Kong, "'Everyone looked real': multinational firm's Hong Kong office loses HK$200 million after scammers stage deepfake video meeting," *South China Morning Post*, Feb. 4, 2024, www.scmp.com/news/hong-kong/law-and-crime/article/3250851/everyone-looked-real-multinational-firms-hong-kong-office-loses-hk200-million-after-scammers-stage.
[5] Koh Ewe, "The Ultimate Election Year: All the Elections Around the World in 2024," *Time*, Dec. 28, 2023, https://time.com/6550920/world-elections-2024/.
[6] Curt Devine, Donie O'Sullivan, Sean Lyngass, "A fake recording of a candidate saying he'd rigged the election went viral. Experts say it's only the beginning," *CNN*, Feb. 1, 2024, www.cnn.com/2024/02/01/politics/election-deepfake-threats-invs/index.html.
[7] Adam Edelman, "States turn their attention to regulating AI and deepfakes as 2024 kicks off," *NBC News*, Jan. 22, 2024, www.nbcnews.com/politics/states-turn-attention-regulating-ai-deepfakes-2024-rcna135122.

jets to one of his campaign videos.[8] Around the same time, Governor DeSantis' super PAC released an ad containing an AI-generated speech by former president Donald Trump.[9] The Republican National Committee also released a 30-second ad that displayed images of disorder and destruction, with a voiceover that described the "consequences" of re-electing President Biden.[10] None of the images in this ad were real.

3) **Content provenance.** Many of the issues associated with deepfakes could be resolved, if only there were a reliable way to identify GenAI content. While at present no single solution exists, there are ongoing efforts to embed information related to "content provenance" – the verifiable history of a piece of content – into both GenAI products and the products of real-life recorders, such as digital cameras. Under this framework, the users of social media platforms would be able to rely on provenance data to identify trustworthy content.

Metadata. The Merriam-Webster Dictionary defines metadata as "data that provides information about other data." In practice, metadata is a structured set of descriptors attached to digital content. A photograph's metadata might include information about the camera used to take the photo, the time and date the photo was taken, and the photo's precise geolocation. A written document's metadata may include details about its author, its creation date, and the number of times the document has been edited. Metadata can be used to verify content authenticity, helping to combat fake news by providing a traceable history of content creation and modification. But metadata can also contain personal information about the individual who creates or modifies a piece of content.

Watermarking. Watermarking is the process of embedding an identifiable marker into digital content. This "watermark" can be visible, such as a logo or text overlay, or it can be invisible, data embedded into a file in a manner that does not noticeably alter the content. As a technology, watermarking is in its infancy. Watermarks can be stripped from content relatively easily by common screenshotting tools, file compression software, and image editing programs like Photoshop. They can also be faked by treating a GenAI system like a copy machine: a real image or video can be fed into a GenAI system and spit back out, unaltered except for the addition of a watermark. The system's user receives a piece of authentic content that has been incorrectly marked as inauthentic, ready to be posted online in order to create confusion and sow discord. Furthermore, while progress has been made towards developing standards for watermarking of images and video, how text should be watermarked is far less clear.

4) **What this bill would do.** This bill has four basic components. First, it requires GenAI system providers include disclosures in content produced by their systems. The bill describes two types of disclosures: manifest and latent. Manifest disclosures are optional, and primarily serve to clearly identify content as AI-generated. Latent disclosures are mandatory, and serve as an imperceptible indicator that content is AI-generated. These disclosures are required to be interpretable by a provider's AI detection tool, detailed below.

---

[8] Ana Faguy, "New DeSantis Ad Superimposes Fighter Jets In AI-Altered Video Of Speech," *Forbes*, May 25, 2023, www.forbes.com/sites/anafaguy/2023/05/25/new-desantis-ad-superimposes-fighter-jets-in-ai-altered-video-of-speech/.
[9] Alex Isenstadt, "DeSantis PAC uses AI-generated Trump voice in ad attacking ex-president," *Politico*, Jul. 17, 2023, www.politico.com/news/2023/07/17/desantis-pac-ai-generated-trump-in-ad-00106695.
[10] GOP, "Beat Biden," Apr. 25, 2023, https://www.youtube.com/watch?v=kLMMxgtxQ1Y.

Second, it requires GenAI system providers create and make freely available AI detection tools that can identify content created by the providers' systems. When a user uploads content to be analyzed, the tool is required to output information related to the system-of-origin of the content, and withhold any personal information included in the disclosure.

Third, it requires that GenAI system providers who license their systems to third parties contractually obligate licensees to retain the systems' disclosure capabilities. If a provider knows that a licensee has stripped a licensed system of its disclosure capabilities, the provider is required to revoke their license. A licensee who has had their license revoked must stop using the system.

Fourth, the bill outlines enforcement mechanisms. Under the bill's enforcement scheme, a GenAI system provider who violates the bill's provisions is liable for a civil penalty of $5000, to be collected in a civil action filed only by the Attorney General. A third party licensee who violates the bill's provisions may have a civil action brought against them by the Attorney General, a district attorney, a county counsel, or a city attorney, for injunctive relief and reasonable attorney's costs and fees.

5) **Author's statement.** According to the author:

> Generative Artificial Intelligence (AI) is advancing at an unprecedented pace, revolutionizing industries and everyday life. However, this rapid progress has also sparked a pressing need for regulatory frameworks to govern its use. While AI offers innovative solutions that drive economic growth and efficiency, it simultaneously presents complex challenges related to misinformation, manipulation, bias, and ownership of generated content.
>
> One of the most concerning aspects of unregulated AI models is their potential to produce "deepfake" content, such as manipulated images and videos, which can be used to deceive or manipulate individuals and society at large. Deepfake pornography and fabricated political messages have already emerged as serious threats, highlighting the urgency of implementing effective regulations.
>
> As AI-generated content approaches a level of realism that makes it indistinguishable from genuine content, the need for transparency becomes paramount. It is crucial for industry leaders to take responsibility for clearly labeling AI-generated content and developing tools that can accurately identify such content, especially when there is no visible disclosure. The transparency SB 942 would provide is essential for establishing trust in AI technologies and ensuring that users can make informed decisions about the content they consume.

6) **Analysis.** As noted in this committee's analysis of AB 3211 (Wicks, 2024), the issue of content authentication can theoretically be solved in three easy steps:

1. Require all content produced by GenAI to be labeled "fake."

2. Require all content produced by cameras and other recording devices to be labeled "real."

3. Require all social media platforms to prominently display these labels.

Passing the current bill would represent meaningful progress towards "step 1" of this plan. SB 942 requires that "disclosures" be included in content produced by GenAI systems. Of the

two disclosures required by the bill – manifest (visible), and latent (imperceptible to the human eye) – the latent disclosure requirement is more important by far. The bill describes the minimum information that these disclosures must convey:

1. The name of the covered provider.
2. The name and version number of the GenAI system used.
3. The time and date of the content's creation or alteration.
4. Which parts of the content were created or altered by the GenAI system.
5. A unique identifier.

It may not be possible to embed all of these details directly into a single disclosure, especially in content that is small or low-resolution. The bill accounts for this in two ways: first, it allows a GenAI provider to instead embed a link to a permanent internet website into the content. This link, which would be comparatively small, could be repeated many times within a single piece of content. This would impart a degree of stability onto the disclosure – cropping or obscuring large parts of the content would not necessarily affect the disclosure's ability to be read. Second, the bill does not actually specify a disclosure mechanism for GenAI content. The bill very carefully avoids the term "watermark" – a GenAI provider may choose to use a watermark, but they may also choose to adopt a non-watermark-based approach. The language surrounding this disclosure is left intentionally ambiguous.

Aside from requiring disclosures in GenAI content, the bill requires GenAI providers develop and make freely available tools to identify content generated by their systems. This approach is fraught: as pointed out by Oakland Privacy's "support if amended" letter, AI detection tools are frequently found to be unreliable:

> How bad are AI detection tools? Pretty bad. Content-At-Scale, one of the largest AI detection tools identified sections of the United States Constitution as highly likely to be AI generated. The most developed tools address text and rely on certain indicators including predictability, variety of sentence length and structure, word repetition, unnatural word usage, and inconsistent verb tense. It goes without saying that many of these are also characteristics of poorly written human-generated text. The tools are dealing in probabilities and while they may effectively flag bad writing that *could* be AI-generated, they are simply not accurate enough to be relied on as evidence of anything.

However, the current bill does not require the creation of tools that detect all AI-generated content; instead, this bill only requires GenAI providers be able to identify their own systems' outputs. Having intimate knowledge of the exact disclosure mechanisms included in their systems' outputs may allow GenAI developers to succeed here, despite the frequent failures of more "general" tools. This is even truer of GenAI systems that exist entirely online: by keeping a record of all content their systems produce (perhaps in a privacy-protective form, like a cryptographic hash), GenAI providers can quickly and easily determine whether user-uploaded content was generated by their systems.

Furthermore, the examples referenced by Oakland Privacy relate primarily to the detection of AI-generated text. The current bill recognizes that embedding disclosures into text is not yet feasible

at scale, and excludes text from the list of digital content that requires disclosure. However, as the inclusion of the phrase "other digital content" could be construed to include text, the authors may wish to edit the language of the bill to clarify whether or not AI-generated text is covered.

Relationship to AB 3211 (Wicks, 2024). AB 3211 would require GenAI providers and the manufacturers of recording devices to include watermarking capabilities in the systems and devices they make available in California, and require social media platforms to label content with provenance info extracted from uploaded content. This bill overlaps with AB 3211 nearly entirely. Where this bill only seeks to address "step 1" in the three-step content authentication plan outlined above, AB 3211 seeks to implement the full plan. Where this bill lacks a particular mechanism for provenance disclosure, AB 3211 specifically requires watermarking. And where this bill's philosophy is to set a "floor" that GenAI providers are able to innovate around and above, AB 3211 is far more prescriptive in its requirements.

Relationship to AB 1791 (Weber, 2024). AB 1791 would require social media platforms to strip provenance data containing personal information from content uploaded to the platform, while retaining provenance data related to system-of-origin. If this bill represents an attempt to fulfill "step 1" in the three-step plan outlined above, AB 1791 represents an attempt to fulfill "step 3". The two bills share the language of "provenance data," "personal provenance data," and "content provenance data," and are broadly compatible as a result.

Effect on open source. Software, unlike hardware, cannot be destroyed or contained. This is especially true of open-source software, which is generally free to use, edit, and otherwise manipulate. The concept of open-source is central to the tech industry's ability to innovate and expand, and the tech industry is central to California's culture and economy. But the current bill – and many other legislative efforts in California – are predicated on the notion that certain GenAI applications are inherently risky. This bill requires disclosure capabilities be built into GenAI systems. This requirement would seem antithetical to the concept of open source. How can this circle be squared?

The question of how to regulate GenAI applications without trampling on the notion of "open-source" is currently bedeviling lawmakers around the world. Take the current bill: SB 942 requires GenAI systems include disclosure capabilities. How can this be guaranteed? Legislators could specifically require that the disclosure capabilities be permanent and un-removable. But open-source requires software be freely alterable – clearly, these two ideas are not compatible. The current bill instead introduces a novel concept for maintaining disclosure capabilities in GenAI systems, even as those systems are shared and altered.

Under the provisions of this bill, any GenAI provider who licenses their system to a third party must require, as a term of the contract, that the third party not remove their system's disclosure capability. This is not a physical limitation: the licensee could, if they truly desired, remove the disclosure capability. But if a covered provider discovers that a licensee has done so, the bill requires that the provider revoke the third party's license. A licensee who continues to use a GenAI system once this license has been revoked (only in response to the removal of the system's disclosure capabilities, not for other reasons) may have a civil action brought against them for injunctive relief and reasonable costs and attorney's fees. At every step in this process, the third party is encouraged to stop using the system to create content without disclosures. On the spectrum of actions that could be taken against the users of open-source GenAI software, this

approach involves a relatively light touch. Other bills seeking to prevent GenAI systems from being used in dangerous ways might consider adopting something similar.

5) **Related legislation.**

AB 1791 (Weber, 2024) would require social media platforms to strip provenance data containing personal information from content uploaded to the platform, while retaining provenance data related to system-of-origin. The bill is pending in Senate Judiciary Committee.

AB 3050 (Low, 2024) would require CDT to issue regulations to establish standards for watermarks to be included in covered AI-generated material. The bill died in this Committee.

AB 3211 (Wicks, 2024) would require GenAI providers and the manufacturers of recording devices to include watermarking capabilities in the systems and devices they make available in California, and require social media platforms to label content with provenance info extracted from uploaded content. The bill is pending in Senate Judiciary Committee.

6) **Committee amendments.** Several committee amendments clarify the bill's disclosure requirements and update enforcement language to reflect the responsibilities of both the licensers and licensees of GenAI systems. The bill's full text is reproduced below, as proposed to be amended:

SECTION 1. Chapter 25 (commencing with Section 22757) is added to Division 8 of the Business and Professions Code, to read:

CHAPTER 25. AI TRANSPARENCY ACT

22757. This chapter shall be known as the California AI Transparency Act.

22757.1. As used in this chapter:

~~(a) "AI detection tool" means the tool required by Section 22757.2.~~

~~(b)~~

~~"Artificial intelligence" or "AI" means a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments by using machine-based inputs and human-based inputs to perceive real and virtual environments, abstract its perceptions into models through analysis in an automated manner, and use model inference to formulate options for information or action.~~

~~(c) "AI-generated content" means any form of digital content that is created with deep learning or machine learning processes.~~

~~(d) "Business" has the same meaning as defined in Section 1798.140 of the Civil Code.~~

~~(e)~~

(a) "Artificial intelligence" or "AI" means an engineered or machine-based system that varies in its level of autonomy and that can, for explicit or implicit objectives, infer from the input it receives how to generate outputs that can influence physical or virtual environments.

(b) "Covered provider" ~~is a business~~ *means a person* that ~~provides~~ *creates, codes, or otherwise produces* a generative~~AI~~ *artificial intelligence* system that ~~has, on average over the preceding 12 months,~~ *has* over 1,000,000 monthly visitors or users and is publicly accessible within the geographic boundaries of the state.

~~(f) "Department" means the Department of Technology.~~

~~(g)~~

(c) "Generative ~~AI system" refers to deep learning models that can generate text, images, and other content based on the data they were trained on.~~ *artificial intelligence" or "GenAI" means an artificial intelligence that can generate derived synthetic content, including text, images, video, and audio, that emulates the structure and characteristics of the system's training data.*

~~(h)~~

(d) "Metadata" means structural or descriptive information about ~~data, including content, format, source, rights, accuracy, provenance, frequency, periodicity, granularity, publisher or responsible party, contact information, method of collection, and other descriptions.~~ *data.*

~~(i) "Model" means a component of an information system that implements artificial intelligence technology and uses computational, statistical, or machine learning techniques to produce outputs from a given set of inputs.~~

~~(j) "Person" means a natural person located within the geographic boundaries of the state.~~

~~(k)~~

(e) "Personal information" has the same meaning as defined in Section 1798.140 of the Civil Code.

(f) "Personal provenance data" means provenance data that contains either of the following:

(1) Personal information.

(2) Unique device, system, or service information that is reasonably capable of being associated with a particular user.

(g) "Provenance data" means data that is embedded into digital content, or that is included in the digital content's metadata, for the purpose of verifying the digital content's authenticity, origin, or history of modification.

(h) "System provenance data" means provenance data that is not reasonably capable of being associated with a particular user and that contains either of the following:

(1) Information regarding the type of device, system, or service that was used to generate a piece of digital content.

(2) Information that provides proof of content authenticity.

22757.2. (a) A covered provider shall create *and make freely available* an AI detection tool ~~by which a person can query the covered provider as to the extent to which text, image, video, audio, or multimedia content was created, in whole or in part, by a generative AI system provided by the covered provider that meets all of the following criteria:~~ *that meets all of the following criteria:*

(1) The tool allows a user to assess whether image, video, audio, or other digital content was created or altered by the covered provider's GenAI system.

(2) The tool allows a user to determine which parts of the content were created or altered by the covered provider's GenAI system.

(3) The tool outputs any system provenance data that is detected in the content.

(4) The tool does not output any personal provenance data that is detected in the content.

~~(1)~~

(5) ~~The AI detection tool shall be~~ *The tool is* publicly accessible ~~and available via a uniform resource locator (URL) on~~ *through* the covered provider's internet website and ~~through~~ its mobile application, as applicable.

~~(2)~~

(6) ~~The AI detection tool shall allow~~ *The tool allows* a ~~person~~ *user* to upload content or *provide* a ~~URL.~~ *uniform resource locator (URL) linking to online content.*

~~(3)~~

(7) ~~The AI detection tool shall support~~ *The tool supports* an application programming interface *(API)* that allows a ~~person~~ *user* to invoke the AI detection tool without visiting the covered provider's internet website.

~~(4)~~

(b) ~~The AI detection tool~~ *A covered provider* shall ~~allow a person to provide~~ *collect user* feedback ~~if the person believes the AI detection tool is not properly identifying content that was created by the covered provider.~~ *related to the efficacy of the covered provider's AI detection tool and incorporate that feedback into any attempt to improve the efficacy of the tool.*

~~(b)~~

(c) ~~In complying with this section, a~~ *A* covered provider shall not do any of the following:

~~(1) Reveal personal information that identifies who utilized the covered provider's generative AI system to create AI-generated content that was submitted to the covered provider's AI detection tool.~~

~~(2)~~

(1) (A) ~~Subject to~~ *Except as provided in* subparagraph (B), collect ~~and~~ *or* retain personal information ~~when a person utilizes~~ *from users of* the covered provider's AI detection tool.

(B) *(i)* A covered provider may collect and retain the contact information of a ~~person~~ *user* who ~~submitted~~ *submits* feedback pursuant to ~~paragraph (4) of subdivision (a).~~ *subdivision (b) if the user opts in to being contacted by the covered provider.*

(ii) User information collected pursuant to clause (i) shall be used only to evaluate and improve the efficacy the covered provider's AI detection tool.

~~(3)~~

(2) Retain any content submitted to the AI detection tool for longer than is necessary to comply with this section.

(3) Retain any personal provenance data from content submitted to the AI detection tool by a user.

22757.3. (a) A covered provider shall *offer the user an option to* include ~~in AI-generated~~ *a manifest disclosure in* image, ~~text,~~ video, *audio,* or ~~multimedia~~ *other digital* content created *or altered* by ~~a generative AI~~ *the covered provider's GenAI* system ~~it provides a visible disclosure~~ that meets all of the following criteria:

~~(1) The disclosure shall include a clear and conspicuous notice, as appropriate for the medium of the content, that identifies the content as generated by AI, such that the disclosure is not avoidable, is understandable to a reasonable person, and is not contradicted, mitigated by, or inconsistent with anything else in the communication.~~

(1) The disclosure identifies content as AI-generated content.

(2) The disclosure is clear, conspicuous, appropriate for the medium of the content, and understandable to a reasonable person.

~~(2)~~

(3) The disclosure ~~shall, to the extent technically feasible,~~ *shall* be permanent or extraordinarily difficult to ~~remove.~~ *remove, to the extent it is technically feasible.*

~~(3) The output's metadata information shall include an identification of the content as being generated by AI, the identity of the tool used to create the content, and the date and time the content was created.~~

~~(b) A covered provider shall include in AI-generated image, audio, video, or multimedia content created by a generative AI system an imperceptible disclosure that is machine detectable and is, to the extent technically feasible, permanent or difficult to remove.~~

~~(c) A covered provider shall implement reasonable procedures to prevent downstream use of a generative AI system it provides without the disclosure required by this section, including by doing both of the following:~~

~~(1) Requiring by contract that third-party licensees of the generative AI system refrain from removing a required disclosure.~~

~~(2) Terminating access to the generative AI system when the covered provider has reason to believe that a third-party licensee has removed a required disclosure.~~

~~(d) At least once every two years, the department shall review this section and make recommendations to the Legislature regarding any amendments needed to account for changing technology and standards.~~

(b) A covered provider shall include a latent disclosure in AI-generated image, video, audio, or other digital content created by the covered provider's GenAI system that meets all of the following criteria:

(1) The disclosure conveys all of the following information, either directly or through a link to a permanent internet website:

(A) The name of the covered provider.

(B) The name and version number of the GenAI system that created or altered the content.

(C) The time and date of the content's creation or alteration.

(D) Which parts of the content were created or altered by the GenAI system.

(E) A unique identifier.

(2) The disclosure is detectable by the covered provider's AI detection tool.

(3) The disclosure is consistent with widely accepted industry standards.

(4) The disclosure is permanent or extraordinarily difficult to remove, to the extent it is technically feasible.

(c) (1) If a covered provider licenses its GenAI system to a third party, the covered provider shall require by contract that the licensee maintain the system's capability to include a disclosure required by subdivision (b) in content the system creates or alters.

(2) If a covered provider knows that a third-party licensee modified a licensed GenAI system such that it is no longer capable of including a disclosure required by subdivision (b) in content the system creates or alters, the covered provider shall revoke the license within 72 hours of discovering the licensee's action.

(3) A third-party licensee shall cease using a licensed GenAI system after the license for the system has been revoked by the covered provider pursuant to paragraph (2).

22757.4. (a) A covered provider that violates this chapter shall be liable for a civil penalty in the amount of five thousand dollars ($5,000) per violation to be collected in a civil action filed only by the Attorney General.

(b) Each day that a covered provider is in violation of this chapter shall be deemed a discrete violation.

(c) For a violation by a third-party licensee of paragraph (3) of subdivision (c) of Section 22757.3, the Attorney General, a district attorney, a county counsel, or a city attorney may bring a civil action for both of the following:

(1) Injunctive relief.

(2) Reasonable attorney's fees and costs.

ARGUMENTS IN SUPPORT:

Taking a "support if amended" position on a previous version of the bill Oakland Privacy previously explains the need for disclosure of AI-generated content:

> We consider generative AI disclosures to be akin to these measures and to be reasonable measures to protect the integrity of public discussions. A buyer beware absolutism threatens to poison all public discourse by causing massive confusion about what is authentic and what is not. It is likely the higher the stakes in a public debate, the more generative AI content will be created to try to push public opinion in one direction or the other. For the sake of the general welfare, disclosure will at least provide a modicum of accuracy to various kinds of difficult and controversial subjects without banning or removing content.

ARGUMENTS IN OPPOSITION:

A coalition of industry associations including Technet, NetChoice, Computer & Communications Industry Association, and CalChamber writes in opposition, describing their desire for a "federal standard" to be adopted:

> 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.

REGISTERED SUPPORT / OPPOSITION:

Support

Protection of The Educational Rights of Kids - Advocacy (PERK Advocacy)
Transparency Coalition.ai

Support If Amended

Oakland Privacy

Opposition

California Chamber of Commerce
Computer and Communications Industry Association
Netchoice
Technet

Analysis Prepared by: Slater Sharp / P. & C.P. / (916) 319-2200