Date of Hearing: April 16, 2026
Fiscal: Yes

ASSEMBLY COMMITTEE ON PRIVACY AND CONSUMER PROTECTION
Rebecca Bauer-Kahan, Chair
AB 2713 (Wicks) – As Amended April 6, 2026

PROPOSED AMENDMENTS

SUBJECT: California AI Transparency Act: system provenance data: inspection

SYNOPSIS

Over the past three years, as generative artificial intelligence (GenAI) technologies have become more realistic and accessible, online content that appears genuine, but that is actually false, has flooded social media and other large online platforms. This unmitigated spread of synthetic content threatens to harm Californians in numerous ways, such as through the proliferation of nonconsensual deepfake pornography, scams, and the distribution of targeted political disinformation.

Over the past two years, California has sought to tackle the issue of GenAI-produced online content through a three-step plan: label all GenAI content "fake," label a significant portion of real content "real," and prominently display these labels online. Beginning this August, SB 942 (Becker, 2024) requires large GenAI platforms to include machine-detectable "latent disclosures" in their outputs. Beginning in 2028, AB 853 (Wicks, 2025) requires "capture devices" such as cameras to include provenance data in their outputs by default. And beginning in 2027, AB 853 (Wicks, 2025) requires large online platforms to prominently display provenance data detected in user-uploaded content.

According to the author of this bill, the inclusion of language in AB 853 permitting users to download content from online platforms has led copyright owners to privately express piracy concerns. This bill attempts to address that concern by clarifying that the download of content by users is "subject to any applicable federal laws." Committee amendments described in Comment #4 go further, replacing the controversial provision with language permitting users to download provenance data attached to content rather than the content itself. Committee amendments additionally prohibit large online platforms from stripping provenance data out of user-downloaded content and make various non-substantive drafting changes.

This bill has no formal support or opposition.

EXISTING LAW:

1) Defines the following terms:

    a. "Artificial intelligence" or "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.

b. "Capture device" to mean a device that can record photographs, audio, or video content, including, but not limited to, video and still photography cameras, mobile phones with built-in cameras or microphones, and voice recorders.

c. "Covered provider" to mean a person that creates, codes, or otherwise produces a generative artificial intelligence system that has over 1,000,000 monthly visitors or users and is publicly accessible within the geographic boundaries of the state.

d. "Digital signature" to mean a cryptography-based method that identifies the user or entity that attests to the information provided in the signed section.

e. "Generative artificial intelligence system" or "GenAI system" to mean 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.

f. "Large online platform" to mean a public-facing social media platform, file-sharing platform, mass messaging platform, or stand-alone search engine that distributes content to users who did not create or collaborate in creating the content that exceeded 2,000,000 unique monthly users during the preceding 12 months. "Large online platform" does not include either a broadband internet access service, as defined in Section 3100 of the Civil Code, or a telecommunications service, as defined in Section 153 of Title 47 of the United States Code.

g. "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.

h. "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 related to content authenticity. (Bus. & Prof. Code § 22757.1.)

2) Beginning August 2, 2026, requires a covered provider to include a latent disclosure in AI-generated image, video, or audio content, or content that is any combination thereof, created by the covered provider's GenAI system that, to the extent technically feasible and reasonable, conveys the name of the covered provider, the name and version number of the GenAI system that created or altered the content, the time and date of the content's creation or alteration, and a unique identifier, either directly or through a link to a permanent internet website. (Bus. & Prof. Code § 22757.3.)

3) Beginning January 1, 2028, requires a capture device manufacturer to, with respect to any capture device the capture device manufacturer first produced for sale in the state on or after January 1, 2028, provide a user with the option to include a latent disclosure in content captured by the capture device, and to embed latent disclosures in content captured by the device by default.

4) Beginning January 1, 2027, requires a large online platform to allow a user to inspect all available system provenance data embedded into or attached to content through any of the following means:

a. Directly through a user interface provided by the large online platform.

b. By allowing the user to download a version of the content with its attached system provenance data.

c. By providing a link to an internet website or another application displaying the content's system provenance data. (Bus. & Prof. Code § 22757.3.1)

5) Beginning January 1, 2027, prohibits a large online platform from knowingly stripping any system provenance data or digital signature from content uploaded or distributed on the large online platform, to the extent technically feasible, provided the provenance data or signature are compliant with widely adopted specifications adopted by an established standards-setting body. (Bus. & Prof. Code § 22757.1.)

THIS BILL:

1) Clarifies a platform may allow a user to download a version of content hosted on the platform along with any attached system provenance data "subject to any applicable federal copyright laws."

COMMENTS:

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

New and emerging developments of generative AI (GenAI) tools have made it easier to create, edit, and doctor images, video, and audio. GenAI technologies can create and manipulate content to look realistic and convincing, which allow bad actors to create harmful content and spread disinformation. With the passage of AB 853, there would be more transparency of AI-generated content in the digital information ecosystem, and users would have more information to understand the source of content.

AB 2713 would provide additional clarification within existing law in order that any requirements to share provenance information on large online platforms would have to be operationalized in a way that protects content creators and allow for large online platforms to provide meaningful information for users on their platforms.

2) **Background.** *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.

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 traditional computer functions, AI can accomplish tasks that are normally performed by humans. AI that is trained on small, specific datasets in order to make recommendations and predictions is sometimes referred to as "predictive AI." This differentiates it from GenAI, which is trained on massive datasets in order to produce detailed text and images. When Netflix suggests a TV show

to a viewer, that 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.

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.

Deepfake pornography. Since its inception, GenAI has been used to subject women and girls to image-based sexual abuse. While high-profile celebrities were most often targeted when this technology was first developed,[1] open-source GenAI models have been exploited to make this technology more accessible and affordable. This has led to a proliferation of websites and phone-based apps that offer user-friendly interfaces for uploading clothed images of real people to generate photorealistic nude images of not only adults, but also children. According to a *New York Times* article:

> Boys in several states have used widely available "nudification" apps to pervert real, identifiable photos of their clothed female classmates, shown attending events like school proms, into graphic, convincing-looking images of the girls with exposed A.I.-generated breasts and genitalia. In some cases, boys shared the faked images in the school lunchroom, on the school bus or through group chats on platforms like Snapchat and Instagram, according to school and police reports.[2]

In February 2024, deepfake nude images of 16 eighth-grade students were circulated among students at a California middle school.[3] Similar reports of abuse have been reported across the country and show no sign of abating.[4] In the first six months of 2024, nudification sites had been

---

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

[2] Natasha Singer, "Teen Girls Confront an Epidemic of Deepfake Nudes in Schools", *The New York Times (*Apr. 8, 2024), https://www.nytimes.com/2024/04/08/technology/deepfake-ai-nudes-westfield-high-school.html.

[3] 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) *accessed at* https://www.dailymail.co.uk/news/article-13295475/Inappropriate-images-California-Fairfax-High-School-AI-deepfake.html.

[4] Tim McNicholas, "New Jersey high school students accused of making AI-generated pornographic images of classmates," *CBS News* (Nov. 2, 2023), https://www.cbsnews.com/newyork/news/westfield-high-school-ai-pornographic-images-students/; Lauraine Langreo, "Students Are Sharing Sexually Explicit 'Deepfakes.' Are Schools Prepared?" *Ed Week* (Sept. 26, 2024), https://www.edweek.org/leadership/studentsare-sharing-sexually-explicit-deepfakes-are-schools-prepared/2024/09; Gabrielle Hunt and Daryl Higgens "AI nudes of Victorian students were allegedly shared online. How can schools and parents respond to deepfake porn?," *The Guardian* (June, 12, 2024), https://www.theguardian.com/australia-news/article/2024/jun/12/ai-nudes-of-victorian-students-were-allegedly-shared-online-how-canschools-and-parents-respond-to-deepfake-porn.

visited over 200 million times.[5] Meanwhile, a 2024 study from Center on Democracy and Technology reports that 40% of students were aware of deepfakes being shared at school, 15% of which depicted an individual in a sexually explicit or intimate manner. In over 60% of these cases, the images were distributed via social media.[6]

Scams. GenAI-powered speech and video is driving a new era in scamming. These AI tools are often trained on publicly available data, and as a result, 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 internet.[7] 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.[8]

In December 2024, the FBI issued a public service announcement warning about the potential dangers posed by GenAI.[9] Among the concerns highlighted was the technology's ability to significantly lower the barrier for producing counterfeit documents, such as fake IDs, passports, and other fraudulent government-issued identification, which could greatly facilitate identity theft. Additionally, GenAI enables the creation of entirely fictitious profiles on social media and dating platforms, which can be used to exploit individuals both financially and emotionally.

Elections. Deepfake technology is being used around the world to spread disinformation and propaganda. 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

---

[5] *People of the State of California v. Sol Ecom, Inc, et al.* (2024) Case No. CGC-24-617237, p. 2, https://www.sfcityattorney.org/wp-content/uploads/2024/08/2024-08-16-First-Amended-Complaint_Redacted.pdf
[6] Elizabeth Laird, Maddy Dwyer and Kristin Woelfel, "In Deep Trouble: Surfacing Tech-Powered Sexual Harassment in K-12 Schools," *Center for Democracy & Technology* (Sept. 26, 2024), https://cdt.org/wp-content/uploads/2024/09/FINAL-UPDATED-CDT-2024-NCII-Polling-Slide-Deck. Pdf.
[7] 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.
[8] 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.
[9] Department of Justice Federal Bureau of Investigations, "Criminals Use Generative Artificial Intelligence to Facilitate Financial Fraud" (Dec. 3, 2024) , https://www.ic3.gov/PSA/2024/PSA241203.

interests, was defeated in September by an opponent who supported closer ties to Moscow and Russian President Vladimir Putin.[10]

Similar deepfakes surfaced in the United States ahead of the 2024 presidential election. In July 2024, Elon Musk shared a video featuring an AI-generated voice clone of then-Vice President Kamala Harris, in which the fabricated voice claimed she was a "diversity hire" due to being a woman of color and that she "did not know the first thing about running a country."[11] Although Musk admitted two days after posting that the video was intended as satire, the potential impact of such content on political campaigns remains a serious concern. In 2023, the Legislature enacted a trio of bills aimed at addressing elections deepfakes: AB 2655 (Berman, Stats. 2024, Ch. 261), which imposes removal and disclosure obligations on large online platforms during the four months leading up to an election; AB 2839 (Pellerin, Stats. 2024, Ch. 262), which prohibits the distribution of election materials containing certain types of deceptively altered digital content; and AB 2355 (Carillo, Stats. 2024, Ch. 260), which requires AI-altered campaign materials to include clear disclosures. The first two bills, however, have been blocked from enforcement in legal proceedings.

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

---

[10] 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.

[11] Ken Bensinger, "Elon Musk Shares Manipulated Harris Video, in Seeming Violation of X's Policies", *The New York Times* (July 27, 2024), https://www.nytimes.com/2024/07/27/us/politics/elon-musk-kamala-harris-deepfake.html

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.

SB 942 and AB 853. Seeking to enact a comprehensive content provenance framework to combat the rapid online proliferation of AI-generated content, the Legislature passed SB 942 (Becker, Stats. 2024, Ch. 291). The law applies to "covered providers" – developers of publicly-accessible GenAI systems with over one million monthly users – requiring them to provide a free, publicly accessible tool that allows users to detect whether audio, image, or video content was generated or altered by their systems. The bill permits visible disclosures and requires latent, machine-detectable provenance data to be embedded into content generated by these provides' GenAI systems.

While SB 942 focused primarily on developers of large GenAI systems, AB 853 (Wicks, Stats. 2025, Ch. 674) imposed disclosure requirements on large online platforms and manufacturers of devices capable of capturing digital content (e.g. digital cameras.) Together, these bills effectively require GenAI developers to disclose that content generated by their systems is "fake" beginning in mid-2026, recording device manufacturers to disclose that content captured by their devices is "real" beginning in early 2028, and large online platforms to prominently disclose this provenance data to users beginning in early 2027.

3) **What this bill would do.** AB 853 provided large online platforms with three options for disclosing content provenance data to users of the platform: (1) directly disclosing the data through a user interface created by the platform, (2) providing a link to an external website displaying the content, and (3) permitting users to download any content with its provenance data attached. According to the author, various groups representing copyright owners have expressed concern with this third mechanism: they claim that enabling users to download content directly from social media platforms will facilitate online piracy. To address this concern, AB 2713 clarifies that users may download a version of content with its attached provenance data "subject to any applicable federal copyright laws."

4) **Committee amendments.** To more directly address the piracy concern raised by copyright owners, the author has agreed to amendments that do all of the following:

- Allow a user to download content provenance directly from a large online platform, in a format that cannot be easily integrated into unrelated content, instead of downloading content with provenance data attached.
- Prohibit a large online platform from knowingly stripping system provenance data or a digital signature from content a user downloads from the large online platform.
- Make various non-substantive drafting changes.

The full text of the bill as proposed to be amended follows:

22757.3.1. (a) A large online platform shall do all of the following:

(1) Detect whether any provenance data ~~that is compliant with widely adopted specifications adopted by an established standards-setting body~~ is embedded into ~~or~~***,*** attached to***, or otherwise associated with*** content distributed on the large online platform.

(2) (A) Provide a user interface to ~~disclose the availability of system provenance data~~ that reliably indicates ~~that the~~ ***whether*** content was generated or substantially altered by a GenAI system or captured by a capture device.

(B) The user interface required by this paragraph shall make clearly and conspicuously available to users information sufficient to identify the content's authenticity, origin, or history of modification, including, but not limited to, all of the following:

(i) Whether provenance data is ~~available~~ ***embedded into, attached to, or otherwise associated with the content***.

(ii) The name of the GenAI system or capture device that created or substantially altered the content, if applicable.

(iii) Whether any digital signatures are ~~available~~ ***embedded into, attached to, or otherwise associated with the content***.

(3) Allow a user to inspect ~~all available~~ ***any*** system provenance data ~~that is compliant with widely adopted specifications adopted by an established standards-setting body~~ ***embedded into, attached to, or otherwise associated with content*** in an easily accessible manner. The large online platform may satisfy this requirement by any of the following means:

(A) ~~Allow the user to inspect~~ ***Displaying*** the system provenance data directly through the large online platform's user interface pursuant to paragraph (2).

(B) Providing a link to an internet website or other application displaying the system provenance data, including a website or application operated by a third party.

(~~B~~***C***) ~~Allow the~~ ***Allowing a*** user to download ~~a version of the content with its attached~~ ***any*** system provenance data ***embedded into, attached to, or otherwise associated with the content***, subject to any applicable federal copyright laws***, in a format that cannot be easily embedded into, attached to, or associated with unrelated content***.~~(C) Provide a link to the content's system provenance data displayed on an internet website or in another application provided either by the large online platform or a third party.~~

(b) A large online platform shall not, to the extent technically feasible, knowingly strip any system provenance data or digital signature ~~that is compliant with widely adopted specifications adopted by an established standards-setting body from content uploaded or distributed on~~ ***uploaded to, distributed on, or downloaded from*** the large online platform.

(c) This section does not require a large online platform to take any action with respect to provenance data, system provenance data, or digital signatures that are not compliant with widely adopted specifications issued by an established standards-setting body.

(~~c~~***d***) This section shall become operative on January 1, 2027.

REGISTERED SUPPORT / OPPOSITION:

Support

None on file.

Opposition

None on file.

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