File No. 240508

Committee Item No. ______________
Board Item No. 22

COMMITTEE/BOARD OF SUPERVISORS

AGENDA PACKET CONTENTS LIST

Committee: ______________ **Date:** ______________
Board of Supervisors Meeting **Date:** May 21, 2024

Cmte	Board	
□	□	**Motion**
□	☒	**Resolution**
□	□	**Ordinance**
□	□	**Legislative Digest**
□	□	**Budget and Legislative Analyst Report**
□	□	**Youth Commission Report**
□	☒	**Introduction Form**
□	□	**Department/Agency Cover Letter and/or Report**
□	□	**MOU**
□	□	**Grant Information Form**
□	□	**Grant Budget**
□	□	**Subcontract Budget**
□	□	**Contract/Agreement**
□	□	**Form 126 – Ethics Commission**
□	□	**Award Letter**
□	□	**Application**
□	□	**Public Correspondence**

OTHER

□	☒	Assembly Bill 2655 - 4/24/24
□	☒	Assembly Bill 2839 - 5/2/24
□	☒	Assembly Bill 3211 - 4/18/24
□	☒	Senate Bill 1228 - 4/25/24
□	☒	CSAC/LCC Position
□	□	
□	□	

Prepared by: Lisa Lew **Date:** May 17, 2024
Prepared by: ______________ **Date:** ______________

[Supporting California State Assembly Bill Nos. 2655 (Berman), 2839 (Pellerin), 3211 (Wicks), and Senate Bill No. 1228 (Padilla) - Artificial Intelligence in Elections]

Resolution supporting California State Assembly Bill No. 2655 introduced by Assembly Member Marc Berman, Assembly Bill No. 2839 introduced by Assembly Member Gail Pellerin, Assembly Bill No. 3211 introduced by Assembly Member Buffy Wicks, and Senate Bill No. 1228 introduced by Senator Steve Padilla to address various technologies of artificial intelligence in elections.

WHEREAS, Artificial Intelligence technology has greatly developed in recent years, causing concerns over the potential impacts on election integrity; and

WHEREAS, There have already been instances across the country of AI technology being used to interfere in elections on all levels; and

WHEREAS, There have been efforts at the state and federal levels, including the Federal Elections Commission (FEC), to introduce legislation to reduce the potential harms of AI on elections; and

WHEREAS, Supervisor Dean Preston held a hearing on AI in local elections at the Rules Committee of the San Francisco Board of Supervisors on May 13, 2024; and

WHEREAS, The Brennan Center released a report on May 8, 2024, entitled ‘How Election Officials Can Identify, Prepare for, and Respond to AI Threats,’ which provides recommendations on how cities can reduce and prevent risks associated with AI technology, and which was discussed at the May 13, 2024, Rules Committee hearing; and

WHEREAS, A State Bill package was identified at the May 13, 2024 Rules Committee hearing as crucial to preventing misleading or deceptive AI-generated material, and proactively helping to ensure that Californians have access to transparent and accurate information; and

WHEREAS, Assembly Bill No. 2655 (AB 2655), introduced by Assembly Member Marc Berman, would require large online platforms to block the posting or sending of materially deceptive and digitally modified or created content related to elections, or to label that content, during specified periods before and after an election; and

WHEREAS, Assembly Bill No. 2839 (AB 2839), introduced by Assembly Member Gail Pellerin, would prohibit the distribution of an advertisement or other election communication that contains materially deceptive and digitally altered or created images, audio, or video files with the intent to influence an election or solicit funds for a candidate or campaign; and

WHEREAS, Assembly Bill No. 3211 (AB 3211), introduced by Assembly Member Buffy Wicks, would require the implementation of watermark and content provenance requirements on artificial intelligence (AI) system providers, camera technology manufacturers, and online platforms, enforceable through administrative penalties assessed by the California Department of Technology (CDT); and

WHEREAS, Senate Bill No. 1228 (SB 1228), introduced by State Senator Steve Padilla, would require large online platforms to seek to verify influential users, to label such accounts and their posts with notes that the user is or is not authenticated by the platform, and authorizes public prosecutors to file prioritized actions to enjoin violations and seek other equitable relief; now, therefore, be it

RESOLVED, That the San Francisco Board of Supervisors expresses its concern about the potential use of AI-generated content to mislead or deceive San Francisco voters and threaten election integrity; and, be it

FURTHER RESOLVED, That the San Francisco Board of Supervisors finds that it is essential for the State of California to take action to prevent the potential use of AI-generated content to mislead or deceive San Francisco voters and threaten election integrity; and, be it

FURTHER RESOLVED, That the San Francisco Board of Supervisors supports, and urges the California State Legislature and Governor to support, AB 2655, AB 2839, AB 3211, and SB 1228; and, be it

FURTHER RESOLVED, That the Board of Supervisors hereby directs the Clerk of the Board to transmit a copy of this resolution to Assemblymember Matt Haney, Assemblymember Phil Ting, State Senator Scott Wiener, Assembly Member Marc Berman, Assembly Member Gail Pellerin, Assembly Member Buffy Wicks, State Senator Steve Padilla, California Senate President Pro Tempore Mike McGuire, California Assembly Speaker Robert Rivas, and Governor Gavin Newsom.

AMENDED IN ASSEMBLY APRIL 24, 2024

AMENDED IN ASSEMBLY APRIL 1, 2024

AMENDED IN ASSEMBLY MARCH 21, 2024

CALIFORNIA LEGISLATURE—2023–24 REGULAR SESSION

ASSEMBLY BILL No. 2655

Introduced by Assembly Members Berman and Pellerin
(Principal coauthor: Assembly Member Cervantes)
(Coauthor: Assembly Member Bennett)

February 14, 2024

An act to amend Section 35 of the Code of Civil Procedure, and to add Chapter 7 (commencing with Section 20510) to Division 20 of the Elections Code, relating to elections.

LEGISLATIVE COUNSEL'S DIGEST

AB 2655, as amended, Berman. Defending Democracy from Deepfake Deception Act of 2024.

Existing law establishes requirements for the conduct of election campaigns, including requirements regarding the endorsement of candidates, political corporations, campaign funds, fair campaign practices, and libel and slander. Existing law, until January 1, 2027, prohibits any person, committee, or other entity from distributing, with actual malice, materially deceptive audio or visual media of a candidate for elective office 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 the election. Existing law requires specified actions pertaining to elections to be given precedence when they are filed in court, including actions involving the registration of voters, the certification of candidates and measures, and election contests, and, until January

1, 2027, actions involving the foregoing prohibition against materially deceptive media.

This bill would establish the Defending Democracy from Deepfake Deception Act of 2024 for the purpose of preventing the online dissemination of manipulated media and disinformation meant to deceive voters and to prevent them from voting. The bill would require a large online platform, as defined, 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. The bill would require a large online platform to label certain additional content inauthentic, fake, or false during specified periods before and after an election.

The bill would require a large online platform to develop procedures for California residents to report content that has not been blocked or labeled in compliance with the act. The bill would also authorize California residents, the Attorney General, and a district attorney or city attorney to seek injunctive relief against a large online platform for noncompliance with the act, as specified, and would assign precedence to such actions when they are filed in court.

The bill would exempt from its provisions a regularly published online newspaper, magazine, or other periodical of general circulation that routinely carries news and commentary of general interest, if the publication complies with specified disclosure requirements. The bill would also exempt content that is satire or parody.

Vote: majority. Appropriation: no. Fiscal committee: yes. State-mandated local program: no.

The people of the State of California do enact as follows:

SECTION 1. Section 35 of the Code of Civil Procedure, as amended by Section 1 of Chapter 343 of the Statutes of 2023, is amended to read:

35. (a) Proceedings in cases involving the registration or denial of registration of voters, the certification or denial of certification of candidates, the certification or denial of certification of ballot measures, election contests, actions under Section 20010 of the Elections Code, actions under Chapter 7 (commencing with Section 20510) of Division 20 of the Elections Code, and actions under Chapter 2 (commencing with Section 21100) of Division 21 of the

Elections Code shall be placed on the calendar in the order of their date of filing and shall be given precedence.

(b) This section shall remain in effect only until January 1, 2027, and as of that date is repealed, unless a later enacted statute, that is enacted before January 1, 2027, deletes or extends that date.

SEC. 2. Section 35 of the Code of Civil Procedure, as amended by Section 2 of Chapter 343 of the Statutes of 2023, is amended to read:

35. (a) Proceedings in cases involving the registration or denial of registration of voters, the certification or denial of certification of candidates, the certification or denial of certification of ballot measures, election contests, actions under Chapter 7 (commencing with Section 20510) of Division 20 of the Elections Code, and actions under Chapter 2 (commencing with Section 21100) of Division 21 of the Elections Code shall be placed on the calendar in the order of their date of filing and shall be given precedence.

(b) This section shall become operative January 1, 2027.

SEC. 3. Chapter 7 (commencing with Section 20510) is added to Division 20 of the Elections Code, to read:

Chapter 7. Defending Democracy from Deepfake Deception Act of 2024

20510. This chapter shall be known and may be cited as the Defending Democracy from Deepfake Deception Act of 2024.

20511. The Legislature finds and declares all of the following:

(a) California is entering its first-ever generative artificial intelligence (AI) election, in which disinformation powered by generative AI will pollute our information ecosystems like never before. Voters will not know what images, audio, or video they can trust.

(b) In a few clicks, using current technology, bad actors now have the power to create a false image of a candidate accepting a bribe or a fake video of an elections official “caught on tape” saying that voting machines are not secure, or to generate the Governor’s voice telling millions of Californians their voting site has changed.

(c) In the lead-up to the 2024 presidential elections, candidates and parties are already creating and distributing deepfake images and audio and video content. These fake images or files can spread to millions of Californians in seconds and skew election results or undermine trust in the ballot counting process.

(d) The labeling information required by this bill is narrowly tailored to provide consumers with factual information about the inauthenticity of particular images, audio, video, or text content in order to prevent consumer deception.

(e) In order to ensure California elections are free and fair, California must, for a limited time before and after elections, prevent the use of deepfakes and disinformation meant to prevent voters from voting and to deceive voters based on fraudulent content.

20512. For purposes of this chapter, the following terms have the following meanings:

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

~~(a)~~

(b) (1) "Elections official" means any of the following persons, but only in their capacity as a person charged with holding or conducting an election, conducting a canvass, assisting with the holding or conducting of an election or a canvas, or performing another duty related to administering the provisions of this code:

(A) An elections official as defined in Section 320.

(B) The Secretary of State and their staff.

(C) A temporary worker, poll worker, or member of a precinct board.

(D) Any other person charged with holding or conducting an election, conducting a canvass, assisting with the holding or conducting of an election or a canvas, or performing another duty related to administering the provisions of this code.

(2) The requirements of this chapter relating to content portraying an elections official apply only if the large online platform knows or should know that the person is an elections official.

~~(b)~~

(c) "Election processes" means any government process related to an election, including, but not limited to, elections, candidates,

vote counting, redistricting, and proceedings or processes of the electoral college.

~~(c)~~

(d) (1) "Materially deceptive and digitally modified or created content" means an image or an audio or video recording or other digital content, including a chatbot, that has been intentionally manipulated such that all of the following conditions are met:

(A) (i) The digital content is the product of digital manipulation, *including, but not limited to,* artificial intelligence~~, or machine learning, including deep learning techniques, that merges, combines, replaces, or superimposes content onto an image or an audio or video recording, creating an image or an audio or video recording that appears authentic, or that otherwise generates an inauthentic image or an audio or video recording~~ that appears authentic,~~ and~~ *but* that contains a false portrayal of any of the following: a candidate for elective office, elected official, elections official, voting machine, ballot, voting site, other property or equipment related to an election, or elections process.

(ii) For purposes of this subdivision, "false portrayal" means the content would cause a reasonable person to have a fundamentally different understanding or impression of the content than the person would have if they were hearing or seeing~~ the unaltered, original~~ *an authentic* version of the content.

(B) The person or entity who attempted to post or send, or who did post or send, the content did so knowing the portrayal was false, or did so with reckless disregard for whether the portrayal was false. If the content is intentionally manipulated and contains a false portrayal as specified in subparagraph (A), there shall be a rebuttable presumption that the person or entity knew the portrayal was false or that they acted with reckless disregard for whether the portrayal was false.

(2) "Materially deceptive and digitally modified or created content" does not include any image or audio or video recording that contains only minor modifications that do not lead to significant changes to the perceived contents or meaning of the content. Minor changes include changes to the brightness or contrast of images, removal of background noise in audio, and other minor changes that do not impact the content of the image or audio or video recording.

~~(d)~~

(e) “Large online platform” means a public-facing internet website, web application, or digital application, including a social network, video sharing platform, advertising network, or search engine that had at least 1,000,000 California users during the preceding 12 months.

20513. (a) Any large online platform, using state-of-the-art, best available tools to detect digitally modified or created content, shall develop and implement procedures for blocking and preventing, and shall, if the large online platform knows or should know that the digitally modified or created content meets the requirements of this section, block and prevent, the posting or sending of any materially deceptive and digitally modified or created content, during the applicable time period or periods set forth in subdivision (c), of any of the following:

(1) A candidate for election portrayed as doing or saying something that the candidate did not do or say.

(2) An elections official portrayed as doing or saying something in connection with the performance of their elections-related duties that the elections official did not do or say.

(3) An elected official portrayed as doing or saying something that influences the election that the elected official did not do or say.

(4) A voting machine, ballot, voting site, or other property or equipment related to an election that is portrayed in a materially false way.

(b) (1) Notwithstanding paragraph (1) of subdivision (a), a large online platform shall not prevent a candidate for an election, during the time period set forth in subdivision (c), from portraying themself as doing or saying something that the candidate did not do or say, but only if the digital content includes a disclosure stating the following: “This _____ has been manipulated.” The blank in this disclosure shall be filled in with whichever of the following terms most accurately describes the media:

(A) Image.

(B) Audio.

(C) Video.

(2) (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,

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.

(B) If the media consists of audio only, the disclosure shall be read in a 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.

(c) (1) Except as provided in paragraph (2), any large online platform shall block and prevent the content described in subdivision (a), and any candidate shall include the disclosure required by subdivision (b), during a period beginning 120 days before the election and through the day of the election.

(2) If the content described in subdivision (a) depicts or pertains to elections officials, or depicts or pertains to a voting machine, ballot, voting site, or other property or equipment related to an election, a large online platform shall block and prevent the content during a period beginning 120 days before the election and ending on the 60th day after the election.

20514. (a) With respect to any materially deceptive and digitally modified or created content that pertains to election processes and that is not subject to Section 20513, a large online platform, using state-of-the-art, best available tools to detect digitally modified or created content, shall develop and implement procedures for labeling such content as inauthentic, fake, or false, and shall, if the large online platform knows or should know that the digitally modified or created content meets the requirements of this section, label such content in this manner, during the applicable time period or periods set forth in subdivision (c).

(b) The label required by subdivision (a) shall permit users to click or tap on it and to inspect all available provenance data about the digitally modified or created content in an easy-to-understand format.

(c) The labeling requirement set forth in subdivision (a) applies during any of the following time periods, to the extent applicable:

(1) The period beginning one year before the election and through the day of the election that is specified in or implicated by the content.

(2) The period beginning one year before the election process and through the final day of the election process that is specified in or implicated by the content.

(3) If the content depicts or pertains to elections officials, the period beginning one year before the election or election process that is specified in or implicated by the content and ending on the 60th day after that election or the 60th day after the final day of that election process, as applicable.

20515. (a) A large online platform shall provide an easily accessible way for California residents to report to that platform content subject to Section 20513 or 20514 that was not blocked or labeled as required. The online platform shall respond to the person who made the report, within 36 hours of the report, describing any action taken or not taken by the online platform with respect to the content.

(b) Any California resident who has made a report to a large online platform under subdivision (a) and who either has not received a response within 36 hours or disagrees with the response, may seek injunctive or other equitable relief against the online platform to compel compliance with this chapter. The court shall award a prevailing plaintiff reasonable attorney’s fees and costs. An action under this subdivision shall be entitled to precedence in accordance with Section 35 of the Code of Civil Procedure.

20516. The Attorney General or any district attorney or city attorney may seek injunctive or other equitable relief against any large online platform to compel compliance with this chapter. The court shall award a prevailing plaintiff reasonable attorney’s fees and costs. An action under this section shall be entitled to precedence in accordance with Section 35 of the Code of Civil Procedure.

20517. This chapter applies to materially deceptive and digitally modified or created content, regardless of the language used in the content. If the language used is not English, the disclosure required by subdivision (b) of Section 20513 and the label required by Section 20514 must appear in the language used as well as in English.

20518. A large online platform that blocks or labels any materially deceptive and digitally modified or created content shall maintain a copy of the digital content for a period of not less than five years from the election or election process specified or implicated in the content and shall make such digital content available to the Secretary of State, the Fair Political Practices Commission, and researchers, if requested.

20519. (a) This chapter does not preclude a large online platform from blocking or labeling any materially deceptive and digitally modified or created content outside of the time periods specified in Sections 20513 and 20514.

(b) This chapter does not preclude any online platform not subject to this chapter from blocking or labeling any materially deceptive and digitally modified or created content.

20520. This chapter does not apply to either of the following:

(a) A regularly published online newspaper, magazine, or other periodical of general circulation that routinely carries news and commentary of general interest, and that publishes any materially deceptive and digitally altered or digitally created image, audio, or video recording that an online platform is required to block or label based on this chapter, if the publication contains a clear disclosure that the materially deceptive and digitally altered or digitally created image or audio or video recording does not accurately represent any actual event, occurrence, appearance, speech, or expressive conduct.

(b) Materially deceptive and digitally altered or digitally created content that constitutes satire or parody.

20521. The provisions of this chapter are severable. If any provision of this chapter or its application is held invalid, that invalidity shall not affect other provisions or applications that can be given effect without the invalid provision or application.

O

AMENDED IN ASSEMBLY MAY 2, 2024

AMENDED IN ASSEMBLY APRIL 11, 2024

CALIFORNIA LEGISLATURE—2023–24 REGULAR SESSION

ASSEMBLY BILL No. 2839

Introduced by Assembly Members Pellerin and Berman
(Principal coauthor: Assembly Member Cervantes)
(Coauthors: Assembly Members Bennett, Jackson, *Quirk-Silva*, Ting, Valencia, Weber, and Wood)
(Coauthors: Senators Becker and Dodd)

February 15, 2024

An act to amend Section 35 of the Code of Civil Procedure, and to add Section 20012 to the Elections Code, relating to elections.

LEGISLATIVE COUNSEL'S DIGEST

AB 2839, as amended, Pellerin. Elections: deceptive media in advertisements.

Existing law prohibits certain distribution of materially deceptive audio or visual media of a candidate within 60 days of an election at which the candidate will appear on the ballot, unless the media includes a disclosure stating that the media has been manipulated, subject to specified exemptions. Existing law authorizes a candidate whose voice or likeness appears in audio or visual media distributed in violation of these provisions to file specified actions, and it requires a court to place such proceedings on the calendar in the order of their date of filing and give the proceedings precedence.

This bill would prohibit a person, committee, or other entity from knowingly distributing an advertisement or other election communication, as defined, that contains certain materially deceptive and digitally altered or digitally created images or audio or video files, as defined, with the intent to influence an election or solicit funds for a candidate or campaign, subject to specified exemptions. The bill would apply this prohibition within 120 days of an election and, in specified cases, 60 days after an election. The bill would authorize a recipient of a materially deceptive and digitally altered or digitally created image or audio or video file distributed in violation of this section, candidate or committee participating in the election, or officer holding an election or conducting a canvass to file a civil action to enjoin the distribution of the media and to seek damages against the person, committee, or other entity that distributed it. The bill would require a court to place such proceedings on the calendar in the order of their date of filing and give the proceedings precedence.

Vote: majority. Appropriation: no. Fiscal committee: no. State-mandated local program: no.

The people of the State of California do enact as follows:

SECTION 1. Section 35 of the Code of Civil Procedure, as amended by Section 1 of Chapter 343 of the Statutes of 2023, is amended to read:

35. (a) Proceedings in cases involving the registration or denial of registration of voters, the certification or denial of certification of candidates, the certification or denial of certification of ballot measures, election contests, actions under Section 20010 or 20012 of the Elections Code, and actions under Chapter 2 (commencing with Section 21100) of Division 21 of the Elections Code shall be placed on the calendar in the order of their date of filing and shall be given precedence.

(b) This section shall remain in effect only until January 1, 2027, and as of that date is repealed, unless a later enacted statute, that is enacted before January 1, 2027, deletes or extends that date.

SEC. 2. Section 35 of the Code of Civil Procedure, as amended by Section 2 of Chapter 343 of the Statutes of 2023, is amended to read:

35. (a) Proceedings in cases involving the registration or denial of registration of voters, the certification or denial of certification of candidates, the certification or denial of certification of ballot measures, election contests, actions under Section 20012 of the Elections Code, and actions under Chapter 2 (commencing with

Section 21100) of Division 21 of the Elections Code shall be placed on the calendar in the order of their date of filing and shall be given precedence.

(b) This section shall become operative January 1, 2027.

SEC. 3. Section 20012 is added to the Elections Code, to read:

20012. (a) The Legislature finds and declares as follows:

(1) California is entering its first-ever artificial intelligence (AI) election, in which disinformation powered by generative AI will pollute our information ecosystems like never before. Voters will not know what images, audio, or video they can trust.

(2) In a few clicks, using current technology, bad actors now have the power to create a false image of a candidate accepting a bribe, or a fake video of an elections official "caught on tape" saying that voting machines are not secure, or generate an artificial robocall in the Governor's voice telling millions of Californians their voting site has changed.

(3) In the lead-up to the 2024 presidential elections, candidates and parties are already creating and distributing deepfake images and audio and video content. These fake images or files can skew election results, even if they use older methods of distribution, such as mail, television, telephone, and text, and undermine trust in the ballot counting process.

(4) In order to ensure California elections are free and fair, California must, for a limited time before and after elections, prevent the use of deepfakes and disinformation meant to prevent voters from voting and deceive voters based on fraudulent content. *The provisions of this bill are narrowly tailored to advance California's compelling interest in protecting free and fair elections.*

(5) The labeling information required by this bill is narrowly tailored to provide consumers with factual information about the inauthenticity of particular images, audio, ~~video~~ *video,* or text content in order to prevent consumer deception.

(b) (1) A person, committee, or other entity shall not, during the time period set forth in subdivision (c), with the intent to influence an election or solicit funds for a candidate or campaign, knowingly distribute an advertisement or other election communication containing a materially deceptive and digitally altered or digitally created image or audio or video file of any of the following:

(A) A candidate portrayed as doing or saying something that the candidate did not do or say.

(B) An officer holding an election or conducting a canvass portrayed as doing or saying something in connection with the election that the officer holding an election or conducting a canvass did not do or say.

(C) An elected official portrayed as doing or saying something in connection with the election that the elected official did not do or say.

(D) A voting machine, ballot, voting site, or other elections-related property or equipment portrayed in a materially false way.

(2) Notwithstanding subparagraph (A) of paragraph (1), a candidate may portray themself as doing or saying something that the candidate did not do or say, but only if the image or audio or video file includes a disclosure stating "This ____ has been manipulated." and complies with the following requirements:

(A) The blank in the disclosure required by paragraph (2) shall be filled with whichever of the following terms most accurately describes the media:

(i) Image.

(ii) Audio.

(iii) Video.

(B) (i) 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, 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.

(ii) If the media consists of audio only, the disclosure shall be read in a 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.

(c) The prohibition in subdivision (b) applies only during the following time periods:

(1) One hundred twenty days before any election.

(2) For people and items set forth in subparagraphs (B) and (D) of paragraph (1) of subdivision (b), 120 days before any election through 60 days after the election, inclusive.

(d) (1) A recipient of a materially deceptive and digitally altered or digitally created image or audio or video file distributed in violation of this section, candidate or committee participating in the election, or officer holding an election or conducting a canvass may seek injunctive or other equitable relief prohibiting the distribution of the materially deceptive and digitally altered or digitally created image or audio or video file in violation of this section. The court shall also award a prevailing plaintiff reasonable attorney's fees and costs. An action under this paragraph shall be entitled to precedence in accordance with Section 35 of the Code of Civil Procedure.

(2) A recipient of a materially deceptive and digitally altered or digitally created image or audio or video file distributed in violation of this section, candidate or committee participating in the election, or officer holding an election or conducting a canvass may bring an action for general or special damages against the person, committee, or other entity that distributed the materially deceptive and digitally altered or digitally created image or audio or video file in violation of this section. The court shall also award a prevailing party reasonable attorney's fees and costs. This subdivision shall not be construed to limit or preclude a plaintiff from securing or recovering any other available remedy at law or equity.

(3) In any civil action alleging a violation of this section, the plaintiff shall bear the burden of establishing the violation through clear and convincing evidence.

(e) (1) This section does not apply to a radio or television broadcasting station, including a cable or satellite television operator, programmer, or producer, that broadcasts any materially deceptive and digitally altered or digitally created image or audio or video file prohibited by this section as part of a bona fide newscast, news interview, news documentary, or on-the-spot coverage of bona fide news events, if the broadcast clearly acknowledges through content or a disclosure, in a manner that can be easily heard or read by the average listener or viewer, that the materially deceptive audio or visual media does not accurately represent any actual event, occurrence, appearance, speech, or expressive conduct.

(2) This section does not apply to a regularly published newspaper, magazine, or other periodical of general circulation, including an internet or electronic publication, that routinely carries news and commentary of general interest, and that publishes any materially deceptive and digitally altered or digitally created image or audio or video file prohibited by this section, if the publication clearly states that the materially deceptive and digitally altered or digitally created image or audio or video file does not accurately represent any actual event, occurrence, appearance, speech, or expressive conduct.

(3) This section does not apply to a materially deceptive and digitally altered or digitally created image or audio or video file that constitutes satire or parody.

(f) For purposes of this section, the following definitions apply:

(1) "Advertisement" means any general or public communication that is authorized or paid for the purpose of supporting or opposing a candidate for elective office or a ballot measure and that is broadcast by or through television, radio, telephone, or text, or disseminated by print media, including billboards, video billboards or screens, and other similar types of advertising.

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

~~(2)~~

(3) "Committee" means a committee as defined in Section 82013 of the Government Code.

~~(3)~~

(4) "Election communication" means any general or public communication not covered under "advertisement" that is broadcast by or through television, radio, telephone, or text, or disseminated by print media, including billboards, video billboards or screens, and other similar types of communications, that concerns any of the following:

(A) A candidate for office or ballot measure.

(B) Voting or refraining from voting in an election.

(C) The canvass of the vote.

~~(4)~~

(5) (A) “Materially deceptive and digitally modified or created image or audio or video file” means an image or an audio or video file that has been intentionally manipulated in a manner such that all of the following conditions are met:

(i) The image or audio or video file is the product of digital ~~manipulation, artificial intelligence, or machine learning, including deep learning techniques, that merges, combines, replaces, or superimposes content onto an image or an audio or video file, creating an image or an audio or video file~~ *manipulation or artificial intelligence* that appears authentic~~, or generates an inauthentic image or an audio or video file that appears authentic.~~ *but that contains a false portrayal of any of the following:*

(I) Candidate for elective office.

(II) Elected official.

(III) Elections official.

(IV) Voting machine.

(V) Ballot.

(VI) Voting site.

(VII) Other property or equipment related to an election or elections process.

~~(ii) (I) The image or audio or video file represents a false portrayal of a candidate for elective office, an elected official, an elections official, or a voting machine, ballot, voting site, or other elections property or equipment.~~

~~(II)~~

(ii) For the purposes of this clause, “a false portrayal of the candidate for elective office, an elected official, an elections official, or a voting machine, ballot, voting site, or other elections property or equipment” means the image or audio or video file would cause a reasonable person *to believe that the content is authentic and* to have a fundamentally different understanding or impression of the expressive content of the image or audio or video file than that person would have if the person were hearing or seeing the~~ unaltered, original~~ *authentic* version of the image or audio or video file.

(iii) The person, committee, or other entity distributed the image or audio or video file knowing the portrayal of the candidate for elective office, the elected official, the elections official, or the voting machine, ballot, voting site, or other elections property or equipment was false or with a reckless disregard for the true portrayal of the candidate, the elected official, the elections official, or the voting machine, ballot, voting site, or other elections property or equipment. This clause is presumed when an image or audio or video file has been intentionally manipulated to represent a false portrayal of the candidate for elective office, the elected official, the elections official, or the voting machine, ballot, voting site, or other elections property or equipment, but may be rebutted.

(B) "Materially deceptive and digitally modified or created image or audio or video file" does not include any image or audio or video file that contains only minor modifications that do not lead to significant changes to the perceived contents or meaning of the content. Minor changes include changes to brightness or contrast of images, removal of background noise in audio, and other minor changes that do not impact the content of the image or audio or video file.

~~(5)~~

(6) "Officer holding an election or conducting a canvass" has the same meaning as in Section 18502.

~~(6)~~

(7) "Recipient" includes a person who views, hears, or otherwise perceives an image or audio or video file that was initially distributed in violation of this section.

(g) The provisions of this section apply regardless of the language used in the advertisement or solicitation. If the language used is not English, the disclosure required by paragraph (2) of subdivision (a) shall appear in the language used in the advertisement or solicitation.

(h) The provisions of this section are severable. If any provision of this section or its application is held invalid, that invalidity shall not affect other provisions or applications that can be given effect without the invalid provision or application.

O

AMENDED IN ASSEMBLY APRIL 18, 2024

AMENDED IN ASSEMBLY MARCH 21, 2024

CALIFORNIA LEGISLATURE—2023–24 REGULAR SESSION

ASSEMBLY BILL No. 3211

Introduced by Assembly Member Wicks

February 16, 2024

An act to add Chapter 41 (commencing with Section 22949.90) to Division 8 of the Business and Professions Code, relating to artificial intelligence.

LEGISLATIVE COUNSEL'S DIGEST

AB 3211, as amended, Wicks. California Provenance, Authenticity and Watermarking Standards.

Existing law requires the Secretary of Government Operations to develop a coordinated plan to, among other things, investigate the feasibility of, and obstacles to, developing standards and technologies for state departments to determine digital content provenance. For the purpose of informing that coordinated plan, existing law requires the secretary to evaluate, among other things, the impact of the proliferation of deepfakes, as defined.

Beginning February 1, 2025, this bill, the California Provenance, Authenticity and Watermarking Standards Act, would require a generative artificial intelligence (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.~~ *provider to, among other things,* *place imperceptible and maximally indelible watermarks containing provenance data into synthetic content produced or significantly modified by a generative AI system that the provider makes available, as those terms are defined.* The bill would require, within 24 hours of discovering a vulnerability or failure in ~~an~~ *a generative* AI system, a generative AI ~~system~~ provider to report the vulnerability or failure to the Department of Technology and to notify other generative AI ~~system~~ providers, as specified. The bill would also require a conversational AI system, as defined, to clearly and prominently disclose to users that the conversational AI system receives synthetic content.

Beginning March 1, 2025, this bill would require 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.~~ *require a user that uploads or distributes content on its platform to disclose whether the content is synthetic content, as specified.*

Beginning January 1, 2026, this bill would require newly manufactured digital cameras and recording devices sold, offered for sale, or distributed in California to offer users the option to place an authenticity watermark and provenance watermark in the content produced by that device. The bill would require the authenticity watermark and provenance watermark to be compatible with widely used industry ~~standards, as specified. If a digital camera or recording device purchased in California prior to January 1, 2026, is capable of receiving a software or firmware update that would enable a user to place an authenticity watermark and provenance watermark on the content created by the device, the bill would require the device's manufacturer to offer that update.~~ *standards. If technically feasible, the bill would require a camera and recording device manufacturer, as defined, to offer to a user of a digital camera or recording device purchased in California prior to January 1, 2026, a software or firmware update enabling the user to place an authenticity watermark and provenance watermark on the content created by the device.*

Beginning January 1, 2026, and annually thereafter, this bill would also require ~~specified entities~~ *generative AI providers and large online* *platforms* to produce a Risk Assessment and Mitigation Report that assesses the risks posed~~ by synthetic content and the harms that have been or could be caused by synthetic content,~~ *and harms caused by synthetic content generated in their systems or hosted on their platforms,* as prescribed. *The bill would require the report to be audited by qualified, independent auditors who are required to assess and either validate or invalidate the claims made in the report, as specified.*

This bill would provide that a violation of its provisions may result in an administrative penalty, assessed by the department, of up to $1,000,000 or 5% of the violator's annual global revenue, whichever is greater. The bill would require the department to adopt regulations as necessary to implement and carry out the purposes of this act and to review and update those regulations as needed.

Vote: majority. Appropriation: no. Fiscal committee: yes. State-mandated local program: no.

The people of the State of California do enact as follows:

SECTION 1. The Legislature finds and declares all of the following:

(a) Generative artificial intelligence (GenAI) technologies are increasingly able to produce inauthentic images, audio, video, and text content in ways that are harmful to society.

(b) In order to reduce the severity of the harms caused by GenAI, it is important for GenAI content to be clearly disclosed and labeled.

(c) Failing to appropriately label GenAI content can skew election results, enable academic dishonesty, and erode trust in the online information ecosystem.

(d) The Legislature should act to adopt standards pertaining to the clear disclosure and labeling of GenAI content, in order to alleviate harms caused by the misuse of these technologies.

(e) The Legislature should push for the creation of tools that allow Californians to assess the authenticity of online content.

(f) The Legislature should require online platforms to label inauthentic content produced by GenAI.

(g) Through these actions, the Legislature can help to ensure that Californians remain safe and informed.

SEC. 2. Chapter 41 (commencing with Section 22949.90) is added to Division 8 of the Business and Professions Code, to read:

Chapter 41. California Provenance, Authenticity, and Watermarking Standards

22949.90. For purposes of this chapter, the following definitions apply:

(a) "AI red-teaming" means a structured testing effort to find flaws and vulnerabilities in an AI system, including, but not limited to, harmful or discriminatory outputs, unforeseen or undesirable system behaviors, limitations, or potential risks associated with misuse of the system.

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

(c) "Authentic content" means images, videos, audio, or text created by human beings without any modifications or with only minor modifications that do not lead to significant changes to the perceived contents or meaning of the content. Minor modifications include, but are not limited to, changes to brightness or contrast of images, removal of background noise in audio, and spelling or grammar corrections in text.

(d) "Conversational AI system" means chatbots and other audio- or video-based systems that can hold humanlike conversations through digital media, including, but not limited to, online calling, phone calling, video conferencing, messaging, application or web-based chat interfaces, or other conversational interfaces. Conversational AI systems include, but are not limited to, chatbots for customer service or entertainment purposes embedded in internet websites and applications.

(e) "Digital signature" means a digital method that allows a user to sign a piece of authentic or synthetic content with their name or device information, verifying that they created the content.

(f) "Generative AI hosting platform" means an online repository or other internet website that makes generative AI systems available for download.

(g) "Generative AI provider" means an organization or individual that creates, codes, substantially modifies, or otherwise produces a generative AI system.

(h) "Generative AI system" means an artificial intelligence system that generates derived synthetic content, including images, videos, audio, text, and other digital content.

(i) "Inauthentic content" means synthetic content that is so similar to authentic content that it could be mistaken as authentic.

(j) "Large online platform" means a public-facing internet website, web application, or digital application, including a social network, video sharing platform, messaging platform, advertising network, or search engine that had at least 1,000,000 California users during the preceding 12 months.

(k) "Maximally indelible watermark" means a watermark that is designed to be as difficult to remove as possible using state-of-the-art techniques and relevant industry standards.

(l) "Provenance data" means data that identifies the origins of synthetic content, including, but not limited to, the following:

(1) The name of the generative AI provider.

(2) The name and version number of the AI system that generated the content.

(3) The time and date of the creation.

(4) The portions of content that are synthetic.

(m) "Synthetic content" means information, including images, videos, audio, and text, that has been produced or significantly modified by a generative AI system.

(n) "Watermark" means information that is embedded into a generative AI system's output for the purpose of conveying its synthetic nature, identity, provenance, history of modifications, or history of conveyance.

(o) "Watermark decoders" means freely available software tools or online services that can read or interpret watermarks and output the provenance data embedded in them.

22949.90.1. (a) A generative AI provider shall do all of the following:

(1) Place imperceptible and maximally indelible watermarks containing provenance data into synthetic content produced or significantly modified by a generative AI system that the provider makes available.

(A) If a sample of synthetic content is too small to contain the required provenance data, the provider shall, at minimum, attempt to embed watermarking information that identifies the content as synthetic and provide the following provenance information in

order of priority, with clause (i) being the most important, and clause (iv) being the least important:

(i) The name of the generative AI provider.

(ii) The name and version number of the AI system that generated the content.

(iii) The time and date of the creation of the content.

(iv) If applicable, the specific portions of the content that are synthetic.

(B) To the greatest extent possible, watermarks shall be designed to retain information that identifies content as synthetic and gives the name of the provider in the event that a sample of synthetic content is corrupted, downscaled, cropped, or otherwise damaged.

(2) Develop downloadable watermark decoders that allow a user to determine whether a piece of content was created with the provider's system, and make those tools available to the public.

(A) The watermark decoders shall be easy to use by individuals seeking to quickly assess the provenance of a single piece of content.

(B) The watermark decoders shall adhere, to the greatest extent possible, to relevant national or international standards.

(3) Conduct AI red-teaming exercises involving third-party experts to test whether watermarks can be easily removed from synthetic content produced by the provider's generative AI systems, as well as whether the provider's generative AI systems can be used to falsely add watermarks to otherwise authentic content. Red-teaming exercises shall be conducted before the release of any new generative AI system and annually thereafter.

(A) If a provider allows their generative AI systems to be downloaded and modified, the provider shall additionally conduct AI red-teaming to assess whether their systems' watermarking functionalities can be disabled.

(B) A provider shall make summaries of its AI red-teaming exercises publicly available in a location linked from the home page of the provider's internet website, using a clearly labeled link that has a similar look, feel, and size relative to other links on the same web page. The provider shall remove from the summaries any details that pose an immediate risk to public safety.

(C) A provider shall submit full reports of its AI red-teaming exercises to the Department of Technology within six months of conducting a red-teaming exercise pursuant to this section.

(b) A generative AI provider may continue to make available a generative AI system that was made available before the date upon which this act takes effect and that does not have watermarking capabilities as described by paragraph (1) of subdivision (a), if either of the following conditions are met:

(1) The provider is able to retroactively create and make publicly available a decoder that accurately determines whether a given piece of content was produced by the provider's system with at least 99 percent accuracy as measured by an independent auditor.

(2) The provider conducts and publishes research to definitively demonstrate that the system is not capable of producing inauthentic content.

(c) Providers and distributors of software and online services shall not make available a system, application, tool, or service that is designed to remove watermarks from synthetic content.

(d) Generative AI hosting platforms shall not make available a generative AI system that does not place maximally indelible watermarks containing provenance data into content created by the system.

(e) (1) Within 24 hours of discovering a vulnerability or failure in a generative AI system, a generative AI provider shall report the vulnerability or failure to the Department of Technology.

(A) A provider shall notify other generative AI providers that may be affected by similar vulnerabilities or failures in a manner that allows the other generative AI provider to harden their own AI systems against similar risks.

(B) A provider shall notify affected parties, including, but not limited to, online platforms, researchers or users who received incorrect results from a watermark decoder, or users who produced AI content that contained incorrect or insufficient watermarking data. A provider shall not be required to notify an affected party whose contact information the provider has not previously collected or retained.

(2) A provider shall make any report to the Department of Technology under this subdivision publicly available in a location linked from the home page of the provider's internet website with a clearly labeled link that has a similar look, feel, and size relative to other links on the same web page. If public disclosure of the

report under this subdivision could pose public safety risks, a provider may instead do either of the following:

(A) Post a summary disclosure of the reported vulnerability or failure.

(B) Delay, for no longer than 30 days, the public disclosure of the report until the public safety risks have been mitigated. If a provider delays public disclosure, they shall document their efforts to resolve the vulnerability or failure as quickly as possible in order to meet the reporting requirements under this subdivision.

(f) (1) A conversational AI system shall clearly and prominently disclose to users that the conversational AI system generates synthetic content.

(A) In visual interfaces, including, but not limited to, text chats or video calling, a conversational AI system shall place the disclosure required under this subdivision in the interface itself and maintain the disclosure's visibility in a prominent location throughout any interaction with the interface.

(B) In audio-only interfaces, including, but not limited to, phone or other voice calling systems, a conversational AI system shall verbally make the disclosure required under this subdivision at the beginning and end of a call.

(2) In all conversational interfaces of a conversational AI system, the conversational AI system shall, at the beginning of a user's interaction with the system, obtain a user's affirmative consent acknowledging that the user has been informed that they are interacting with a conversational AI system. A conversational AI system shall obtain a user's affirmative consent prior to beginning the conversation.

(3) Disclosures and affirmative consent opportunities shall be made available to a user in the language in which the conversational AI system is communicating with the user.

(4) The requirements under this subdivision shall not apply to conversational AI systems that do not produce inauthentic content.

(g) This section shall become operative on February 1, 2025.

22949.90.2. (a) For purposes of this section, the following definitions apply:

(1) "Authenticity watermark" means a watermark of authentic content that includes the name of the device manufacturer.

(2) "Camera and recording device manufacturer" means the makers of a device that can record photographic, 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.

(3) “Provenance watermark” means a watermark of authentic content that includes details about the content, including, but not limited to, the time and date of production, the name of the user, details about the device, and a digital signature.

(b) (1) Beginning January 1, 2026, newly manufactured digital cameras and recording devices sold, offered for sale, or distributed in California shall offer users the option to place an authenticity watermark and provenance watermark in the content produced by that device.

(2) A user shall have the option to remove the authenticity and provenance watermarks from the content produced by their device.

(3) Authenticity watermarks shall be turned on by default, while provenance watermarks shall be turned off by default.

(4) Newly manufactured digital cameras and recording devices subject to the requirements of this subdivision shall clearly inform a user of the existence of the authenticity and provenance watermarks settings upon the user’s first use of the camera or the recording function on the recording device.

(A) When a camera or audio recording application is open, a newly manufactured digital camera or recording device shall have a clear indicator that a watermark is being applied.

(B) A newly manufactured digital camera or recording device shall allow the user to adjust the watermarks settings.

(5) Authenticity and provenance watermarks shall, if enabled, be applied to authentic content produced using third-party applications that bypass default camera or recording applications in order to offer camera or audio recording functionalities.

(c) The authenticity watermark and provenance watermark, as required by subdivision (b), shall be compatible with widely used industry standards.

(d) Beginning January 1, 2026, a camera and recording device manufacturer shall offer a software or firmware update enabling a user to place an authenticity watermark and provenance watermark on the content created by the device to a user of a digital camera or recording device purchased in California prior to January 1, 2026, if technically feasible.

22949.90.3. (a) Beginning March 1, 2025, a large online platform shall use labels to prominently disclose the provenance data found in watermarks or digital signatures in content distributed to users on its platforms.

(1) The labels shall indicate whether content is fully synthetic, partially synthetic, authentic, authentic with minor modifications, or does not contain a watermark.

(2) A user shall be able to click or tap on a label to inspect provenance data in an easy-to-understand format.

(b) The disclosure required under subdivision (a) shall be readily legible to an average viewer or, if the content is in audio format, shall be clearly audible. A disclosure in audio content shall occur at the beginning and end of a piece of content and shall be presented in a prominent manner and at a comparable volume and speaking cadence as other spoken words in the content. A disclosure in video content should be legible for the full duration of the video.

(c) A large online platform shall use state-of-the-art techniques to detect and label synthetic content that has had watermarks removed or that was produced by generative AI systems without watermarking functionality.

(d) (1) A large online platform shall require a user that uploads or distributes content on its platform to disclose whether the content is synthetic content.

(2) A large online platform shall include prominent warnings to users that uploading or distributing synthetic content without disclosing that it is synthetic content may result in disciplinary action.

(3) A large online platform may provide users with an option to indicate that the user is uncertain whether the content they are uploading or distributing is synthetic content. If a user uploads or distributes content and indicates that they are uncertain of whether the content is synthetic content, a large online platform shall indicate that the uploaded or distributed content is possibly synthetic and shall display that indication in a manner that is visible or audible to viewers or listeners of the content.

(e) A large online platform shall use state-of-the-art techniques to detect and label text-based inauthentic content that is uploaded by users.

(f) A large online platform shall make accessible a verification process for users to apply a digital signature to authentic content. The verification process shall include options that do not require disclosure of personal identifiable information.

22949.90.4. (a) (1) Beginning January 1, 2026, and annually thereafter, generative AI providers and large online platforms shall produce a Risk Assessment and Mitigation Report that assesses the risks posed and harms caused by synthetic content generated by their systems or hosted on their platforms.

(2) The report shall include, but not be limited to, assessments of the distribution of AI-generated child sexual abuse materials, nonconsensual intimate imagery, disinformation related to elections or public health, plagiarism, or other instances where synthetic or inauthentic content caused or may have the potential to cause harm.

(b) The report required under subdivision (a) shall be audited by qualified, independent auditors who shall assess and either validate or invalidate the claims made in the report. Auditors shall use state-of-the-art techniques to assess reports, and shall adhere to relevant national and international standards.

22949.90.5. A violation of this chapter may result in an administrative penalty, assessed by the Department of Technology, of up to one million dollars ($1,000,000) or 5 percent of the violator's annual global revenue, whichever is greater.

22949.90.6. Within 90 days of the date upon which this act takes effect, the Department of Technology shall adopt regulations to implement and carry out the purposes of this chapter. The department shall review and update its regulations relating to the implementation of this chapter as needed, including, but not limited to, adopting specific national or international standards for provenance, authenticity, watermarking, and digital signatures, as long as the standards do not weaken the provisions of this chapter.

22949.91. The provisions of this chapter are severable. If any provision of this chapter or its application is held invalid, that invalidity shall not affect other provisions or applications that can be given effect without the invalid provision or application.

~~SECTION 1. The Legislature finds and declares all of the following:~~

~~(a) In light of the widespread adoption of generative artificial intelligence (AI) technologies that are increasingly able to generate inauthentic images, audio, video, and text content, sometimes called “deepfakes,” it is increasingly important that the provenance of this content be clearly disclosed and labeled in ways that could both prevent harms, reduce the severity of harms, and also make it more difficult and costly for bad actors to cause these harms.~~

~~(b) (1) The harms caused by inauthentic content that is presented as authentic span a wide gamut of fields. These harms can be costly and deeply damaging to individuals and society. Some prominent categories of harm include scams and fraud, child sexual abuse material (CSAM) and nonconsensual intimate imagery (NCII), disinformation, and plagiarism and academic integrity.~~

~~(A) The Federal Trade Commission reports that people lost $8.8 billion to scams in 2022. This is only expected to increase with broad access to generative AI tools, with one study already finding a 1,265-percent increase in phishing scam emails since the fourth quarter of 2022.~~

~~(B) A recent study found that a version of one of the most popular AI image generator tools was trained on CSAM, which facilitates production in derivative models of both CSAM and NCII of children. Another recent study found 34 different AI “undressing” tools that produce realistic naked images based on clothed photographs of individuals. Many of these tools only work on women. One recent case involved a high school where 30 different teenage girls were victims of these tools.~~

~~(C) From politics to public safety, disinformation can have a multiplicity of negative impacts on society. It can skew election results, as may have been the case with a convincing audio deepfake in Slovakia. In the lead up to the 2024 United States elections, we have already seen candidate- and party-produced advertisements using deepfake audio and video content. A deepfake of an explosion at the Pentagon caused a brief but significant stock market dip.~~

~~(D) Educational institutions have struggled to adopt clear best practices for assigning students at-home writing assignments since the widespread use of high-quality AI text generation tools.~~

~~(2) More broadly, erosion of trust in the information ecosystem, sometimes referred to as “truth decay,” can be shown to increase polarization and stands to be greatly exacerbated by our ongoing~~

~~failure to establish norms for clearly disclosing and labeling the provenance of digital media.~~

~~(c) (1) For the above-described reasons, this act provides for the phased introduction of the California Provenance, Authenticity and Watermarking Standards (PAWS), which aim to provide the public with tools to understand how the content that they see across digital media was produced and if it is, in fact, authentic.~~

~~(2) The PAWS require all producers of media to embed maximally indelible and privacy-preserving content provenance data into the content that they generate, whether AI generated or authentic. The PAWS also require all large online platforms to display clearly understandable labels on content that alert users to its provenance, or to the absence of available provenance data.~~

~~(3) The provenance data required by PAWS is narrowly tailored to provide consumers with factual information about the authenticity or inauthenticity of images, audio, video, or text content in order to prevent consumer deception.~~

~~(4) While the PAWS alone will not fully eliminate the harms described above, it has the capacity to dramatically reduce these harms by signaling to the would-be audiences for and victims of these harms that inauthentic content will not be easily mistaken for authentic content. The PAWS will also significantly reduce the burden on large online platforms that would like to display content provenance data to their users, but are not able to do so because of the lack of an industrywide standard for embedding of provenance data in their content.~~

~~SEC. 2. Chapter 41 (commencing with Section 22949.90) is added to Division 8 of the Business and Professions Code, to read:~~

~~Chapter 41. California Provenance, Authenticity and Watermarking Standards~~

~~22949.90. For purposes of this chapter, the following definitions apply:~~

~~(a) “AI red-teaming” means a structured testing effort to find flaws and vulnerabilities in an AI system, often in a controlled environment and in collaboration with developers of artificial intelligence (AI) and is most often performed by dedicated “red teams” that adopt adversarial methods to identify flaws and vulnerabilities, including, but not limited to, harmful or~~

~~discriminatory outputs from an AI system, unforeseen or undesirable system behaviors, limitations, or potential risks associated with misuse of the system.~~

~~(b) "AI system" means a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs, including, but not limited to, predictions, content, recommendations, or decisions that may influence physical or virtual environments.~~

~~(c) "Authentic content" means images, videos, audio clips, or text created by human beings without any modifications or with only minor modifications that do not lead to significant changes to the perceived contents or meaning of the content. Minor modifications include, but are not limited to, changes to brightness or contrast of images, removal of background noise in audio, and spelling or grammar corrections in text.~~

~~(d) "Camera and recording device manufacturers" means the makers of 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.~~

~~(e) "Conversational AI systems" means chatbots and other audio- or video-based systems that can hold humanlike conversations through digital media, including, but not limited to, online calling, phone calling, video conferencing, messaging, application or web-based chat interfaces, or other conversational interfaces. Conversational AI systems includes chatbots for customer service or entertainment purposes embedded in internet websites and applications.~~

~~(f) "Digital signature" means a digital method that allows a user to sign a piece of authentic or synthetic content with their name or device information, verifying that they created the content, and that is included in the content's provenance data.~~

~~(g) "Generative AI system" means the class of AI models that emulates the structure and characteristics of input data to generate derived synthetic content, including images, videos, audio, text, and other digital content. The synthetic content generated may, but does not necessarily have to, be of the same content type as the input data.~~

~~(h) "Generative AI system distributors" means organizations or individuals that distribute generative AI systems, or substantial~~

~~components thereof, such as model weights, in ways that can be downloaded and used by individuals locally on their own hardware, or modified or incorporated into other products or services.~~

~~(i) "Generative AI system providers" means organizations or individuals that make AI systems, or substantial components thereof, available on the market, put them into service, provide them as standalone models or embed them in other systems or products, or provide them under free and open source licenses as a service, or through other large online platforms. Generative AI system distributors include repositories or hosting internet websites that make AI systems available for download, even if those repositories are not the original makers of the AI systems they make available, and providers of conversational AI systems.~~

~~(j) "Inauthentic content" means synthetic content that is so similar to authentic content that it could be mistaken as authentic.~~

~~(k) "Large online platform" means a public-facing internet website, web application, or digital application, including a social network, video sharing platform, messaging platform, advertising network, or search engine that had at least 1,000,000 California users during the preceding 12 months.~~

~~(*l*) "Legacy generative AI systems" means generative AI systems created and released before the date upon which this act takes effect.~~

~~(m) "Maximally indelible watermarks" means watermarks that are as difficult to remove as is possible with currently available techniques and that are tested through AI red-teaming.~~

~~(n) "Provenance data" means data that includes the name of the AI system that generated the content, the underlying AI models that were part of the AI system, the time and date of the creation of the content, and, if applicable, which specific portions of the content are synthetic.~~

~~(o) "Synthetic content" means information, including images, videos, audio clips, and text, that has been significantly modified or generated by algorithms, including by AI.~~

~~(p) "Tamper-evident" means a type of watermark that contains provenance data that cannot be modified without leaving evidence of tampering.~~

~~(q) "Watermark" means embedded information, typically difficult to remove, into outputs created, including into photographs, videos, audio clips, or text, for the purposes of~~

~~verifying the authenticity of the output or the identity or characteristics of its provenance, modifications, or conveyance.~~

~~(r) "Watermark decoders" means freely available software tools or online services that can read or interpret watermarks and output the provenance data embedded in them.~~

~~22949.90.1. (a) A generative AI system provider shall do all of the following:~~

~~(1) Place imperceptible and maximally indelible watermarks containing provenance data into content created by an AI system that the generative AI system provider makes available. If the content is manipulated in ways that are intended to remove a watermark, or if a content sample, including a cropped portion of an image, a truncated segment of a generated piece of audio or video content, or a small amount of text content, is damaged or becomes too small to contain the required provenance data, the generative AI system provider shall, at minimum, embed watermarking information that identifies the content as synthetic and gives the name of the generative AI system provider.~~

~~(2) Provide downloadable software tools or online services to determine whether a piece of content was created with the provider's system and make those tools available to all large online platforms and the public.~~

~~(A) A generative AI system provider shall produce the software tools or online services in ways that are easy to use manually by individuals seeking to quickly assess the provenance of a single piece of content and to use in an efficient and automated fashion by online platforms or researchers seeking to assess the provenance of hundreds or thousands of pieces of content per minute.~~

~~(B) A generative AI system provider shall produce the software tools or online services to be interoperable to the greatest extent possible with decoders made available by other providers or organizations. The decoders shall adhere, to the greatest extent possible, to relevant national or international standards, if available.~~

~~(3) Conduct AI red-teaming exercises that involve third-party experts to test whether watermarks can be easily removed from their synthetic content and whether AI systems could be used to falsely add watermarks to authentic content that indicates that it is inauthentic.~~

~~(A) If a generative AI system provider intends to allow their systems to be downloaded and modified, the provider shall conduct~~

~~AI red-teaming to assess whether the systems' watermarking functionalities can be disabled to generate deceptive, inauthentic content.~~

~~(B) A generative AI system provider shall make summaries of its AI red-teaming exercises publicly available in a location linked from the home page of the generative AI system provider's internet website with a clearly labeled link that has a similar look, feel, and size relative to other links on the same web page and shall remove from the summaries details that may pose immediate public security. A generative AI system provider shall submit full reports of its AI red-teaming exercises to the Department of Technology within six months from the date upon which this act takes effect, and annually thereafter.~~

~~(b) A generative AI system provider may continue providing legacy generative AI systems that do not have watermarking capabilities as described by paragraph (1) of subdivision (a) if either of the following applies:~~

~~(1) The generative AI system provider is able to retrospectively create and make publicly available a decoder that accurately determines whether content generated by the legacy generative AI system is synthetic with at least 99 percent accuracy as measured by an independent auditor.~~

~~(2) The generative AI system provider conducts and publishes research to definitively demonstrate that the legacy generative AI system is not capable of producing inauthentic content of a quality that is sufficiently realistic to be potentially mistaken for authentic content.~~

~~(c) A generative AI system provider, generative AI system distributor, or other provider, or a distributor of software or online services shall not make available or distribute an AI system, application, tool, or service that has the capacity to remove watermarks from synthetic content or an AI system that has the capacity to remove watermarks from synthetic content but was not explicitly designed for that purpose.~~

~~(d) (1) If a generative AI system provider distributes a generative AI system in a way that allows the generative AI system to be modified by others, the generative AI system provider shall ensure that the generative AI system does not allow for removal of the system's watermarking functionality.~~

~~(2) A generative AI system provider shall not distribute a generative AI system or make a generative AI system available for use if the system's watermarking functionality can be removed by others.~~

~~(e) (1) Within 24 hours of discovering a vulnerability or failure in an AI system, a generative AI system provider shall report the vulnerability or failure to the Department of Technology. A generative AI system provider shall also notify other generative AI system providers that may be affected by similar vulnerabilities or failures in a manner that allows the other generative AI system provider to harden their own AI systems against similar risks.~~

~~(2) A generative AI system provider shall also notify any affected party. Affected parties include, but are not limited to, an online platform, researchers or users who received incorrect results from a watermark decoder, or users who produced AI content that contained incorrect or insufficient watermarking data.~~

~~(3) A generative AI system provider shall make any report to the Department of Technology under this subdivision publicly available in a location linked from the home page of the generative AI system provider's internet website with a clearly labeled link that has a similar look, feel, and size relative to other links on the same web page.~~

~~(4) If public disclosure of the report under this subdivision could pose public safety risks, a generative AI system provider may instead do either of the following:~~

~~(A) Post a summary disclosure of the reported vulnerability or failure.~~

~~(B) Delay, for no longer than 30 days, the public disclosure of the report until the public safety risks have been mitigated. If a generative AI system provider delays public disclosure, the provider shall document and demonstrate that they moved as quickly as possible to resolve the vulnerability or failure and meet the reporting requirements under this subdivision.~~

~~(f) (1) A conversational AI system shall clearly and prominently disclose to users that the conversational AI system receives synthetic content.~~

~~(A) In visual interfaces, including, but not limited to, text chats or video calling, a conversational AI system shall place the disclosure required under this subdivision in the interface itself~~

~~and maintain the disclosure's visibility in a prominent location throughout any interaction with the interface.~~

~~(B) In audio-only interfaces, including, but not limited to, phone or other voice calling systems, a conversational AI system shall verbally make the disclosure required under this subdivision at the beginning and end of a call.~~

~~(2) In all conversational interfaces of a conversational AI system, the conversational AI system shall, at the beginning of a user's interaction with the system, obtain a user's affirmative consent acknowledging that the user has been informed that they are interacting with a conversational AI system. A conversational AI system shall obtain a user's affirmative consent prior to beginning the conversation.~~

~~(3) Disclosures and affirmative consent opportunities shall be made available to a user in the language in which the conversational AI system is communicating with the user.~~

~~(4) The requirements under this subdivision shall not apply to conversational AI systems that produce content that could not be reasonably mistaken as authentic.~~

~~(g) This section shall become operative on February 1, 2025.~~

~~22949.90.2. (a) For purposes of this section, the following definitions apply:~~

~~(1) "Authenticity watermark" means a watermark of authentic content that includes the name of the device manufacturer.~~

~~(2) "Provenance watermark" means a watermark of authentic content that includes details about the content, including, but not limited to, the time and date of production, the name of the user, details about the device, and a digital signature.~~

~~(b) (1) Beginning January 1, 2026, newly manufactured digital cameras and recording devices sold, offered for sale, or distributed in California shall offer users the option to place an authenticity watermark and provenance watermark in the content produced by that device.~~

~~(2) A user shall have the option to remove the authenticity and provenance watermarks from the content produced by their device.~~

~~(3) Authenticity watermarks shall be turned on by default, while provenance watermarks shall be turned off by default.~~

~~(4) Newly manufactured digital cameras and recording devices subject to the requirements of this subdivision shall clearly inform a user of the existence of the authenticity and provenance~~

~~watermarks settings upon the user’s first use of the camera or the recording function on the recording device.~~

~~(A) When a camera or audio recording application is open, a newly manufactured digital camera or recording device shall have a clear indicator that a watermark is being applied.~~

~~(B) A newly manufactured digital camera or recording device shall allow the user to adjust the watermarks settings.~~

~~(5) The authenticity watermark and provenance watermark shall persist in content produced using newly manufactured digital cameras or recording devices even if the content is created or recorded in third-party applications that offer camera or audio recording functionalities.~~

~~(c) The authenticity watermark and provenance watermark, as required in subdivision (b), shall be compatible with widely used industry standards, including the Coalition for Content Provenance and Authenticity’s content credentials.~~

~~(d) Beginning January 1, 2026, a camera and recording device manufacturer shall offer a software or firmware update enabling a user to place an authenticity watermark and provenance watermark on the content created by the device to a user of a digital camera or recording device purchased in California prior to January 1, 2026. This requirement shall only apply if a digital camera or recording device purchased in California prior to January 1, 2026, is capable of receiving a software or firmware update that would enable the user to place an authenticity watermark and provenance watermark on the content created by the device.~~

~~22949.90.3. (a) Beginning March 1, 2025, a large online platform shall use labels to prominently disclose the provenance data found in watermarks or digital signatures in content distributed to users on its platforms by making use of the data contained in watermarks and digital signatures embedded in content using widely used industry standards, including the Coalition for Content Provenance and Authenticity’s content credentials, and synthetic content decoders provided by generative AI system providers.~~

~~(1) The labels shall indicate whether content is fully synthetic, partially synthetic, authentic, authentic with minor modifications, or does not contain a watermark.~~

~~(2) A user shall be allowed to click or tap on a label to inspect all available provenance data in an easy-to-understand format.~~

~~(b) The disclosure required under subdivision (a) shall be readily legible to an average viewer or, if the content is in audio format, shall be clearly audible. A disclosure in audio and video content shall occur at the beginning and end of a piece of content and shall be presented in a prominent manner and at a comparable volume and speaking cadence as other spoken words in the content.~~

~~(c) A large online platform shall use state-of-the-art techniques to detect and label synthetic content that has had watermarks removed or was produced by AI systems without watermarking functionality.~~

~~(d) (1) A large online platform shall require a user that uploads or distributes content on its platform to disclose whether the uploaded or distributed content is synthetic content.~~

~~(2) A large online platform shall include prominent warnings to users that uploading or distributing synthetic content without disclosing that it is synthetic content is not permissible and will result in disciplinary action by the large online platform.~~

~~(3) A large online platform may provide users with an option to indicate that the user is uncertain whether the content they are uploading or distributing is synthetic content. If a user uploads or distributes content and indicates that they are uncertain of whether the content is synthetic content, a large online platform shall indicate that the uploaded or distributed content is possibly synthetic and shall display that indication in a manner that is visible or audible to viewers or listeners of the content.~~

~~(e) (1) A large online platform shall use state-of-the-art techniques to detect and label inauthentic text content that is uploaded or distributed by individual users or networks of users.~~

~~(2) A large online platform may use a variety of methods to detect inauthentic text content, including, but not limited to, the following:~~

~~(A) Bulk analysis of collected text content from users or networks of users.~~

~~(B) Analysis of user behavioral signals indicating usage of synthetic content.~~

~~(C) Assessing large quantities of text generated by users or networks of users for watermarked content.~~

~~(D) Considering whether a user's typing cadence indicates authenticity or automation.~~

~~(E) Verification that users are matched to unique device identifications such as a subscriber identity module (SIM) card, international mobile equipment identity (IMEI), or multifactor authentication (MFA).~~

~~(3) A large online platform shall also consider account age, login frequency, connection to other identity verification services, frequency of content uploading or distributing, authenticity of original media content, and other on-platform behaviors by a user that could be used to detect undisclosed inauthentic content production.~~

~~(4) If a large online platform discovers that a user has uploaded or distributed inauthentic content and the user did not disclose that the uploaded or distributed content is synthetic content pursuant to subdivision (d), the large online platform shall disable the account of the user that uploaded or distributed the undisclosed inauthentic content.~~

~~(f) A large online platform shall make accessible a verification process for users to apply a digital signature to content created by a human being. The verification process shall include options to verify in a variety of methods that do not necessarily require disclosure of personal identifiable information, including, but not limited to, uploading a government-issued identification and matching picture identification or verifying that a user possesses a unique device with a SIM card and active phone number.~~

~~22949.90.4. (a) (1) Beginning January 1, 2026, and annually thereafter, generative AI system providers, generative AI system distributors, and large online platforms shall produce a Risk Assessment and Mitigation Report that assesses the risks posed by synthetic content and the harms that have been or could be caused by synthetic content.~~

~~(2) The report shall include, but is not limited to, assessments on the distribution of AI-generated child sexual abuse materials, nonconsensual intimate imagery, disinformation related to elections or public health, plagiarism, or other instances where synthetic or inauthentic content caused or may have the potential to cause harm.~~

~~(3) The report shall be audited by qualified, independent auditors who shall assess and either validate or invalidate the claims made in the report. An auditor shall assess the report by using state-of-the-art techniques and adhering to national and~~

~~international standards for the auditing of AI systems as they become available.~~

~~(b) The Department of Technology may authorize independent researchers associated with educational institutions or civil society organizations approved by the Department of Technology to access special researcher tools designed to facilitate large-scale and efficient analysis of content and application programming interfaces (APIs) from generative AI system providers and large online platforms for the purposes of generating test content, studying the efficacy of labeling and effects on users, and evaluating the overall effectiveness of this chapter in preventing harms caused by inauthentic content.~~

~~22949.90.5. A violation of this chapter may result in an administrative penalty, assessed by the Department of Technology, of up to one million dollars ($1,000,000) or 5 percent of the violator's annual global revenue, whichever is greater.~~

~~22949.90.6. Within 90 days of the date upon which this act takes effect, the Department of Technology shall adopt regulations as necessary to implement and carry out the purposes of this chapter. The department shall review and update its regulations relating to the implementation of this chapter as needed, including, but not limited to, adopting specific national or international standards for provenance, authenticity, watermarking, and digital signatures, so long as the standards do not weaken the provisions of this chapter.~~

~~22949.91. The provisions of this chapter are severable. If any provision of this chapter or its application is held invalid, that invalidity shall not affect other provisions or applications that can be given effect without the invalid provision or application.~~

O

AMENDED IN SENATE APRIL 25, 2024

AMENDED IN SENATE APRIL 10, 2024

AMENDED IN SENATE MARCH 18, 2024

SENATE BILL No. 1228

Introduced by Senator Padilla

February 15, 2024

An act to add Chapter 22.9 (commencing with Section 22684) to Division 8 of the Business and Professions Code~~, and to amend Section 35 of the Code of Civil Procedure,~~ relating to ~~online~~ *social media* platforms.

LEGISLATIVE COUNSEL'S DIGEST

SB 1228, as amended, Padilla. ~~Large online platforms: user~~ *User* identity authentication.

Existing law generally regulates online platforms, including by requiring a social media company, as defined, to post terms of service for each social media platform owned or operated by the social media company in a manner reasonably designed to inform all users of the social media platform of the existence and contents of the terms of service.

~~Existing law requires certain actions in cases involving the registration or denial of registration of voters, among others, to be placed on the calendar in the order of their date of filing and given precedence.~~

This bill would require a ~~large online~~ *social media* platform, as defined, to seek to verify the name, telephone number, and email address of an influential user, as defined, by a means chosen by the ~~large online~~ *social media* platform and would require the *social media* platform to seek to verify the identity of a highly influential user, as defined, by asking to review the highly influential user's government-issued identification.

This bill would require a ~~large online~~ *social media* platform to note on the profile page of an influential or highly influential user, in type at least as large and as visible as the user's name, whether the user has been authenticated pursuant to those provisions, as prescribed, and would require the platform to attach to any post of an influential or highly influential user a notation that would be understood by a reasonable person as indicating that the user is authenticated or unauthenticated, as prescribed.

This bill would authorize the Attorney General or any district attorney or city attorney to seek injunctive or other equitable relief against a ~~large online~~ *social media* platform to compel compliance with the ~~bill and would require those actions to be placed on the calendar in the order of their date of filing and given precedence.~~ *bill.*

Vote: majority. Appropriation: no. Fiscal committee: yes. State-mandated local program: no.

The people of the State of California do enact as follows:

SECTION 1. Chapter 22.9 (commencing with Section 22684) is added to Division 8 of the Business and Professions Code, to read:

Chapter 22.9. Verify or Flag

22684. For the purposes of this chapter:

(a) "Highly influential user" means a user of a ~~large online~~ *social media* platform that meets any of the following criteria:

(1) Content authored, created, or produced by the user has been seen by more than 100,000 users within a seven-day period over all of the accounts that they control or administer on the platform.

(2) Accounts controlled or administered by the user have more than 30,000 followers.

(3) The user ranks in the top 3 percent of users by amount of content viewed by users on the platform within a seven-day period over all of the accounts that the user controls or administers on the platform.

(b) "Influential user" means a user of a ~~large online~~ *social media* platform that meets any of the following criteria:

(1) Content authored, created, or produced by the user has been seen by more than 50,000 users within a seven-day period over all of the accounts that the user controls or administers on the platform.

(2) Accounts controlled or administered by the user have more than 15,000 followers.

(3) The user ranks in the top 6 percent of users by amount of content viewed by users on the platform within a seven-day period over all of the accounts that the user controls or administers on the platform.

~~(c) “Large online platform” means a public-facing internet website, web application, or digital application, including a social network, video sharing platform, messaging platform, advertising network, or search engine that had at least 1,000,000 California users during the preceding 12 months.~~

~~(d)~~

(c) “Sensitive personal information” has the same meaning as defined in Section 1798.140 of the Civil Code.

(d) “Social media platform” has the same meaning as defined in Section 22675.

22684.1. (a) (1) (A) A ~~large online~~ *social media* platform shall seek to verify an influential user’s name, telephone number, and email address by a means chosen by the large online platform.

(B) A ~~large online~~ *social media* platform shall seek to verify a highly influential user’s identity by asking to review the highly influential user’s government-issued identification.

(2) A ~~large online~~ *social media* platform shall not use the identification information provided by a user for the purpose of complying with this subdivision for any purpose other than compliance with this subdivision.

(3) (A) Subject to subparagraph (B), a ~~large online~~ *social media* platform shall protect any identification information provided by an influential user in compliance with this section using, at a minimum, the standard of the industry used to protect the confidential information of users unless the ~~large online~~ *social media* platform makes that information public in the normal course of a user’s use of the ~~large online~~ *social media* platform.

(B) A ~~large online~~ *social media* platform shall not allow a user’s sensitive personal information to become public.

(b) A ~~large online~~ *social media* platform shall note on the profile page of an influential or highly influential user, in type at least as large and as visible as the user's name, either of the following:

(1) "This user has been authenticated," or some similar phrase, if the user has complied with the ~~large online~~ *social media* platform's identification process required by subdivision (a).

(2) "This user is unauthenticated," or some similar phrase, if the user has failed to comply with the ~~large online~~ *social media* platform's identification process required by subdivision (a).

(c) (1) A ~~large online~~ *social media* platform shall attach to any post of an influential or highly influential user a notation that would be understood by a reasonable person as indicating that the user is authenticated or unauthenticated.

(2) For a post from an unauthenticated influential or highly influential user, the notation required by paragraph (1) shall be visible for at least two seconds before the rest of the post is visible and then shall remain visible with the post.

(d) A ~~large online~~ *social media* platform shall allow its users to opt out of receiving any posts, information, or other distributions from a user who is not authenticated.

(e) A ~~large online~~ *social media* platform shall maintain proof that it has complied with the verification requirements of this section but may refrain from storing or maintaining the verification information or documentation.

22684.2. (a) The Attorney General or any district attorney or city attorney may seek injunctive or other equitable relief against a ~~large online~~ *social media* platform to compel compliance with this chapter.

(b) In an action filed pursuant to this section, the court shall award a prevailing plaintiff reasonable attorney's fees and costs.

22684.3. (a) This chapter does not preclude a ~~large online~~ *social media* platform from developing and implementing policies that seek to verify the identity of users who are not influential or highly influential users or from verifying the identity of influential or highly influential users with additional proof.

(b) This chapter does not preclude a ~~large online~~ *social media* platform from requiring that its users comply with any identification verification required by the ~~company.~~ *social media platform.*

(c) This chapter does not preclude a ~~large online~~ *social media* platform from requiring that all users, including influential or highly influential users, be identified as either “this user has been authenticated” or “this user is unauthenticated” or similar phrases.

(d) This chapter does not preclude a ~~large online~~ *social media* platform from requiring or prohibiting its users to be publicly identified.

(e) This chapter does not require that a ~~large online~~ *social media* platform provide less exposure or visibility to the posts made or digital media content created by users who decline to provide identity verification.

(f) The provisions of this chapter are severable. If any provision of this chapter or its application is held invalid, that invalidity shall not affect other provisions or applications that can be given effect without the invalid provision or application.

(g) This chapter does not apply to a social media platform that had less than 1,000,000 California users during the preceding 12 months.

~~SEC. 2. Section 35 of the Code of Civil Procedure, as amended by Section 1 of Chapter 343 of the Statutes of 2023, is amended to read:~~

~~35. (a) Proceedings in cases involving the registration or denial of registration of voters, the certification or denial of certification of candidates, the certification or denial of certification of ballot measures, election contests, actions under Section 20010 of the Elections Code, actions under Chapter 2 (commencing with Section 21100) of Division 21 of the Elections Code, and actions under Chapter 22.9 (commencing with Section 22684) of Division 8 of the Business and Professions Code shall be placed on the calendar in the order of their date of filing and shall be given precedence.~~

~~(b) This section shall remain in effect only until January 1, 2027, and as of that date is repealed, unless a later enacted statute, that is enacted before January 1, 2027, deletes or extends that date.~~

~~SEC. 3. Section 35 of the Code of Civil Procedure, as amended by Section 2 of Chapter 343 of the Statutes of 2023, is amended to read:~~

~~35. (a) Proceedings in cases involving the registration or denial of registration of voters, the certification or denial of certification of candidates, the certification or denial of certification of ballot measures, election contests, actions under Chapter 2 (commencing~~

~~with Section 21100) of Division 21 of the Elections Code, and actions under Chapter 22.9 (commencing with Section 22684) of Division 8 of the Business and Professions Code shall be placed on the calendar in the order of their date of filing and shall be given precedence.~~

~~(b) This section shall become operative January 1, 2027.~~

O

From: Kilgore, Preston (BOS)
To: BOS Legislation, (BOS)
Cc: Preston, Dean (BOS); Somera, Alisa (BOS)
Subject: Re: Resolution supporting State Bills to Address Artificial Intelligence in Elections
Date: Tuesday, May 14, 2024 6:13:22 PM

Thank you, Arthur. To my knowledge, the California State Association of Counties, League of California Cities, or the National League of Cities have *not* taken a position on these bills.

Preston Kilgore
Pronouns: He/Him
Chief of Staff | District 5
Supervisor Dean Preston
Sign up for the District 5 Newsletter here!

From: BOS Legislation, (BOS) <bos.legislation@sfgov.org>
Sent: Tuesday, May 14, 2024 2:34 PM
To: BOS Legislation, (BOS) <bos.legislation@sfgov.org>; Kilgore, Preston (BOS) <preston.kilgore@sfgov.org>
Cc: Preston, Dean (BOS) <dean.preston@sfgov.org>; Somera, Alisa (BOS) <alisa.somera@sfgov.org>
Subject: RE: Resolution supporting State Bills to Address Artificial Intelligence in Elections

Hello Preston,

Per Board Rule 2.8.2, please confirm that organizations such as the California State Association of Counties, League of California Cities, or the National League of Cities have *not* taken a position on these bills. If they have, please provide a copy of their statement for completeness of the file

Thanks,

Arthur Khoo
Office of the Clerk of the Board
San Francisco Board of Supervisors
1 Dr. Carlton B. Goodlett Place, Room 244
San Francisco, CA 94102
(415) 554-4447 | (415) 554-5163
arthur.khoo@sfgov.org | www.sfbos.org

Disclosures: *Personal information that is provided in communications to the Board of Supervisors is subject to disclosure under the California Public Records Act and the San Francisco Sunshine Ordinance. Personal information provided will not be redacted. Members of the public are not required to provide personal identifying information when they communicate with the Board of Supervisors and its committees. All written or oral communications that members of the public submit to the Clerk's Office regarding pending legislation or hearings will be made available to all members of the public for inspection and copying. The Clerk's Office does not redact any information from these submissions. This means that personal information—including names, phone numbers, addresses and similar information that a member of the public elects to submit to the Board and its committees—may appear on the Board of Supervisors website or in other public documents that members of the public may inspect or copy.*

From: BOS Legislation, (BOS) <bos.legislation@sfgov.org>
Sent: Tuesday, May 14, 2024 2:32 PM
To: Kilgore, Preston (BOS) <preston.kilgore@sfgov.org>; BOS Legislation, (BOS) <bos.legislation@sfgov.org>
Cc: Preston, Dean (BOS) <dean.preston@sfgov.org>; Somera, Alisa (BOS) <alisa.somera@sfgov.org>
Subject: RE: Resolution supporting State Bills to Address Artificial Intelligence in Elections

Hi Preston,

Since the item is requested to be placed on the For Adoption Without Committee Reference of the agenda, pursuant to Board Rule 2.1.2, please confirm that these matters are routine, not contentious in nature, and of no special interest.

Also, please provide copies of the State Assembly Bills as referenced in the subject Resolution. Plus, we are also seeking your Supervisor's approval for the introduction.

Regards,

Arthur Khoo
Office of the Clerk of the Board
San Francisco Board of Supervisors
1 Dr. Carlton B. Goodlett Place, Room 244
San Francisco, CA 94102
(415) 554-4447 | (415) 554-5163
arthur.khoo@sfgov.org | www.sfbos.org

Disclosures: *Personal information that is provided in communications to the Board of Supervisors is subject to disclosure under the California Public Records Act and the San Francisco Sunshine Ordinance. Personal information provided will not be redacted. Members of the public are not required to provide personal identifying information when they communicate with the Board of Supervisors and its committees. All written or oral communications that members of the public submit to the Clerk's Office regarding pending legislation or hearings will be made available to all members of the public for inspection and copying. The Clerk's Office does not redact any information from these submissions. This means that personal information—including names, phone numbers, addresses and similar information that a member of the public elects to submit to the Board and its committees—may appear on the Board of Supervisors website or in other public documents that members of the public may inspect or copy.*

From: Kilgore, Preston (BOS) <preston.kilgore@sfgov.org>
Sent: Tuesday, May 14, 2024 2:26 PM
To: BOS Legislation, (BOS) <bos.legislation@sfgov.org>
Cc: Preston, Dean (BOS) <dean.preston@sfgov.org>; Somera, Alisa (BOS) <alisa.somera@sfgov.org>
Subject: Resolution supporting State Bills to Address Artificial Intelligence in Elections

Good afternoon,

Please find an introduction form and a resolution supporting State Bills to Address Artificial Intelligence in Elections from Supervisor Preston.

If you have any questions or concerns,

Please do not hesitate to reach out.

Thanks

Preston Kilgore
Pronouns: He/Him
Chief of Staff | District 5
Supervisor Dean Preston
Sign up for the District 5 Newsletter here!

Introduction Form

(by a Member of the Board of Supervisors or the Mayor)

I hereby submit the following item for introduction (select only one):

□ 1. For reference to Committee (Ordinance, Resolution, Motion or Charter Amendment)

☒ 2. Request for next printed agenda (For Adoption Without Committee Reference)
(Routine, non-controversial and/or commendatory matters only)

□ 3. Request for Hearing on a subject matter at Committee

□ 4. Request for Letter beginning with "Supervisor □ inquires…"

□ 5. City Attorney Request

□ 6. Call File No. □ from Committee.

□ 7. Budget and Legislative Analyst Request (attached written Motion)

□ 8. Substitute Legislation File No. □

□ 9. Reactivate File No. □

□ 10. Topic submitted for Mayoral Appearance before the Board on □

The proposed legislation should be forwarded to the following (please check all appropriate boxes):

□ Small Business Commission □ Youth Commission □ Ethics Commission

□ Planning Commission □ Building Inspection Commission □ Human Resources Department

General Plan Referral sent to the Planning Department (proposed legislation subject to Charter 4.105 & Admin 2A.53):

□ Yes □ No

(Note: For Imperative Agenda items (a Resolution not on the printed agenda), use the Imperative Agenda Form.)

Sponsor(s):

Supervisor Preston

Subject:

Resolution supporting State Bills to Address Artificial Intelligence in Elections

Long Title or text listed:

Resolution Supporting AB 2655 (Berman), AB 2839 (Pellerin), AB 3211 (Wicks), and SB 1228 (Padilla) to Address Artificial Intelligence in Elections.

Signature of Sponsoring Supervisor: □