Date of Hearing: April 23, 2024

ASSEMBLY COMMITTEE ON PRIVACY AND CONSUMER PROTECTION
Rebecca Bauer-Kahan, Chair
AB 1856 (Ta) – As Amended March 13, 2024

AS PROPOSED TO BE AMENDED

SUBJECT: Disorderly conduct: distribution of intimate images

SYNOPSIS

As generative artificial intelligence (GenAI) has become more accessible, it has been adopted for a number of nefarious purposes, including scamming, the propagation of political disinformation, and the generation of nonconsensual deepfake pornography. Nonconsensual deepfake pornography is incredibly damaging to individuals and to society. Teenage males in California's middle and high schools can now use cheap, phone-based "nudification" applications to digitally undress their underage female classmates, and the likenesses of female celebrities and public figures are being regularly adapted into high-resolution (but fake) pornographic scenarios.

Existing law makes it a misdemeanor to intentionally distribute intimate imagery of another identifiable person without that person's consent. This crime is also commonly known as "revenge porn." This bill would expand California's revenge porn laws to include deepfake imagery.

This bill is author-sponsored and supported by a number of public safety associations, as well as SAG-AFTRA. It is opposed by ACLU California Action and the California Public Defenders Association. This bill was previously passed unanimously out of Assembly Public Safety Committee.

SUMMARY: Provides that an individual who intentionally distributes nonconsensual deepfake pornography is subject to a misdemeanor. Specifically, **this bill**:

1) Requires that a person who intentionally distributes a deepfake containing various listed content, and who has knowledge or should have knowledge of certain things, is guilty of disorderly conduct if the deepfake causes the person depicted to suffer distress.

   a) Listed content includes any of the following:

      i) An intimate body part or parts of an identifiable person.

      ii) The person depicted engaged in an act of sexual intercourse, sodomy, oral copulation, sexual penetration.

      iii) Masturbation by the person depicted.

   b) Required knowledge includes both of the following:

      i) That the person depicted did not consent to the distribution of the deepfake.

ii) That the distribution will cause serious emotional distress.

2) Defines "distribute" to mean making an image available to another person through any medium, including, but not limited to, exhibiting it in public, giving possession of the image, or through the use of the internet, email, or text messaging.

3) Defines "deepfake" to mean any audio or visual media in an electronic format, including, but not limited to, an image, motion picture film, or video recording, that it created or altered such that it would falsely appear to a reasonable observer to be an authentic record of the actual speech or conduct of the individual depicted in the recording.

EXISTING LAW:

1) Makes it a misdemeanor for a person to intentionally distribute an image of the intimate body parts of another identifiable person, or of the person depicted engaged in a sex act, under circumstances in which the persons agreed or understood that the image would remain private, and the person distributing the image knows or should know that the distribution of the image will cause serious emotional distress, and the person depicted suffers that distress. This crime is also commonly known as "revenge porn." (Pen. Code, § 647.)

   a) States that a person intentionally distributes an image when that person personally distributes the image, or arranges, specifically requests, or intentionally causes another person to distribute that image.

   b) Provides that distribution of the image is not a violation of the law if:

      i) The distribution is made in the course of reporting an unlawful activity;

      ii) The distribution is made in compliance with a subpoena or other court order for use in a legal proceeding; or,

      iii) The distribution is made in the course of a lawful public proceeding.

   c) Defines "intimate body part" to mean "any portion of the genitals, the anus and, in the case of a female, also includes any portion of the breasts below the top of the areola that is either uncovered or clearly visible through clothing."

2) Defines "deepfake" as any audio or visual content that has been generated or manipulated by artificial intelligence which would falsely appear to be authentic or truthful and which features depictions of people appearing to say or do things they did not say or do without their consent. (Gov. Code, § 11547.5.)

FISCAL EFFECT: As currently in print, this bill is keyed fiscal.

COMMENTS:

1) **Artificial Intelligence and Generative Artificial Intelligence.** Artificial intelligence (AI) refers to the mimicking of human intelligence by artificial systems, such as computers. AI uses algorithms – sets of rules – to transform inputs into outputs. Inputs and outputs can be anything a computer can process: numbers, text, audio, video, or movement. AI that are trained on small, specific datasets in order to make recommendations and predictions are sometimes referred to as

"predictive AI." This differentiates them from GenAI, which are trained on massive datasets in order to produce detailed text and images. When Netflix suggests a TV show to a viewer, the recommendation is produced by predictive AI that has been trained on the viewing habits of Netflix users. When DALL-E generates high-resolution, lifelike images, it uses GenAI that has been trained on ~250 million text-image pairs.

2) **Deepfake pornography.** The creation of text, imagery, video, and audio by GenAI has the potential to change the world by automating repetitive tasks and fostering creativity. When employed by bad actors, however, these capabilities have the potential to invade privacy and disrupt the lives of Californians. Since its inception, GenAI has been used to create nonconsensual pornography, more accurately referred to by sexual assault experts as image-based sexual abuse. According to a recent New York Times article, phone-based apps allowing teenage boys to digitally "nudify" their classmates have become increasingly accessible and affordable:

> Boys in several states have used widely available "nudification" apps to pervert real, identifiable photos of their clothed female classmates, shown attending events like school proms, into graphic, convincing-looking images of the girls with exposed A.I.-generated breasts and genitalia. In some cases, boys shared the faked images in the school lunchroom, on the school bus or through group chats on platforms like Snapchat and Instagram, according to school and police reports.[1]

The harms of AI-powered image-based sexual abuse are already being felt in California:

> A third school in Southern California has been hit with allegations of digitally manipulated images of students circulating around campus . . . "Sixteen eighth-grade students were identified as being victimized, as well as five egregiously involved eighth-grade students," Superintendent Michael Bregy wrote. While Bregy acknowledged that children "are still learning and growing, and mistakes are part of the process," he affirmed disciplinary measures had been taken and noted that the incident was swiftly contained. The district vowed to hold accountable any other students "found to be creating, disseminating, or in possession of AI-generated images of this nature."[2]

Women are the primary targets of these efforts, and no one appears to be immune: in the run-up to the 2024 Super Bowl, a series of images involving Taylor Swift began to appear on the social media platform X (formerly Twitter). These images were removed over the following days, but the damage had been done:

> "We are too little, too late at this point, but we can still try to mitigate the disaster that's emerging," says Mary Anne Franks, a professor at George Washington University Law School and president of the Cyber Civil Rights Initiative. Women are "canaries in the coal mine" when it comes to the abuse of artificial intelligence, she adds. "It's not just going to be

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

[2] Mackenzie Tatananni, "'Inappropriate images' circulate at yet another California high school, as officials grapple with how to protect teens from AI porn created by classmates," *Daily Mail,* Apr. 11, 2024, https://www.dailymail.co.uk/news/article-13295475/Inappropriate-images-California-Fairfax-High-School-AI-deepfake.html

> the 14-year-old girl or Taylor Swift. It's going to be politicians. It's going to be world leaders. It's going to be elections."[3]

The harm inflicted on women and girls by this technology cannot be understated. In a recent *Guardian* article by gender equity expert and journalist, Luba Kassova, she argues that "nonconsensual deepfake pornography has become a growing human rights crisis." In her article she asks readers to:

> Imagine finding that someone has taken a picture of you from the internet and superimposed it on a sexually explicit image available online. Or that a video appears showing you having sex with someone you have never met.
>
> Imagine worrying that your children, partner, parents or colleagues might see this and believe it is really you. And that your frantic attempts to take it off social media keep failing, and the fake "you" keeps reappearing and multiplying. Imagine realising that these images could remain online for ever and discovering that no laws exist to prosecute the people who created it.[4]

The problem has become so pervasive that the United States Department of Justice recently launched the first the first national 24/7 helpline for survivors of image-based sexual abuse.[5] According to RAINN, a non-profit anti-sexual assault organization, more than 100,000 deepfake images and videos are posted on the internet every day.[6] The *2023 State of Deepfakes* report found in its survey of American men that 74 percent of deepfake pornography users did not feel guilty about their consumption. According to the report's authors, this finding suggests that deepfake pornographic content is becoming normalized and accepted. Further, of those surveyed almost one-third of those surveyed stated that they did not think that deepfake pornography hurt anyone as long as it was only used for their personal interest.[7]

3) **What this bill would do.** This bill would provide that an individual who intentionally distributes nonconsensual deepfake pornography is subject to a misdemeanor. The language of the bill mirrors the section immediately above it in code, which relates to the nonconsensual distribution of intimate imagery – also known as "revenge porn."

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

> As with any new technology, artificial intelligence can improve people's lives. However, AI can also be used to inflict harm through dangerous and unregulated "deepfakes". The weaponization of deepfakes to create and distribute revenge pornography can have a massive

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

4 Kassova, Luba. "Tech bros need to realise deepfake porn ruins lives – and the law has to catch up," *The Guardian* (Mar. 1, 2024) https://www.theguardian.com/global-development/2024/mar/01/tech-bros-nonconsensual-sexual-deepfakes-videos-porn-law-taylor-swift.

[5] Travers, Karen and Emmanuelle Saliba. "Fake explicit Taylor Swift images: White House is 'alarmed'," *ABC News* (Jan. 26, 2024) https://abcnews.go.com/US/white-house-calls-legislation-regulate-ai-amid-explicit/story?id=106718520.

[6] *Ibid.*

[7] *2023 State of Deepfakes: Realities, Threats, and Impact.* Home Security Heroes. https://www.homesecurityheroes.com/state-of-deepfakes/#deepfake-porn-survey.

> impact on the economy, national security, and individual harm. The Legislature must take action to protect victims from extortion, humiliation, and harassment that can come from artificially generated pornography.

5) **Analysis**. Nonconsensual deepfake pornography is incredibly damaging to individuals and to society. This bill represents a meaningful attempt to combat these issues by expanding California's "revenge porn" laws to include deepfake materials.

Defining "deepfake." This bill does not adopt either of the definitions for "deepfake" being used in California. The authors are justified in this decision – and it is worth briefly commenting on why this is so.

The California Privacy Protection Agency provides the following definition:

> "Deepfake" means manipulated or synthetic audio, image, or video content that depicts a consumer saying or doing things they did not say or do and that are presented as truthful or authentic without the consumer's knowledge and permission.

Section 11547.6 of Government Code provides the following definition:

> "Deepfake" means audio or visual content that has been generated or manipulated by artificial intelligence which would falsely appear to be authentic or truthful and which features depictions of people appearing to say or do things they did not say or do without their consent.

These definitions share two issues. First, whether a given piece of content is labeled a "deepfake" should not depend on the truth of its content. During the 2023 Republican Primary, presidential candidate Ron Desantis released an advertisement containing an AI-generated speech by former president Donald Trump.[8] The deepfake's content was "true," in the sense that it was a word-for-word recreation of a post Trump had made on social media. Trump may not have spoken the words out loud, but they were still his speech. Was the audio that Desantis generated not, then, a deepfake? According to both of these definitions, it was not. The AI-generated audio depicted a person – Trump – saying something that he truthfully did say.

The second issue relates to the technical role of consent in defining whether a given piece of content may be considered a "deepfake." Consent is, of course, an essential component of determining when synthetically-generated content should be subject to punishment. But, strictly speaking, the content's fundamentally synthetic nature – as distinguished from the potentially emotionally devastating impact the content may inflict on the real person whose likeness is used therein – serves as a better basis for a technical definition of "deepfake."

This bill's definition of "deepfake" falls into neither of these traps:

> "Deepfake" means any audio or visual media in an electronic format . . . that is created or altered in a manner that it would falsely appear to a reasonable observer to be an authentic record of the actual speech or conduct of the individual depicted in the recording media.

[8] Alex Isenstadt, "DeSantis PAC uses AI-generated Trump voice in ad attacking ex-president," *Politico*, Jul. 17, 2023, www.politico.com/news/2023/07/17/desantis-pac-ai-generated-trump-in-ad-00106695.

According to this definition, consent plays no role in whether a given piece of content is a "deepfake." In other words, deepfake content will forever be considered deepfake content, and real content cannot become fake. Furthermore, Desantis's AI-generated audio is clearly a "deepfake" under this definition, as the audio is not "an authentic record of actual speech." The phrase "authentic record" is key: the speech may have actually occurred, but the record of it is manufactured.

Finally, it is worth noting that this definition does not contain the words "artificial intelligence." This aligns it with the Privacy Agency's definition, but places it at odds with the definition that appears in Government Code. It is the opinion of this Committee that definitions should be made technology-neutral whenever possible; this bill manages to define "deepfake" without tying it to a definition of "artificial intelligence," and the bill is better for it.

First Amendment considerations. The Assembly Public Safety Committee's analysis of this bill thoroughly discusses it in the context of the First Amendment. According to the Assembly Public Safety Committee:

> [. . .] A former version of California's "revenge porn" law (Pen. Code, § 647, subd. (j)(4)(iii)) survived First Amendment scrutiny in ***People v. Iniguez*** (2016) 247 Cal.App.4th Supp. 1 (*Iniguez*).
>
> There, the defendant argued the statute was overbroad, violating free speech. Under the overbreadth doctrine, a defendant "may challenge a statute not because their own rights of free expression are violated, but because the very existence of an overbroad statute may cause others not before the court to refrain from constitutionally protected expression. [Citations.]" (*In re M.S.* (1995) 10 Cal.4th 698, 709.) To avoid being overbroad, "statutes attempting to restrict or burden the exercise of First Amendment rights must be narrowly drawn and represent a considered legislative judgment that a particular mode of expression has to give way to other compelling needs of society." (*Broadrick v. Oklahoma* (1973) 413 U.S. 601, 611–612 [citations omitted].)
>
> Assuming, without deciding a person has a free speech right to distribute such images, the *Iniguez* court concluded former subdivision (j)(4)(iii) of Penal Code section 647.6 was not constitutionally overbroad because its requirement that a person intend to cause distress served to narrow the law. (*People v. Iniguez, supra*, 247 Cal.App.4th Supp. at pp. 7-8.) The court noted this rendered the law inapplicable should a person act under a mistake of fact or by accident. (*Id.* at pp. 7-8.)
>
> The *Iniguez* court also explained that "it is not just *any* images that are subject to the statute, but only those which were taken under circumstances where the parties agreed or understood the images were to remain private. The government has an important interest in protecting the substantial privacy interests of individuals from being invaded in an intolerable manner." (*People v. Iniguez, supra*, 247 Cal.App.4th Supp. at p. 8 [citation omitted].) The court stated, "It is evident that barring persons from intentionally causing others ***serious emotional distress through the distribution of photos of their intimate body parts*** is a compelling need of society." (Emphasis added.) (*Ibid.*)
>
> Additionally, in *Iniguez* at. pp. 10-11, the defendant also argued insufficient evidence supported his conviction because he had failed to "distribute" the photo by posting it on Facebook. The court concluded, however, "there is no indication in section 647, subdivision

> (j)(4), that the term "distribute[s]" was intended to have a technical legal meaning, or to mean anything other than its commonly used and known definition of "to give or deliver (something) to people." (*Id*. at p. 10; See also, Merriam Webster online definition of "distribute.") The court further noted, "Legislative analyses of the Senate bill that enacted section 647, subdivision (j)(4), are replete with indications that posting images on public Web sites was precisely one of the evils the statute sought to remedy." (*Ibid*.)

This bill is mostly identical to the revenge porn provision that the California Supreme Court considered in *Iniguez*,[9] including the requirement that the "person distributing the deepfake knows or should know that the person depicted did not consent to the distribution and that the distribution of the deepfake will cause serious emotional distress, and the person depicted suffers that distress." The bill differs in two respects. First, it expands the definition of "distribute" – defined under existing law as "exhibiting in public or giving possession" – for purposes of revenge porn and deepfakes to include "making the image available to another person through any medium, including, but not limited to, exhibiting it in public, giving possession of the image, or through the use of internet, email, or text messaging." This appears to be an elaboration on modern variations of "giving possession." Second, the bill – and the process by which deepfakes are created, more generally – does not apply to circumstances in which the parties understood the image to be private. But the life-like depiction of a person's intimate parts still constitutes a major invasion of privacy.

Notwithstanding *Iniguez*, ACLU California Action, in opposition, asserts that the bill is unconstitutional because it doesn't meet the constitutional standard for defamation of public figures:

> … [T]he First Amendment protects nearly all speech, with only a handful of notable exceptions. One of those exceptions is "defamation." But there are numerous constitutional requirements that the Supreme Court has imposed before speech can be prohibited – even speech that is false and may harm someone's reputation and/or may cause emotional distress. (See generally *Hustler Magazine, Inc. v. Falwell*, 485 U.S. 46 (1988) and *New York Times Co. v. Sullivan*, 376 U.S. 254 (1964).) The constitutional requirements that are most relevant here are that even false speech against a public figure, such as a politician, cannot be prohibited unless the plaintiff can show by clear and convincing evidence that the speaker acted with actual malice, i.e., that the speaker knew that the speech was false or acted with "reckless disregard" of its falsity. (*New York Times v. Sullivan*, 376 U.S. at 279-86.)
>
> Assembly Bill 1856 does not take into account these constitutional safeguards. Under the bill, someone who distributed a deepfake depiction of a politician who had staked their reputation on support for family values having sex with a

[9] The predecessor revenge porn statute considered by the court was substantially similar to the current one. It read "(A) Any person who photographs or records by any means the image of the intimate body part or parts of another identifiable person, under circumstances where the parties agree or understand that the image shall remain private, and the person subsequently distributes the image taken, with the intent to cause serious emotional distress, and the depicted person suffers serious emotional distress. [¶] (B) As used in this paragraph, intimate body part means any portion of the genitals, and in the case of a female, also includes any portion of the breasts below the top of the areola, that is either uncovered or visible through less than fully opaque clothing." (§ 647, former (j)(4).) (*Id.* at p. 6.)

> sex worker even if the person who distributed the deepfake believed it was authentic, would be subject to criminal penalties**.** Indeed, this constitutional problem is exacerbated by the bill's depiction of "deepfake" to be "any image, motion picture film, or video recording, *that is created or altered in a manner that it would falsely appear to a reasonable observer to be an authentic record of the actual speech or conduct of the individual depicted in the recording*." In other words, someone who distributed the deep fake but did not create it may very well not know that it is fake.
>
> Nor does the bill's requirement that the speaker know that distribution will "cause serious emotional distress, and that the person depicted suffers that distress" obviate the constitutional requirement for "actual malice." In *Hustler*, 485 U.S. 46 (1988), the United States Supreme Court held that the First Amendment protected speech directed at public figures even if it caused severe emotional distress unless "the publication contains a false statement of fact which was made with 'actual malice,' i.e., with knowledge that the statement was false or with reckless disregard as to whether or not it was true." (*Id*. at 56.)

6) **Committee amendments.** Three minor amendments have been introduced in collaboration with the authors. The first fixes an unintentional language omission:

> (ii) A person who intentionally distributes or causes to be distributed a deepfake of an intimate body part or parts of another identifiable person, or a deepfake of the person depicted engaged in an act of sexual intercourse, sodomy, oral copulation, sexual penetration, or a deepfake of masturbation by the person depicted or in which the person depicted participates, ***under circumstances in which*** the person distributing the deepfake knows or should know that the person depicted did not consent to the distribution and that the distribution of the deepfake will cause serious emotional distress, and the person depicted suffers that distress.

The second modifies the definition of "distribute" to include all content, not just images:

> (i) "Distribute" means making ~~the image~~ ***content*** available to another person through any medium, including, but not limited to, exhibiting it in public, giving possession of the ~~image~~ ***content***, or through the use of internet, email, or text messaging.

The third modifies the definition of "deepfake" to include physical media.

> (ii) "Deepfake" means any audio or visual media ~~in an electronic format,~~ including, ~~without limitation, any~~ ***but not limited to, an*** image, motion picture film, or video recording, that is created or altered ~~in a manner~~ ***such*** that it would falsely appear to a reasonable observer to be an authentic record of the actual speech or conduct of the individual depicted in the ~~recording~~ ***media.***

7) **Related legislation.** AB 1721 (Ta, 2023) would have made it a crime for a person to knowingly distribute nonconsensual deepfake pornography. This bill was held in Assembly Public Safety Committee.

AB 1831 (Berman and Sanchez, 2024) would expand existing prohibitions on the creation and distribution of child pornography to include artificial intelligence-generated child pornography. This bill is currently pending in this Committee.

AB 3050 (Low, 2024) would make an AI-generating entity that creates a nonconsensual deepfake using a person's name, voice, signature, photograph, or likeness, in any manner, liable for damages sustained as a result. This bill is currently pending in this Committee.

SB 1235 (Gonzalez, 2024) would convene a working group to study the impact of deepfakes on state and local government, businesses and the workforce, education, and residents of the state. This bill is currently pending in Senate Education Committee.

ARGUMENTS IN SUPPORT:

California State Sheriffs' Association writes:

> The advancement of AI technology has exacerbated the prevalence and severity of revenge porn – with the line of what is real and what is generated blurring together, putting unsuspecting victims at risk of their image being exploited. Current revenge porn statutes and remedies have flaws and are insufficient. While these are not easy problems to solve, the Legislature can and should criminalize AI-generated revenge porn. AB 1856 would provide a stronger disincentive to create pornographic deepfakes.

California District Attorneys Association writes:

> Distribution of deepfake images is increasing across the state in high schools and middle schools and often involves the online posting of deepfake images to shame, embarrass, harass, and intimidate victims. These type of incidents can cause lasting emotional trauma and distress.

SAG-AFTRA writes:

> Our members have been targets of this despicable conduct since the inception of deepfake technology. This abuse of this technology has negatively impacted careers, relationships, and the emotional wellbeing of its victims. We have sponsored legislation creating civil penalties, but it has not been enough. Our hope is that real criminal penalties and accountability will help.

ARGUMENTS IN OPPOSITION:

California Public Defenders Association writes:

> AB 1856 would likely run afoul of the First Amendment. As noted in the Assembly Public Safety Committee analysis of AB 1280 (Grayson) 2019 which also sought to criminalize deepfake recordings of adult sexual activity, while courts have found that laws criminalizing deepfakes involving child pornography serve a compelling governmental interest, prohibitions of depictions of adult sexual activity are not afforded the same protection.

ACLU California Action writes:

> **[**We] fear that AB 1856 will result in further criminalization of youth, particularly youth of color, who engage in the proscribed behavior. A recent study found that 73% of teenagers 17 years old or younger had been exposed to online pornography. Young people access pornography in a variety of ways, including by sharing it with one another and through social media. Under AB 1856, a young person could be convicted of a crime and sentenced to jail if they share deepfake pornography with a peer under the circumstances outlined in the bill. While this conduct may very well be inappropriate and sometimes harmful, sending young people to jail appears overly punitive and counterproductive.

REGISTERED SUPPORT / OPPOSITION:

Support

California District Attorneys Association
California State Sheriffs' Association
Peace Officers Research Association of California (PORAC)
SAG-AFTRA

Opposition

ACLU California Action
California Public Defenders Association

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