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

AB 2355 (Wendy Carrillo)
Version: June 11, 2024
Hearing Date: July 2, 2024
Fiscal: Yes
Urgency: No
CK

SUBJECT

Political advertisements: artificial intelligence

DIGEST

This bill requires committees that create, publish, or distribute a political advertisement that contains any image, audio, or video that is generated or substantially altered using artificial intelligence (AI) to include a disclosure in the advertisement disclosing that the content has been so altered.

EXECUTIVE SUMMARY

Certain forms of media – audio recordings, video recordings, and still images – can be powerful evidence of what truly took place. While such media have always been susceptible to some degree of manipulation, until recently, 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 synthetic content – audio, images, text, and video recordings that are not real, but that are so realistic that they are virtually impossible to distinguish from authentic content, including so-called "deepfakes."

In the context of election campaigns, such deepfakes can be weaponized to deceive voters into thinking that a candidate said or did something which the candidate did not. In an attempt to prevent deepfakes from altering the outcome of an election in this way, this bill requires committees to disclose when certain images, video, or audio in political advertisements have been generated or altered by AI.

The bill is author-sponsored. It is supported by various organizations, including the Los Angeles Area Chamber of Commerce and Oakland Privacy. No timely opposition has been received by the Committee. The bill passed out of the Senate Elections and Constitutional Amendments Committee on a 6 to 0 vote.

PROPOSED CHANGES TO THE LAW

Existing law:

1) Provides that "Congress shall make no law… abridging the freedom of speech..." (U.S. Const., amend. 1.)

2) Applies the First Amendment to the states through operation of the Fourteenth Amendment. (*Gitlow v. New York* (1925) 268 U.S. 652; *NAACP v. Alabama* (1925) 357 U.S. 449.)

3) Provides that no provider or user of an interactive computer service shall be treated for liability purposes as the publisher or speaker of any information provided by another information content provider. (47 U.S.C. § 230.)

4) Defines "materially deceptive audio or visual media" as an image or an audio or video recording of a candidate's appearance, speech, or conduct that has been intentionally manipulated in a manner such that both of the following conditions are met:
   a) the image or audio or video recording would falsely appear to a reasonable person to be authentic; and
   b) the image or audio or video recording would cause a reasonable person to have a fundamentally different understanding or impression of the expressive content of the image or audio or video recording than that person would have if the person were hearing or seeing the unaltered, original version of the image or audio or video recording. (Elec. Code § 20010(e).)

5) Prohibits a person, committee, or other entity from distributing with actual malice materially deceptive audio or visual media of a candidate with the intent to injure the candidate's reputation or to deceive a voter into voting for or against the candidate within 60 days of an election at which a candidate for elective office will appear on the ballot, as specified, and unless specified conditions are met. (Elec. Code § 20010(a).)

6) Exempts audio or visual media that includes a disclosure stating: "This _____ has been manipulated." Requires the blank in the disclosure to be filled with a term that most accurately describes the media, as specified. Requires the following disclosures for visual and audio-only media:
   a) for visual media, the text of the disclosure shall appear in a size that is easily readable by the average viewer and no smaller than the largest font size of other text appearing in the visual media. If the visual media does not include any other text, then the disclosure shall appear in a size that is easily readable by the average viewer. Requires, for visual media that is video, the disclosure to be displayed throughout the duration of the video;

b) for audio-only media, the disclosure shall be read in the clearly spoken manner and in a pitch that can be easily heard by the average listener, at the beginning of the audio, at the end of the audio, and, if the audio is greater than two minutes in length, interspersed within the audio at intervals of not greater than two minutes each. (Elec. Code § 20010(b).)

7) Permits a candidate for elective office whose voice or likeness appears in a materially deceptive audio or visual media distributed in violation of the above provisions, to seek injunctive or other equitable relief prohibiting the distribution of audio or visual media in violation of the provisions of this bill. (Elec. Code § 20010(c)(1).)

9) Permits a candidate for elective office whose voice or likeness appears in materially deceptive audio or visual media distributed in violation of the provisions of this bill to bring an action for general or special damages against the person, committee, or other entity that distributed the materially deceptive audio or visual media, as specified. Requires the plaintiff to bear the burden of establishing the violation through clear and convincing evidence in any civil action alleging a violation of the provisions of this bill, as specified. (Elec. Code § 21101(c)(2).)

13) Provides that the provisions of the above statute shall not be construed to alter or negate any rights, obligations, or immunities of an interactive service provider under the federal Communications Decency Act. (Elec. Code § 20010(d)(1).)

This bill:

1) Provides that if a committee creates, originally publishes, or originally distributes a qualified political advertisement, the qualified political advertisement shall include, in a clear and conspicuous manner, the following disclosure: "This _____ has been generated or substantially altered using artificial intelligence." The blank must indicate if it is audio, an image, or video.

2) Defines "qualified political advertisement" as an advertisement that contains any image, audio, or video that is generated or substantially altered using artificial intelligence. Any image, audio, video, or other media is "generated or substantially altered using artificial intelligence" if either of the following conditions are met:
    a) The visual or audio media is entirely created using artificial intelligence and would falsely appear to a reasonable person to be authentic.
    b) The visual or audio media is materially altered by artificial intelligence such that the alteration would cause a reasonable person to have a fundamentally different understanding of the altered media when comparing it to an unaltered version.

3) Requires the text of the disclosure, for visual media, to appear in a size that is easily readable by the average viewer and no smaller than the largest font size of other text appearing in the visual media. If the visual media does not include any other text, the disclosure shall appear in a size that is easily readable by the average viewer. For visual media that is video, the disclosure shall appear for the duration of the video. If the media consists of audio only, the disclosure shall be read in a clearly spoken manner at the beginning or end of the advertisement and in a pitch and tone substantially similar to the rest of the advertisement.

4) Authorizes the Fair Political Practices Commission (FPPC), if a committee does not comply, to seek injunctive relief to compel compliance and pursue any administrative or civil remedies available.

5) Clarifies that it does not alter or negate any rights, obligations, or immunities of an interactive service provider under Section 230 of Title 47 of the United States Code.

COMMENTS

1. Blurring reality: AI-generated content

Generative AI is a type of artificial intelligence that can create new content, including text, images, code, or music, by learning from existing data. Generative AI models can produce realistic and novel artifacts that resemble the data they were trained on, but do not copy it. For example, generative AI can write a poem, draw a picture, or compose a song based on a given prompt or theme. Generative AI enables users to quickly generate new content based on a variety of inputs. Generative AI models use neural networks to identify the patterns and structures within existing data to generate new and original content.

The world has been in awe of the powers of this 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, 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 bodies. The deepfake images can be produced using a cellphone."[4] 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."[5]

Relevant here, AI and specifically generative AI can spread misinformation regarding elections with ease, both in California and across the world:

> Artificial intelligence is supercharging the threat of election disinformation worldwide, making it easy for anyone with a smartphone and a devious imagination to create fake – but convincing – content aimed at fooling voters.
>
> It marks a quantum leap from a few years ago, when creating phony photos, videos or audio clips required teams of people with time, technical skill and money. Now, using free and low-cost generative artificial

---

[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 June 21, 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/.
[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] 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.

> intelligence services from companies like Google and OpenAI, anyone can create high-quality "deepfakes" with just a simple text prompt.
>
> A wave of AI deepfakes tied to elections in Europe and Asia has coursed through social media for months, serving as a warning for more than 50 countries heading to the polls this year.
>
> "You don't need to look far to see some people ... being clearly confused as to whether something is real or not," said Henry Ajder, a leading expert in generative AI based in Cambridge, England.
>
> The question is no longer whether AI deepfakes could affect elections, but how influential they will be, said Ajder, who runs a consulting firm called Latent Space Advisory.
>
> As the U.S. presidential race heats up, FBI Director Christopher Wray recently warned about the growing threat, saying generative AI makes it easy for "foreign adversaries to engage in malign influence."[6]

On that last note, 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.[7] The examples are endless:

> Former President Donald Trump, who is running in 2024, has shared AI-generated content with his followers on social media. A manipulated video of CNN host Anderson Cooper that Trump shared on his Truth Social platform on Friday, which distorted Cooper's reaction to the CNN town hall this past week with Trump, was created using an AI voice-cloning tool.
>
> A dystopian campaign ad released last month by the Republican National Committee offers another glimpse of this digitally manipulated future. The online ad, which came after President Joe Biden announced his reelection campaign, and starts with a strange, slightly warped image of Biden and the text "What if the weakest president we've ever had was re-elected?"

---

[6] Ali Swenson & Kelvin Chan, *Election disinformation takes a big leap with AI being used to deceive worldwide* (March 14, 2024) Associated Press, https://apnews.com/article/artificial-intelligence-elections-disinformation-chatgpt-bc283e7426402f0b4baa7df280a4c3fd.

[7] 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.

> A series of AI-generated images follows: Taiwan under attack; boarded up storefronts in the United States as the economy crumbles; soldiers and armored military vehicles patrolling local streets as tattooed criminals and waves of immigrants create panic.
>
> "An AI-generated look into the country's possible future if Joe Biden is re-elected in 2024," reads the ad's description from the RNC.
>
> The RNC acknowledged its use of AI, but others, including nefarious political campaigns and foreign adversaries, will not, said Petko Stoyanov, global chief technology officer at Forcepoint, a cybersecurity company based in Austin, Texas. Stoyanov predicted that groups looking to meddle with U.S. democracy will employ AI and synthetic media as a way to erode trust.[8]

Legislatures across the country are pushing legislation that would address this looming threat.

2. AI generated or altered content in political advertisements

According to the author:

> Since the broad public release of generative AI applications to create sound, video, photos, and text since 2022, we have seen widespread adoption of and noticeable technological improvement in these tools. In a world where fabricated material is easier to create than ever before, protections are needed to ensure that content created by digital tools is properly labelled. Sensible regulation of this type of digital content balances free speech protections with the need to protect and uphold faith in our electoral democracy by updating the disclosure requirements in the Political Reform Act.

This bill institutes baseline protections in political advertisements. It requires a clear and conspicuous disclosure when an image, audio, or video in a political advertisement is generated or substantially altered using AI. This means that the content is either entirely created using AI and would falsely appear to a reasonable person to be authentic or it is materially altered by AI such that the alteration would cause a reasonable person to have a fundamentally different understanding of the altered media when comparing it to an unaltered version.

---

[8] David Klepper & Ali Swenson, *AI-generated disinformation poses threat of misleading voters in 2024 election* (May 14, 2023) PBS News, https://www.pbs.org/newshour/politics/ai-generated-disinformation-poses-threat-of-misleading-voters-in-2024-election.

Given the narrow scope of the requirement and the fact that no content is prohibited, but simply requires disclosure as to its provenance, there is likely little concern regarding offending the First Amendment, and no such concerns have been raised with regard to the current version of the bill.

The bill borrows from existing law. In anticipation of the possibility that deepfakes might be used to try to influence the outcome of the 2020 election, California enacted AB 730 (Berman, Ch. 493, Stats. 2019). AB 730 prohibited the use of deepfakes depicting a candidate for office within 60 days of the election unless the deepfake is accompanied by a prominent notice that the content of the audio, video, or image has been manipulated. (Elec. Code § 20010(a),(b).) Additionally, AB 730 authorized a candidate who was falsely depicted in a deepfake to seek rapid injunctive relief against further publication and distribution of the deepfake. (Code Civ. Proc. §35(a); Elec. Code § 20010(c).) AB 972 (Berman, Ch. 745, Stats. 2022) later extended the sunset date placed on these provisions to January 1, 2027.

3. Stakeholder positions

Oakland Privacy writes in support:

> Generative artificial intelligence can now create fake (i.e. artificial) content that can seem to decisively indicate that someone said something they didn't say, was at a location they never visited or that statistics and other factual material that impacts policy are not accurate when they are. Unlike mere "claims" that can be rebutted; generative AI can provide what seems like dispositive evidence of truth or falsity, but is a mere computer projection of bits and bytes. This can wreak havoc on a voter's ability to research what is true and what is not and make their decisions accordingly.
>
> Assembly Bill 2355 seeks to help voters in this position by simply requiring that synthetic content be labeled as such, so it is therefore harder to use generative AI election content as evidence of truth or falsity. This provision is literally structured in the same manner as long-standing California election law that requires the labeling of paid political advertisements. As such, the requirement is straightforward, understandable to those that would have to abide by it, doesn't depend on unreliable technologies like watermarking, and shouldn't confuse voters or require them to understand a lot about AI to benefit from the legislation.
>
> Because AB 2355 is transparency legislation, it does not trigger significant First Amendment concerns and would not suppress or censor any

> properly-labeled content nor is it likely to be gamed by candidates to gain an advantage in a hard-fought election.

Writing in support, the Los Angeles Area Chamber of Commerce states:

> It is understood that Generative Artificial Intelligence can create fake content that can seem to decisively indicate that someone said something they didn't say, was at a location they never visited or that statistics and other factual material that impacts policy are not accurate when they are. This "deepfake" content can be extremely misleading and can negatively impact elections. AB 2355 seeks to help voters by simply requiring that synthetic content used in advertisements be labeled as such, so it is harder to use generative AI election content as evidence of truth or falsity. The potential threat posed by manipulated media to future elections' integrity is more significant now than it has ever been. Action must be taken in order to ensure election integrity.

SUPPORT

California Clean Money Campaign
California Contract Cities Association
Computer & Communications Industry Association
Los Angeles Area Chamber of Commerce
Oakland Privacy
Software & Information Industry Association
Technet

OPPOSITION

None received

RELATED LEGISLATION

Pending Legislation:

SB 942 (Becker, 2024) establishes the California AI Transparency Act, requiring covered providers to create and make freely available an AI detection tool to detect content as AI-generated and to include disclosures in content generated by the provider's system. SB 942 is currently in the Assembly Judiciary Committee.

SB 970 (Ashby, 2024) 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. SB 970 was held on suspense in the Senate Appropriations Committee.

AB 2655 (Berman, 2024) establishes the Defending Democracy from Deepfake Deception Act of 2024, which requires a large online platform to block the posting or sending of materially deceptive and digitally modified or created content related to elections, during specified periods before and after an election. It requires these platforms to label certain additional content inauthentic, fake, or false during specified periods before and after an election and to provide mechanisms to report content. AB 2655 is currently in this Committee.

AB 2839 (Pellerin, 2024) prohibits a person, committee, or other entity from knowingly distributing an advertisement or other election communication that contains materially deceptive content, as defined and specified, with malice, except as provided, within 120 days of a California election, and in specified cases, 60 days thereafter. AB 2839 is currently in this Committee.

AB 2930 (Bauer-Kahan, 2024) requires, among other things, a deployer and a developer of an automated decision tool to 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. 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 substantial 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 this Committee.

AB 3211 (Wicks, 2024) establishes the California Provenance, Authenticity and Watermarking Standards Act, which requires a generative AI system provider 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. AB 3211 is currently in the Senate Appropriations Committee.

Prior Legislation:

AB 972 (Berman, Ch. 745, Stats. 2022) *See* Comment 2.

AB 730 (Berman, Ch. 493, Stats. 2019) *See* Comment 2.

PRIOR VOTES:

Senate Elections Committee (Ayes 6, Noes 0)
Assembly Floor (Ayes 64, Noes 0)
Assembly Appropriations Committee (Ayes 13, Noes 0)
Assembly Judiciary Committee (Ayes 9, Noes 1)
Assembly Privacy and Consumer Protection Committee (Ayes 9, Noes 1)
Assembly Elections Committee (Ayes 7, Noes 1)
****