Date of Hearing: April 22, 2025
Fiscal: Yes

ASSEMBLY COMMITTEE ON PRIVACY AND CONSUMER PROTECTION
Rebecca Bauer-Kahan, Chair
AB 853 (Wicks) – As Amended March 28, 2025

SUBJECT: California AI Transparency Act

SYNOPSIS

As generative artificial intelligence (GenAI) becomes more accessible, online content that appears real, but that 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, reputational harms, and the distribution of targeted election disinformation. One strategy to combat this is to embed within content generated by AI and captured in real life provenance data that enables viewers to authenticate content.

Last session, SB 942 (Becker, Stats. 2024, Ch. 291) required GenAI developers to ensure that content created using their tools must include provenance data that can be readily accessed through AI detection tools. That legislation was a crucial first step towards creating a regime in which all content can be verified for its authentication.

This bill, sponsored by The California Initiative for Technology & Democracy (CITED), represents the next key step. The bill requires large online platforms to develop a way for users to easily access provenance data of uploaded content. The bill would also require capture device manufacturers to include features on their products that enable users to include provenance data in the content that they capture. These requirements, coupled with SB 942, would create a comprehensive disclosure and detection framework that would enable the large-scale classification of content as either authentic or artificial.

This bill is opposed by TechNet and the Computer and Communications Industry Association (CCIA), who argue that the content provenance and watermarking are immature technologies that should not be subjected to prescriptive regulation, and that the bill as written is ambiguous in certain respects.

If passed by this Committee, this bill will next be heard by the Assembly Judiciary Committee

THIS BILL:

1) Defines the following terms:

    a. "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.

    b. "Capture device manufacturer" to mean a person who produces a capture device for sale in the state.

c. "Large online platform" to mean a public-facing social media platform, content-sharing platform, messaging platform, advertising network, or standalone 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.

2) Requires that a large online platform do the following:

   a. Retain any available provenance data in content provided to, or posted on, the large online platform.

   b. Make available to a consumer of content on the large online platform either of the following:

      i. The provenance data.

      ii. A conspicuous indicator that provenance data is available.

3) Require that a capture device manufacturer, with respect to any capture device the capture device manufacturer produces for sale in the state, do the following:

   a. Include in the capture device's default capture app the ability for a user to enable the inclusion of provenance data in the user's captured content.

   b. Ensure secure hardware-based provenance capture is available to third-party applications.

4) Establishes that large online platforms and capture device manufacturers can be held liable for violations of this bill.

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 "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.

4) 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.)

5) 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. "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.

c. "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.

d. "Latent" to mean present but not manifest.

e. "Manifest" to mean easily perceived, understood, or recognized by a natural person.

f. "Metadata" to mean structural or descriptive information about data.

g. "Personal information" to have the same meaning as defined in Section 1798.140 of the Civil Code.

h. "Personal provenance data" to mean provenance data that contains either of the following:

    i. Personal information.

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

i. "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.

j. "System provenance data" to mean provenance data that is not reasonably capable of being associated with a particular user and that contains either of the following:

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

    ii. Information related to content authenticity. (Bus. & Prof. Code § 22757.1.)

6) Requires that a covered provider make available an AI detection tool at no cost to the user that meets all of the following criteria:

    a. The tool allows a user to assess whether image, video, or audio content, or content that is any combination thereof, was created or altered by the covered provider's GenAI system.

    b. The tool outputs any system provenance data that is detected in the content.

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

d. The tool is publicly accessible. A covered provider may impose reasonable limitations on access to the tool to prevent, or respond to, demonstrable risks to the security or integrity of its GenAI system.

e. The tool allows a user to upload content or provide a uniform resource locator (URL) linking to online content.

f. The tool supports an application programming interface that allows a user to invoke the tool without visiting the covered provider's internet website. (Bus. & Prof. Code [hereafter, BPC] § 22757.2(a).)

7) Requires that a covered provider collect user feedback related to the efficacy of the covered provider's AI detection tool and incorporate relevant feedback into any attempt to improve the efficacy of the tool. (BPC § 22757.2(b).)

8) Prohibits a covered provider from do any of the following:

a. Except as provided with some exceptions collect or retain personal information from users of the covered provider's AI detection tool.

b. A covered provider may collect and retain the contact information of a user who submits feedback if the user opts in to being contacted by the covered provider.

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

d. Retain any personal provenance data from content submitted to the AI detection tool by a user. (BPC § 22757.2(c).)

9) Requires that a covered provider offer the user the option to include a manifest disclosure in image, video, or audio content, or content that is any combination thereof, created or altered by the covered provider's GenAI system that meets all of the following criteria:

a. The disclosure identifies content as AI-generated content.

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

c. The disclosure is permanent or extraordinarily difficult to remove, to the extent it is technically feasible. (BPC § 22757.3(a).)

10) Requires that a covered provider 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 meets all of the following criteria:

a. To the extent that it is technically feasible and reasonable, the disclosure conveys all of the following information, either directly or through a link to a permanent internet website:

i. The name of the covered provider.

ii. The name and version number of the GenAI system that created or altered the content.

iii. The time and date of the content's creation or alteration.

iv. A unique identifier.

v. The disclosure is detectable by the covered provider's AI detection tool.

b. The disclosure is consistent with widely accepted industry standards.

c. The disclosure is permanent or extraordinarily difficult to remove, to the extent it is technically feasible. (BPC § 22757.3(b).)

11) Requires that 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 in content the system creates or alters. (BPC § 22757.3(c).)

12) Establishes that 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 required disclosure in content the system creates or alters, the covered provider must revoke the license within 96 hours of discovering the licensee's action. (BPC § 22757.3(c).)

13) Establishes that a third-party licensee shall cease using a licensed GenAI system after the license for the system has been revoked by the covered provider. (BPC § 22757.3(c).)

14) Establishes a civil penalty for covered providers who do not abide requirements for latent and manifest disclosures on GenAI content. (BPC § 22757.4.)

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.
>
> AB 853 will help provide more transparency of AI-generated content in the digital information ecosystem and would provide more information to understand the source of content and discern what is real and what is inauthentic. This bill will help mitigate some of the harmful impacts of AI-generated content.

2) **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.

3) **Ctrl+Alt+Deceive: 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***.** The creation of text, 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 invade privacy and disrupt the lives of Californians. Since its inception, GenAI has been used to create nonconsensual pornography, more accurately referred to by sexual assault experts as image-based sexual abuse, almost entirely against women and girls.

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:

---

[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/.

> 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 abuses, almost always against girls, have been reported across the country and show no sign of abating.[4] In the first six months of 2024, these sites had been 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] This provides a potent means of amplifying deepfake nonconsensual pornography, extending the content's reach by, in effect, and crowdsourcing abuse, potentially reaching thousands or even millions of viewers.

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 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

---

[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.

[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.

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 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.

Given the high risk that generative AI poses in spreading campaign disinformation, 17 states have enacted laws addressing the use of deepfakes online, and at least 40 states have considered legislative action.[12] Last year 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

---

[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.

[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

[12] National Conference of State Legislatures, "Deceptive Audio or Visual Media ('Deepfakes') 2024 Legislation" (Nov. 22, 2024), https://www.ncsl.org/technology-and-communication/deceptive-audio-or-visual-media-deepfakes-2024-legislation

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. Notably, the first two bills are currently facing a legal challenge.[13]

4) **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.

5) **What this bill would do.** The challenge of content authentication could theoretically be solved with three steps:

1. Require that all GenAI-derived content be labeled as "fake."

2. Require all content produced by recording devices be labeled as "real."

3. Require social media platforms to clearly present these labels.

Last year, SB 942 (Becker) was chaptered to address concerns around labeling GenAI-produced content as "fake." The bill requires developers of GenAI systems with over one million users to embed latent disclosures within content generated using their systems. Additionally, those

---

[13] Tori Guidry, "CALIFORNIA AI LAW HIT WITH A CONSTITUTIONAL CHALLENGE: X Corp. Attempts To Take Down California's Law Regulating AI-Generated Political Content", *The National Law Review* (Nov. 20, 2024), https://natlawreview.com/article/california-ai-law-hit-constitutional-challenge-x-corp-attempts-take-down

developers must provide a publicly accessible and free AI detection tool capable of identifying the embedded disclosures. While SB 942 is prescriptive regarding the type of provenance data that must be detectable by such tools, it allows the industry to establish best practices for ensuring GenAI-produced content can be reliably identified.

This bill incorporates elements of AB 3211 (Wicks, 2024) to establish a more comprehensive framework for authenticating content. Specifically, it builds upon the foundation set by SB 942 by requiring manufacturers of recording devices to provide users with a tool to embed provenance data into images and audio captured by those devices. Additionally, it mandates that large online platforms display the provenance of shared audio and visual content. Together, these measures address the second and third key points in the effort to enhance transparency and trust in digital media.

This bill would require manufacturers of capture devices, such as cameras, smartphones with cameras, scanners, audio recorders, and other devices capable of storing and transferring digital media, to provide users with the ability to embed provenance data in the content they capture. This approach offers a particularly effective means of content authentication. Unlike GenAI tools, which can produce limitless amounts of synthetic content and are increasingly accessible due to open-source code, capture devices are produced by a limited number of manufacturers. By focusing compliance efforts on these manufacturers, enforcement becomes more feasible. Additionally, the volume of content generated by capture devices is significantly lower than that produced by GenAI systems, further increasing the likelihood that captured content can be reliably authenticated.

This bill would also require large online platforms, such as Instagram and X, to provide users with a readily accessible method for inspecting the provenance data of content shared on their platforms. Given that most individuals engage with digital content through these platforms, it is both practical and impactful to place a duty on them to help users determine the authenticity of the content they encounter. Under current law, the responsibility falls on the viewer to seek out and use AI detection tools to verify content. This bill would shift that burden, establishing a more uniform and accessible framework for identifying content provenance directly within the platforms themselves.

6) **Technical Feasibility.** The policy goals of this bill would complete a framework for large-scale detection of content provenance online. However, this raises important questions about whether the necessary technology currently exists to support such goals. The Coalition for Content Provenance and Authenticity (C2PA), a consortium of eleven industry organizations, including Microsoft, Adobe, and the BBC, is actively working to develop standards and best practices for content provenance.[14] C2PA collaborates closely with the Content Authenticity Initiative (CAI), a broader coalition of over 4,000 civil society, media, and technology organizations.[15] The CAI publishes open-source tools that enable the embedding of provenance data across various content types, helping to lay the groundwork for the technical infrastructure needed to support this bill's aims.[16]

---

[14] Information about the C2PA can be found at https://c2pa.org/
[15] Information about CAI can be found at https://contentauthenticity.org/
[16] Open source code for embedding disclosures in content from CAI can be afound at https://opensource.contentauthenticity.org/docs/sb-algs/

Currently, the CAI notes that provenance disclosures embedded using their open-source code can be stripped from content when shared on platforms that do not support compatible software. However, this issue could be addressed by supplementing disclosures with invisible watermarks or digital fingerprints that link to an external database where users can access the associated provenance data. Meta, which operates Facebook, Instagram, and WhatsApp and is a member of the C2PA, already utilizes the C2PA standard to label AI-generated content.[17] However, users are still unable to access the specific provenance data. Given that the company already uses this metadata internally, it would likely be feasible to embed links or tools directly within the content, allowing users to easily inspect the same provenance information.

Regarding capture devices, both Nikon and Leica have developed cameras capable of embedding provenance data directly into captured images.[18] Though, as noted by TechNet and CCIA in their opposition letter, only a few cameras have these capabilities and no known video camera has been created to incorporate these technologies. The CAI does, however, provides open-source code for embedding provenance data across a range of file types, including images, videos, and audio. The availability of open code alongside this bill should help create the industry impetus to incorporate disclosures into their devices. Nevertheless, no smartphones on the market offer the ability to embed provenance data into captured content. However, Samsung recently announced that its upcoming phone release will include a tool for accessing provenance data on content encountered while using the device.[19] Despite the current absence of this feature in smartphones, the successful implementation in cameras indicates that capture hardware can indeed be designed to support the embedding of provenance data.

7) **Policy Considerations.** Going forward, the author may wish to consider clarifying the specific types of provenance information that should be embedded in captured content. Under SB 942, a prescriptive framework was established for GenAI-generated content, requiring inclusion of details such as the name of the GenAI developer, the system used to generate the content, the date of creation, and a unique content identifier. A similar standardized approach could be applied to capture devices to ensure consistency in provenance data. This might include information such as the manufacturer and model of the capture device, the date and time of content creation, and a unique identifier linked to the captured file.

As noted above, the incorporation of provenance data into content captured on cameras and in phones is a relatively nascent technology. The author may wish to consider including a delayed operative date for this requirement to give industry a buffer to incorporate this technology into their devices.

Additionally, the author may wish to examine what provenance data must be retained by large online platforms. As currently drafted, the bill would require platforms to retain any available provenance data associated with content uploaded or shared on their services. However, this could potentially include personal provenance data, raising privacy concerns. The language

---

[17] Monika Bickert, "Our Approach to Labeling AI-Generated Content and Manipulated Media", *Meta* (Sept. 12, 2024), https://about.fb.com/news/2024/04/metas-approach-to-labeling-ai-generated-content-and-manipulated-media/
[18]Jaron Schneider, "Nikon Will Add C2PA Content Credentials to the Z6 III by Next Year", *PetaPixel* (Oct. 14, 2024), https://petapixel.com/2024/10/14/nikon-will-add-c2pa-content-credentials-to-the-z6-iii-by-next-year/
[19] Devesh Beri, "A New Feature in Samsung Galaxy S25 Phones Helps Spot AI-Edited Photos", *Yahoo Tech (*Feb. 6, 2025), https://tech.yahoo.com/phones/articles/feature-samsung-galaxy-s25-phones-130000863.html

could be refined to require platforms to retain only system provenance data, such as device type, AI system, or creation timestamp that cannot be easily linked to individual content creators.

At the same time, the author may also consider whether there should be an option for including personal provenance data. Photojournalists, professional photographers, and videographers, for instance, may wish to embed personal information to assert authorship and demonstrate ownership of their work. Balancing the need for privacy with the desire for proper attribution will be crucial in shaping fair and effective provenance policies.

ARGUMENTS IN SUPPORT: California Initiative for Technology & Democracy (CITED) co-sponsors of the bill, write in support:

> Last year, the California Legislature passed SB 942, the AI Transparency Act, which created the first-in-the-nation rules requiring generative AI providers to implement content provenance for AI-generated content. When this law takes effect in 2026, the public will be able to use AI detection tools to identify the source of AI-generated content. SB 942 represents an important foundation in our effort to rebuild trust in our information ecosystem.
>
> But more must be done to stem the tide of mis- and dis-information in the age of AI. AB 853 builds upon the framework of the AI Transparency Act by adding several critical interventions, first at the point of content creation and then at the point of dissemination.
>
> At the point of content creation, AB 853 would enable human-created authentic content to be differentiated from AI-generated synthetic content by requiring cameras and recording devices sold in California to include an option to place provenance information on the content that the device produces. This provenance information, together with existing provenance requirements for generative AI under the AI Transparency Act, would allow the public to easily differentiate between human vs. AI-generated content.
>
> Thereafter, at the point of content dissemination, AB 853 would require social media and other online platforms to display the source of the content shared on their platforms by leveraging the underlying provenance data. By requiring clear, factual labeling of the source of online content, AB 853 would equip the public with a tool to make their own judgment about what information they deem to be trustworthy.
>
> With the rapid proliferation of GenAI tools, the public must be equipped with the necessary tools to distinguish the content we see online in order to restore trust in our democracy and our society. For these reasons, CITED is proud to sponsor and support AB 853.

ARGUMENTS IN OPPOSITION: In opposition to the bill, TechNet and the Computer and Communications Industry Association argue:

> Manufacturers of capture devices and large online platforms are actively leading the development of industry-wide standards for provenance and watermarking technology. These technologies remain emergent and complex, and the relevant standards are still being refined by the organizations best positioned to ensure their success.
>
> For camera and recording device manufacturers, AB 853 imposes technically infeasible and commercially impracticable requirements regarding any capture device. The state of the technology has not yet matured to enshrine this requirement in law at this time. There are

only a handful of devices on the market today that have the capability to include provenance information by default, and it is unclear if there are any video cameras on the market with these capabilities. We would suggest a standard that is more permissive by phasing in these requirements over time and by requiring a manufacturer to provide at least one product or offering that allows users to incorporate provenance information into nonsynthetic content - rather than requiring it in every recording device. This more targeted approach would ensure that, if there are significant costs to developing and incorporating the technology, those costs do not have to be passed onto consumers who do not want them. Consumers could choose this functionality if desired, and the requirements could be phased in over time. In addition, capture device requirements should not apply in B2B, where the needs, considerations, and use cases are sharply different than in the consumer context.

Similarly, large online platforms are deeply engaged in collaborative, multistakeholder efforts to build a scalable and interoperable ecosystem for provenance data. Many are members of key industry organizations, like the Coalition for Content Provenance and Authenticity (C2PA), and are working to establish consensus-driven standards that will give consumers meaningful transparency into synthetic content. Many of these platforms have already made public commitments and investments to build out technical infrastructure aligned with these evolving standards.

By contrast, imposing rigid compliance obligations at this stage risks sidelining these robust industry-led initiatives. Narrowly targeting large platforms also overlooks the broader ecosystem, particularly fringe or noncompliant actors that are unlikely to voluntarily adhere to provenance standards—effectively weakening the bill's intended impact.

[…]

The bill's scope remains unclear regarding platform obligations for third-party or embedded content. Whether large online platforms would be held responsible for the visibility or preservation of provenance information in media hosted elsewhere— such as content embedded from or linked to external sites is ambiguous. While the operative language refers to content "posted," the expansive definition of "online platform" introduces confusion.

For these reasons, we respectfully oppose AB 853.

REGISTERED SUPPORT / OPPOSITION:

Support

California Initiative on Technology and Democracy (Cited) (Sponsor)

Opposition

Computer & Communications Industry Association
Technet-technology Network

Analysis Prepared by: John Bennett / P. & C.P. / (916) 319-2200