LOS ANGELES & SAN FRANCISCO
Daily Journal
www.dailyjournal.com
MONDAY, JANUARY 27, 2020

PERSPECTIVE

Dealing with deepfakes

California and social media platforms take aim at deepfakes while federal lawmakers take a 'studied' approach

By Melanie Howard and Adam Shapiro

As the presidential election draws nearer, several deepfakes have surfaced manipulating videos and statements of politicians currently in office as well as candidates campaigning for election in 2020. Deepfakes are video and audio recordings that have been digitally manipulated so that the subjects appear to say or do something that they did not say or do. This is on the radar for state and federal legislators, in large part because of the significant gaps in the protections that existing laws afford individuals victimized by the creation and distribution of deepfakes. The technology to create deepfakes has become increasingly available and can be used by anyone. The popular FakeApp application is free to download, and uses deep-learning neural networks and face-mapping software to create very realistic final products. In addition to the continuing concerns surrounding the use of deepfakes to superimpose faces on actors in pornographic films, lawmakers have become focused on the potential malicious or nefarious uses of deepfakes in connection with political campaigns and elections.

In an effort to close these gaps and combat the proliferation of deepfakes, California recently enacted two new laws:

• Assembly Bill 602 bans the sexually explicit depiction of individuals who appear, as a result of digital or electronic technology, to be performing sexual acts that they did not actually perform.

• Assembly Bill 730 bans the distribution of audio or video that gives a false, damaging impression of a politician's words or actions. The law applies to candidates within 60 days of an election, with certain exceptions, and sunsets in 2023.

Both laws provide a private right of action for targets of these manipulated videos to seek injunctive relief to prevent further dissemination of the videos, as well as damages from the offender. AB 602 provides for statutory damages up to $150,000. AB 760 allows victims to seek general or special damages, as well as attorneys' fees. Both laws also exempt certain speech that would be protected by the U.S. Constitution, highlighting the inherent tension with First Amendment concerns.

In its analysis of AB 602, the California Senate Judiciary Committee recognized that existing intellectual property, defamation and criminal laws provide an imperfect solution to addressing the potential harms to those depicted in deepfake videos. The committee concluded that '[w] hile a person could certainly allege a cause of action for defamation based on the creation of a deepfake, such a cause of action would not address the issue of unauthorized use of a person's likeness." Seeking relief under California's right of publicity law would present challenges because the claimant "would have to demonstrate that the altered image produced via deepfake was, in fact, a 'likeness' of themselves, which could be complicated by the fact that the image was created by combining that image with the image of a third person." Neither does copyright law provide a clear path to relief, as an actor appearing in a film "generally has no standing … to sue if someone else alters the film to make it look like the actor is engaged in a new or different act, including a sexual act."

As enacted, California's new laws provide some protections for targets of deepfakes, but they are narrowly tailored. For example, AB 730 only applies to "materially deceptive" deepfakes that are distributed: (1) within 60 days of an election at which the targeted candidate will appear on the ballot; (2) with "actual malice"; and (3) with the "intent to injure the candidate's reputation" or "to deceive a voter into voting for or against the candidate."

As originally drafted by Assemblyman Berman, AB 602 criminalized the distribution of "deceptive videos" by individuals who knew or reasonably should have known that the video would deceive any person who views the recording, or that it would "defame, slander, or embarrass the subject of the recording." Because there was no requirement of publication of the "deceptive video," the Assembly Judiciary Committee lodged objections on First Amendment grounds, and the bill was revised to follow on AB 2643, which codified California's revenge porn statute (Cal. Civ. Code Section 1708.85). Existing California law made it a criminal act to distribute, with the intent to cause serious emotional distress, sexually explicit images of another person, and AB 2643 created a private right of action for victims.

While California has been a pioneer in the efforts to prevent damaging uses of deepfakes, other states have taken the lead in criminalizing the creation and distribution of these videos. Virginia was one of the first states to criminalize distribution of revenge porn deepfakes. As of July 1 of last year, Virginia made the distribution of nonconsensual "falsely created" explicit images and videos a Class 1 misdemeanor, with a penalty of up to a year in jail and a fine of $2,500. Texas took the first stance against use of deepfakes in the context of elections. Texas' S.B. 751 (effective Sept. 1 of last year) makes it a crime to create any video "with the intent to deceive, that appears to depict a real person performing an action that did not occur in reality," and, separately to publish or distribute it within 30 days of an election "with intent to injure a candidate or influence the result of an election." Violators face a Class A misdemeanor, with a sentence of up to a year in jail and a fine of $4,000.

Federal Legislation

While Congress has considered legislation aimed at researching and investigating the misuse of deepfake technology, the prospect of any federal legislation with an effective enforcement mechanism and meaningful remedies for targets of these videos still seems remote.

President Donald Trump signed the National Defense Authorization Act for Fiscal Year 2020, a law that explicitly addresses deepfakes, in December. The NDAA requires the director of national intelligence to: (1) submit an annual report to Congress regarding

any potential national security risks of deepfakes (which the bill identifies as "machine manipulated media" but does not define beyond that) and any actual or potential use of deepfakes by foreign governments; (2) notify Congress of any attempted use of deepfakes by foreign entities to meddle with U.S. elections or political processes; and (3) commence a competition for new technology that can automatically detect deepfakes, with awards up to $5 million.

Although the NDAA signals that federal lawmakers are focusing on the potential harms of deepfakes in elections, and the lack of an existing legal framework within which to address them, the law does not require anything more than reporting on deepfakes, and does not actually criminalize the creation of deepfakes.

The proposed Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act (H.R. 3230), introduced in the House of Representatives in June 2019, is more comprehensive in its attempt to combat deepfakes. H.R. 3230 proposes to prohibit the creation with the intent to distribute deepfakes of any kind, unless the video or audio includes "an embedded digital watermark clearly identifying such record as containing altered audio or visual elements." With respect to more egregious forms of deepfakes (including pornography, videos featuring electoral candidates, videos intended to incite physical harm or violence, and videos intended to perpetuate criminal conduct related to fraud), the bill provides for statutory fines up to $150,000 and possible prison time up to five years. The bill was referred to the Subcommittee on Crime, Terrorism and Homeland Security in June, but has not progressed out of that committee.

The House also adopted the Identifying Outputs of Generative Adversarial Networks Act (H.R. 4355), which requires the National Science Foundation to support research on "manipulated or synthesized content and information authenticity." The House transmitted H.R. 4355 to the Senate at the end of 2019, where the bill was referred to the Committee on Commerce, Science and Transportation.

The Senate has also made efforts to pass impactful deepfake legislation. The Malicious Deep Fake Prohibition Act of 2018 (S. 3805), was intended to prohibit the creation of deepfakes and also provided for criminal penalties. The bill died in Congress at the end of the 2018 legislative session. More recently, the Senate passed S. 2065, the Deepfake Report Act, which would require the Department of Homeland Security to report annually on the use of deepfake technology and assess how foreign governments and domestic groups are using deepfakes to harm national security. Similar to the NDAA, however, this bill authorizes fact gathering but does not provide any meaningful protections to targets of deepfakes.

Social Media Platform Policies

Federal lawmakers have also publicly called upon social media platforms to increase enforcement efforts against deepfakes posted online.

Facebook announced on Jan. 6 that media will be subject to removal from its platform if: (1) the media has been edited "in ways that aren't apparent to an average person and would likely mislead someone into thinking that the subject of the video said words that they did not actually say" and (2) it is the product of artificial intelligence or similar technology that "replaces or superimposes content onto a video, making it appear to be authentic."

Some have criticized Facebook's policy as too narrow and advocate for broader prohibitions addressing not just the sophisticated deepfake technology, but low-tech yet still misleading "cheapfakes" or "shallowfakes" such as the slowed-down video of House Speaker Nancy Pelosi that went viral last year.

Twitter has sought public input on its approach to synthetic and manipulated media. According to the Twitter Blog, Twitter's draft definition of synthetic and manipulated media ("any photo, audio, or video that has been significantly altered or fabricated in a way that intends to mislead people or changes its original meaning") is intended to pull in both deepfakes and shallowfakes. Violation of the policy would not result in automatic removal of the post. Rather Twitter proposes to place a notice next to tweets that share synthetic or manipulated media, warn users before those tweets are shared and add a link to an informational article where users can learn more about manipulated media.

Facebook, together with Amazon, Microsoft and others, appears to be tackling the issues from a technological standpoint as well. These companies have sponsored the Deep Fake Detection Challenge. Funded by $10 million in grants, the challenge aims to bolster technological resources to combat the prevalence and perniciousness of deepfakes, inviting participants "to build innovative new technologies that can help detect deepfakes and manipulated media."

The Fight Continues…

The technology to create deepfakes — and shallowfakes — is available and accessible, and these altered media are being distributed with increasing frequency on social media and other platforms, especially as this year's election cycle ramps up. While the potential use of fakes by foreign actors and others to impact critical national elections is on their radar, federal lawmakers seem to be taking a "studied" approach to the issues. State legislatures, leading tech companies and social media platforms seem more willing to take affirmative steps toward preventing the creation and dissemination of deepfakes and addressing the harm that they cause for their victims. In 2020, more states may follow the lead of California, Texas and Virginia in enacting legislation to combat the harms caused by deepfakes, but it will likely take longer before any comprehensive federal legislation emerges aimed at fighting the malicious misuse of this technology. ■

Melanie Howard *is the chair of the Intellectual Property Protection team, and a member of the Privacy, Security & Data Innovations team at Loeb & Loeb LLP.*

Adam Shapiro *is an associate at the Los Angeles office of Loeb & Loeb LLP.*