SENATE JUDICIARY COMMITTEE
Senator Thomas Umberg, Chair
2023-2024 Regular Session

SB 970 (Ashby)
Version: January 25, 2024
Hearing Date: April 9, 2024
Fiscal: Yes
Urgency: No
CK

SUBJECT

Artificial intelligence technology

DIGEST

This bill ensures that media manipulated or generated by artificial intelligence (AI) technology is incorporated into the right of publicity law and criminal false impersonation statutes. The bill requires those providing access to such technology to provide a warning to consumers about liability for misuse. The bill also requires Judicial Council to create screening procedures to identify written evidence altered or created by AI technology and to provide educational materials to court users and personnel on identifying such materials.

EXECUTIVE SUMMARY

Certain forms of media – audio recordings, video recordings, and still images – can be powerful evidence of the truth. While such media have always been susceptible to some degree of manipulation, fakes were relatively easy to detect. The rapid advancement of AI technology, specifically the wide-scale introduction of generative AI models, has made it drastically cheaper and easier to produce so-called "deepfakes," audio, images, and video recordings that are not real, but that are so realistic that they are virtually impossible to distinguish from authentic content.

This bill works to ensure that AI manipulated media is incorporated into existing laws involving the false impersonation, or use of likeness, of another, namely the right to publicity and false impersonation laws. The bill also tasks Judicial Council with screening for such content in the courts and educating key stakeholders in identifying it. To ensure consumers are on notice of these laws, those selling or providing access to the AI technology designed to create deepfakes are required to warn consumers that misuse can result in civil or criminal liability. This bill is author-sponsored. No timely support or opposition has been received. Should this pass out of this Committee, it will then be referred to the Senate Public Safety Committee.

PROPOSED CHANGES TO THE LAW

Existing law:

1) Establishes California's right of publicity law, which provides that any person who knowingly uses another's name, voice, signature, photograph, or likeness, in any manner, on or in products, merchandise, or goods, or for purposes of advertising or selling, or soliciting purchases of, products, merchandise, goods or services, without such person's prior consent, shall be liable for any damages sustained by the person or persons injured as a result thereof. (Civ. Code § 3344(a).)

2) Subjects a person in violation to liability to the injured party for the greater of the actual damages suffered or statutory damages of $750, and any profits from the unauthorized use that are attributable to the use and are not taken into account in computing the actual damages. Punitive damages may also be awarded to the injured party or parties. The prevailing party shall also be entitled to attorney's fees and costs. (Civ. Code § 3344(a).)

3) Provides that any person who knowingly and without consent credibly impersonates another actual person through or on a website or by other electronic means for purposes of harming, intimidating, threatening, or defrauding another person is guilty of a public offense punishable by a fine and/or imprisonment. (Pen. Code § 528.5.)

4) Provides that every person who falsely personates another in either their private or official capacity, and in that assumed character carries out specified actions, is punishable by a fine and/or imprisonment. (Pen. Code § 529.)

5) Provides that every person who falsely personates another, in either their private or official capacity, and in such assumed character receives any money or property, knowing that it is intended to be delivered to the individual so personated, with intent to convert the same to their own use, or to that of another person, or to deprive the true owner thereof, is punishable in the same manner and to the same extent as for larceny of the money or property so received. (Pen. Code § 530.)

This bill:

1) Requires a person or entity that sells or provides access to any AI technology that is designed to create any synthetic media to provide a consumer warning that misuse of the technology may result in civil or criminal liability for the user.

2) Requires the Department of Consumer Affairs (DCA) to determine the acceptable form and content of the consumer warning required above.

3) Subjects violations to a civil penalty not to exceed $25,000 for each day that the technology is provided to or offered to the public without a consumer warning in a civil action brought by the DCA. The civil penalties collected are to be deposited into the General Fund.

4) Provides, for purposes of the Right to Publicity statute, that a voice or photograph that is synthetic media is deemed to be the voice or photograph of the person depicted, if a reasonable person would believe that the synthetic media is the genuine voice or photograph of that person.

5) Requires Judicial Council to develop and implement screening procedures for writings introduced as evidence to identify those writings that are synthetic media. Judicial Council must also develop and make available to the public educational materials to assist judges, attorneys, and law enforcement officers in understanding and identifying synthetic media and evidence that has been tampered with by means of AI technology.

6) Defines the following terms:

    a) "Artificial intelligence" or "AI" means the simulation of human intelligence processes by computer systems or other machines.
    b) "Synthetic media" means audio, video, or images that have been generated or manipulated by AI technologies to appear to be genuine audio or video recordings or photographic images. Synthetic media includes videos commonly referred to as deepfakes.
    c) "Video cloning technology" means technology that utilizes AI, specifically deep learning and generative adversarial networks (GANs), to create or modify video content in a manner that appears to be an actual recording.
    d) "Voice cloning technology" means technology that utilizes AI to replicate a human voice in a manner that seems to be an actual human voice, including the actual voice of a specific identifiable person.

7) Provides that for the purposes of all Penal Code provisions for which the false impersonation of another is a required element, including, without limitation, Sections 528.5, 529, and 530, the use of video or voice cloning technology with the intent to impersonate another is deemed to be a false personation. Intent to impersonate can be inferred if the synthetic media produced would lead a reasonable person to believe that it is a genuine recording of, or the actual voice of, the person that it is presenting to be.

COMMENTS

1. Blurring reality: AI-generated content

The world has been in awe of the powers of generative AI since the widespread introduction of AI systems such as ChatGPT. However, the capabilities of these advanced systems leads to a blurring between reality and fiction. The Brookings Institution lays out the issue:

> Over the last year, generative AI tools have made the jump from research prototype to commercial product. Generative AI models like OpenAI's ChatGPT and Google's Gemini can now generate realistic text and images that are often indistinguishable from human-authored content, with generative AI for audio and video not far behind. Given these advances, it's no longer surprising to see AI-generated images of public figures go viral or AI-generated reviews and comments on digital platforms. As such, generative AI models are raising concerns about the credibility of digital content and the ease of producing harmful content going forward.
>
> Against the backdrop of such technological advances, civil society and policymakers have taken increasing interest in ways to distinguish AI-generated content from human-authored content.[1]

One expert at the Copenhagen Institute for Future Studies estimates that should large generative-AI models run amok, up to 99 percent of the internet's content could be AI-generated by 2025 to 2030.[2] The problematic applications are seemingly infinite, whether it be deepfakes to blackmail or shame victims, misinformation in elections, false impersonations to commit fraud, or other nefarious purposes. Infamously, in January of this year, Taylor Swift was the victim of sexually explicit, nonconsensual deepfake images using AI that were widely spread across social media platforms.[3] Perhaps more disturbingly, a trend has emerged in schools of students creating such images: "At schools across the country, people have used deepfake technology combined with real images of female students to create fraudulent images of nude

---

[1] Siddarth Srinivasan, *Detecting AI fingerprints: A guide to watermarking and beyond* (January 4, 2024) Brookings Institution, https://www.brookings.edu/articles/detecting-ai-fingerprints-a-guide-to-watermarking-and-beyond/#:~:text=Google%20also%20recently%20announced%20SynthID,model%20to%20detect%20the%20watermark. All internet citations are current as of April 2, 2024.

[2] Lonnie Lee Hood, *Experts Say That Soon, Almost The Entire Internet Could Be Generated by AI* (March 4, 2022) The Byte, https://futurism.com/the-byte/ai-internet-generation.

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

bodies. The deepfake images can be produced using a cellphone."[4] In February of this year, voters in New Hampshire received robocalls that are purported to have used an AI voice resembling President Joe Biden advising them against voting in the presidential primary and saving their vote for the November general election.[5] Recently, a former federal judge urged the federal judiciary's Advisory Committee on Evidence Rules to update evidentiary rules regarding the admissibility of evidence believed to be AI generated.[6] But, in addition to concerns about the potential for AI-generated evidence to be admitted is the reverse, false claims that real evidence is synthetic. As more of the population becomes aware of the potential to realistically fake images, video, and text, some will use the skepticism that creates to challenge the authenticity of real content, a phenomena coined the "liar's dividend."[7]

2. Taking action to identify synthetic content and address its usage

Last month, the European Parliament signed the European Union AI Act. It highlights these very issues and obligates developers and deployers to assist in ensuring, to the extent feasible, that individuals are able to distinguish between original and AI-generated or manipulated content. The Act states:

> A variety of AI systems can generate large quantities of synthetic content that becomes increasingly hard for humans to distinguish from human-generated and authentic content. The wide availability and increasing capabilities of those systems have a significant impact on the integrity and trust in the information ecosystem, raising new risks of misinformation and manipulation at scale, fraud, impersonation and consumer deception. In light of those impacts, the fast technological pace and the need for new methods and techniques to trace origin of information, it is appropriate to require providers of those systems to embed technical solutions that enable marking in a machine readable format and detection that the output has been generated or manipulated by an AI system and not a human. Such techniques and methods should be sufficiently reliable, interoperable, effective and robust as far as this is technically feasible, taking into account available techniques or a combination of such

---

[4] Hannah Fry, Laguna Beach High School investigates 'inappropriate' AI-generated images of students (April 2, 2024) Los Angeles Times, https://www.latimes.com/california/story/2024-04-02/laguna-beach-high-school-investigating-creation-of-ai-generated-images-of-students.

[5] Em Steck & Andrew Kaczynski, *Fake Joe Biden robocall urges New Hampshire voters not to vote in Tuesday's Democratic primary* (January 22, 2024) CNN, https://www.cnn.com/2024/01/22/politics/fake-joe-biden-robocall/index.html.

[6] Avalon Zoppo, *Threat of AI-Generated 'Deepfake' Evidence Needs Judiciary's Attention, Former Judge Says* (October 27, 2023) The National Law Journal, https://www.law.com/nationallawjournal/2023/10/27/threat-of-ai-generated-deepfake-evidence-needs-judiciarys-attention-former-judge-says/?slreturn=20240303000917.

[7] Bobby Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security* (July 14, 2018) 107 California Law Review 1753 (2019), https://ssrn.com/abstract=3213954.

> techniques, such as watermarks, metadata identifications, cryptographic methods for proving provenance and authenticity of content, logging methods, fingerprints or other techniques, as may be appropriate. When implementing this obligation, providers should also take into account the specificities and the limitations of the different types of content and the relevant technological and market developments in the field, as reflected in the generally acknowledged state-of-the-art. Such techniques and methods can be implemented at the level of the system or at the level of the model, including general purpose AI models generating content, thereby facilitating fulfilment of this obligation by the downstream provider of the AI system. To remain proportionate, it is appropriate to envisage that this marking obligation should not cover AI systems performing primarily an assistive function for standard editing or AI systems not substantially altering the input data provided by the deployer or the semantics thereof.

It also specifically obligates deployers who use an AI system to generate or manipulate image, audio or video content that "appreciably resembles existing persons, places or events and would falsely appear to a person to be authentic (deep fakes), should also clearly and distinguishably disclose that the content has been artificially created or manipulated by labelling the artificial intelligence output accordingly and disclosing its artificial origin."

There is currently an arms race in techniques for distinguishing between synthetic and authentic content and companies are declaring their commitment to identifying such content. There are various methods for deciphering AI-generated or altered content, although none are foolproof and all require updates as technology advances:

> There are several approaches that have been proposed for detecting AI-generated content. The four most prominent approaches are watermarking (in its various forms), which is the embedding of an identifiable pattern in a piece of content to track its origin; content provenance, which securely embeds and maintains information about the origin of the content in its metadata; retrieval-based detectors, where all AI-generated content is stored in a database that can be queried to check the origin of content; and post-hoc detectors, which rely on machine learning models to identify subtle but systematic patterns in AI-generated content that distinguish it from human-authored content.[8]

Recently, Meta has committed to "label images that users post to Facebook, Instagram and Threads when we can detect industry standard indicators that they are AI-

[8] *See* footnote 1.

generated."[9] A group of tech companies, including Adobe, Google, and Microsoft, has established the Coalition for Content Provenance and Authenticity (C2PA) to address "the prevalence of misleading information online through the development of technical standards for certifying the source and history (or provenance) of media content."[10] OpenAI announced that it will add C2PA metadata to images created with ChatGPT and the API for the DALL-E 3 model.

In fact, many companies have already voluntarily committed to follow specified guidelines. As described in the White House fact sheet:

> President Biden [convened] seven leading AI companies at the White House [] – Amazon, Anthropic, Google, Inflection, Meta, Microsoft, and OpenAI – to announce that the Biden-Harris Administration has secured voluntary commitments from these companies to help move toward safe, secure, and transparent development of AI technology.
>
> Companies that are developing these emerging technologies have a responsibility to ensure their products are safe. To make the most of AI's potential, the Biden-Harris Administration is encouraging this industry to uphold the highest standards to ensure that innovation doesn't come at the expense of Americans' rights and safety.
>
> These commitments, which the companies have chosen to undertake immediately, underscore three principles that must be fundamental to the future of AI – safety, security, and trust – and mark a critical step toward developing responsible AI. As the pace of innovation continues to accelerate, the Biden-Harris Administration will continue to remind these companies of their responsibilities and take decisive action to keep Americans safe.[11]

The most relevant commitment is focused on earning the public's trust by ensuring individuals are aware of when content is AI generated:

> **Develop and deploy mechanisms that enable users to understand if audio or visual content is AI-generated, including robust provenance, watermarking, or both, for AI-generated audio or visual content**

---

[9] Nick Clegg, *Labeling AI-Generated Images on Facebook, Instagram and Threads* (February 6, 2024) Meta, https://about.fb.com/news/2024/02/labeling-ai-generated-images-on-facebook-instagram-and-threads/.
[10] *Overview*, Coalition for Content Provenance and Authenticity, https://c2pa.org/.
[11] *FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI* (July 21, 2023) The White House, https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/.

> Companies making this commitment recognize that it is important for people to be able to understand when audio or visual content is AI-generated. To further this goal, they agree to develop robust mechanisms, including provenance and/or watermarking systems for audio or visual content created by any of their publicly available systems within scope introduced after the watermarking system is developed. They will also develop tools or APIs to determine if a particular piece of content was created with their system. Audiovisual content that is readily distinguishable from reality or that is designed to be readily recognizable as generated by a company's AI system—such as the default voices of AI assistants—is outside the scope of this commitment. The watermark or provenance data should include an identifier of the service or model that created the content, but it need not include any identifying user information. More generally, companies making this commitment pledge to work with industry peers and standards-setting bodies as appropriate towards developing a technical framework to help users distinguish audio or visual content generated by users from audio or visual content generated by AI.

3. Ensuring synthetic content is incorporated into existing legal frameworks

This bill seeks to ensure that existing laws are equipped to handle the implications of the explosion of AI-generated content.

First, the bill amends California's right to publicity statute. That law provides that any person who knowingly uses another's name, voice, signature, photograph, or likeness, in any manner, on or in products, merchandise, or goods, or for purposes of advertising or selling, or soliciting purchases of, products, merchandise, goods or services, without such person's prior consent shall be liable for any damages sustained by the person or persons injured as a result thereof. The bill adds that a voice or photograph that is synthetic media is deemed to be the voice or photograph of the person depicted, if a reasonable person would believe that the synthetic media is the genuine voice or photograph of that person. The author explains: "This means that if a synthetic voice or image is so realistic that a reasonable person would believe it to be genuine, it is considered the same as the actual voice or image of the person it depicts, offering protection against unauthorized AI-generated content."

"Synthetic media" is defined as audio, video, or images that have been generated or manipulated by AI technologies to appear to be genuine audio or video recordings or photographic images. This includes videos commonly referred to as deepfakes. In order to create some uniformity in the central terms used in AI regulation, the author has agreed to amend this definition to align with that used in President Biden's executive order.

The bill also amends the Penal Code to provide that for any provisions therein that have false impersonation as a required element, the use of video or voice cloning technology with intent to impersonate another is deemed to be a false impersonation. While this provision will be more thoroughly examined in the Senate Public Safety Committee, the additional definitions for the underlying AI technology at play may be unnecessary and may unintentionally narrow what is covered by the bill. The author has agreed to simplify the provision and simply refer to "synthetic media" used with intent to impersonate.

Next, the bill obligates those that sell or provide access to AI technology designed to create synthetic media must provide a warning to consumers about the potential liability that could result from the misuse of that technology. This serves to put consumers on notice that their use of the technology has consequences, hopefully deterring the abuses discussed in detail above.

For this section, the bill requires that the Department of Consumer Affairs (DCA) determine the form and content of this warning. However, there is no phase-period for the obligation or a specific timeline for DCA to create the warning. The author has agreed to amendments setting a deadline for DCA, obligating them to publicly post the guidelines for the warning on a publicly accessible website, and setting an effective date of the obligations on July 1, 2026.

DCA is authorized to bring a civil action against those in violation for civil penalties of up to $25,000 for each day the technology is provided or offered to the public without the warning. Penalties are to be deposited into the General Fund.

Finally, the bill seeks to address the issue cited above about the growing concern that AI-generated or altered media will be introduced as authentic evidence in criminal and civil actions and that challenges on such grounds, founded or not, will be made. The bill requires the Judicial Council to develop and implement screening procedures for writings introduced as evidence to identify that evidence that is synthetic media. Judicial Council is also required to develop and make available to the public educational materials to assist judges, attorneys, and law enforcement officers in understanding and identifying synthetic media and evidence that has been tampered with through AI technology.

While the goal is clear, the difficulty of, and the technical expertise required for, such screening arguably is not something appropriately placed in the hands of Judicial Council to immediately implement. The author has agreed to start this important process by requiring Judicial Council to, no later than January 1, 2026, review the impact of AI on the introduction of evidence in court proceedings and develop any necessary rules of court to assist courts in assessing claims that evidence that is being introduced has been generated by or manipulated by AI.

4. Defining AI

Given the immense potential but attendant challenges and dangers of advancing AI technologies, the Legislature is currently considering dozens of bills on the subject of regulating and fostering AI. The first challenge is determining exactly what we mean by the term. A more thorough discussion of the various definitions that have been crafted by national and international entities for AI can be found in this Committee's analysis of SB 1047 (Wiener, 2024). In order to gain both the benefit of the expertise and compromise that went into formulating those definitions and the efficiencies that come with harmonization, the Committee, with a variety of stakeholders, including the author, have come up with the following definition to begin this process and to amend into the bill:

> "Artificial intelligence" means an engineered or machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs that can influence physical or virtual environments and that may operate with varying levels of autonomy.

SUPPORT

None received

OPPOSITION

None received

RELATED LEGISLATION

Pending Legislation:

SB 942 (Becker, 2024) establishes the California AI Transparency Act, which, among other things, requires a covered provider, as defined, to create 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, as defined, provided by the covered provider that meets certain criteria. Covered providers are required to include in AI-generated content a visible disclosure that, among other things, includes a clear and conspicuous notice, that identifies the content as generated by AI. SB 942 requires a covered provider to register with CDT and provide them a URL to any AI detection tool it has created. SB 942 is currently in this Committee.

SCR 17 (Dodd, 2023) affirms the California Legislature's commitment to President Biden's vision for a safe AI and the principles outlined in the "Blueprint for an AI Bill of Rights" and expresses the Legislature's commitment to examining and implementing

those principles in its legislation and policies related to the use and deployment of automated systems. SCR 17 is currently in the Assembly Privacy and Consumer Protection Committee.

AB 2930 (Bauer-Kahan, 2024) requires, among other things, a deployer and a developer of an automated decision tool to, on or before January 1, 2026, and annually thereafter, perform an impact assessment for any automated decision tool the deployer uses that includes, among other things, a statement of the purpose of the automated decision tool and its intended benefits, uses, and deployment contexts. The assessments must be provided to the Civil Rights Department within 7 days of a request. AB 2930 requires a deployer to, at or before the time an automated decision tool is used to make a consequential decision, notify any natural person that is the subject of the consequential decision that an automated decision tool is being used to make, or be a controlling factor in making, the consequential decision and to provide that person with, among other things, a statement of the purpose of the automated decision tool.
AB 2930 is currently in the Assembly Privacy and Consumer Protection Committee.

AB 3211 (Wicks, 2024) establishes the California Provenance, Authenticity and Watermarking Standards Act, which requires a generative AI system provider, as defined, to take certain actions to assist in the disclosure of provenance data to mitigate harms caused by inauthentic content, including placing imperceptible and maximally indelible watermarks containing provenance data into content created by an AI system that the generative AI system provider makes available. AB 3211 also requires a large online platform, as defined, to, among other things, use labels to prominently disclose the provenance data found in watermarks or digital signatures in content distributed to users on its platforms, as specified. The bill would require a large online platform to use state-of-the-art techniques, including, but not limited to, analysis of user behavioral signals indicating usage of synthetic content, to detect and label inauthentic text content that is uploaded or distributed by individual users or networks of users. AB 3211 is currently in the Assembly Privacy and Consumer Protection Committee.

Prior Legislation:

SB 444 (Umberg, 2019) would have requested the Regents of the University of California (UC) to enact a resolution authorizing the law schools at UC Berkeley and UC Irvine to participate in a pilot project to develop AI or machine-learning solutions to address access to justice issues faced by self-representing litigants in their respective courts. The bill died in the Assembly Higher Education Committee.

AB 1576 (Calderon, 2019) would have required the Secretary of Government Operations to appoint participants to an AI working group to evaluate the uses, risks, benefits, and legal implications associated with the development and deployment of AI by California-based businesses. The bill was held on the Senate Appropriations Committee suspense file.

SJR 6 (Chang, Res. Ch. 112, Stats. 2019) urged the President and the Congress of the United States to develop a comprehensive AI Advisory Committee and to adopt a comprehensive AI policy.

****