A short take on deepfakes

AUTHORS

 Lara Guest

 Molly Reynolds

Our laws and legal system have always struggled to keep pace with technology. From cryptocurrency to self-driving cars, legal and regulatory gaps emerge each time technology progresses.

The latest challenge is “deepfakes”—videos created by artificial intelligence and deep-learning technology that either superimpose an individual’s face onto another person’s body or manipulate existing video and/or audio footage of a person. The result is a fake—but realistic-looking—video that depicts an individual doing or saying things that they have never done.

Since their popularization in 2017, deepfakes have been employed for a variety of purposes. In the political arena, for example, they have been used satirically to provoke dialogue on key issues—like the video produced by a Belgian political party in 2018 of Donald Trump crudely encouraging Belgium to withdraw from the Paris Climate Agreement. Used this way, deepfakes serve a useful social purpose and are protected by Canadian copyright laws, which recognize satire and parody as important and critical tools.[1] But deepfakes have also been used in nefarious ways that raise questions about privacy and control in the era of artificial intelligence.

Such concerns were clearly illustrated in November 2017, when a series of deepfakes that superimposed celebrities faces into pornographic videos were released. Importantly, these videos were created and disseminated without the subject’s consent. Many of the videos have not been removed and others like them have been created, despite some websites’ promises to remove them.

The development of deepfakes has become easier in the past year, with the development of new “user friendly” online deepfake platforms. In this era of technological change, deepfakes present significant risks to both individuals and businesses, and it is presently unclear whether Canadian laws are equipped to provide recourse to deepfake victims.

Personal privacy

The availability of civil causes of action for individual victims of deepfakes has not yet been tested in Canada. Existing law surrounding the nonconsensual disclosure of intimate images (or NCDII—sometimes called “revenge porn”) provides a useful starting point. However, despite the similarity of subject matter, the fact that deepfakes often use public photos in fictitious situations presents a challenge for plaintiffs.

Intentional infliction of emotional distress

A claim for intentional infliction of emotional distress is the most promising privacy tort for individual victims.[2]

A claim for the intentional infliction of emotional distress will be established where there is conduct that a) is flagrant and outrageous, b) was calculated to produce harm, and c) results in a visible and provable injury.[3] This test does not require that the defendant be motivated by malice; a claim may succeed where it is clearly foreseeable that posting

such a video would result in harm.

The first two elements of this test are established where pornographic deepfake videos are distributed without consent. The sexually explicit nature of the videos makes them "outrageous" and a reasonable person would anticipate that releasing a video like this could cause reputational and psychological harm.

The third element—a visible and provable and injury—would likely require a plaintiff to show psychological or monetary injuries, like those flowing from the loss of an opportunity or employment as a result of the deepfake. While these injuries would be case specific, courts are increasingly recognizing the impact that the dissemination of sexually explicit content can have on an individual.[4] In this respect, it does not matter that the video is a deepfake—the intentional infliction of emotional distress would appropriately capture those cases where a plaintiff suffered a real injury, even if the video itself was fake.

Defamation

Another promising cause of action is defamation. Defamation consists of any written or printed words or of any audible or visible matters or acts which has the effect of injuring the reputation of the person to whom it refers. In order to succeed in a defamation action, the plaintiff must prove on the balance of probabilities that the impugned content: a) was defamatory; b) was made in reference to the plaintiff; and c) was published or disseminated.[5]

The tort of defamation is often less applicable to cases of NCDII, due to the fact that these images are "genuine," and therefore the defense of truth may apply.[6] However, in the context of deepfakes, defamation may be easier to establish. This is because the content is manufactured and falsified.

Breach of confidence and public disclosure of private facts

Other causes of action that have been used to secure damages in NCDII cases are less likely to succeed in the context of deepfakes.

The test to establish breach of confidence requires a) that the information have the necessary quality of confidence about it; b) that the information have been imparted in circumstances imparting an obligation of confidence; and c) that there was an unauthorized use of that information to the detriment of the party communicating it.[7] Similarly, to establish the tort of public disclosure of private facts, the plaintiff must show that the defendant gave publicity to a matter concerning the private life of the plaintiff.[8]

It may be difficult for a plaintiff to claim that the information used to create a deepfake had the necessary quality of confidence or was part of his or her private life, when the video depicts a fictitious situation. Further, deepfakes often draw on images and videos that have already been put into a public space, such as Facebook or YouTube, and are intended for public viewing. The issue in the context of a deepfake, unlike NCDII, is not the image itself, but rather how that image is being used and manipulated.

Therefore, there may be little recourse under these privacy torts for deepfake victims, even though the harmful effects may be just as severe as the non-consensual disclosure of genuine intimate images.

False light

Arguably, the cause of action best equipped to deal with deepfakes is one that does not yet exist in Canada: the tort of false light. This cause of action is used in many jurisdictions in the United States and provides damages for publications that put the plaintiff in a false light in the public's eye.[9]

Canadian courts have recognized privacy protections akin to "false light" in the past. In *Aubry v. Éditions Vice-Versa Inc.*, the Supreme Court of Canada accepted that the right to one's image is an element of the right to privacy under section 5 of the Québec *Charter of Human Rights and Freedoms*.[10] However, *Aubry* appears to have to restricted the potential for such claims to celebrities or other individuals who profit from the reproduction of their likeness (akin to a property right). If the tort of false light develops in Canada in line with the *Aubry* jurisprudence, it is possible that a victim of a deepfake may need to prove economic harm due to the loss of chance to profit from the reproduction of their image.

Impact on businesses

The potential impacts of deepfakes on businesses go beyond the obligations to identify and remove such videos. In theory, the risks posed by deepfakes have the ability to impact companies and corporate transactions of all sizes, even those outside the technology industry. This is because deepfakes can present significant reputational risks to corporations.

In any corporate transaction, the reputation of a company and of those who manage it is important. In the mergers and acquisitions context, for example, an acquirer wants to ensure that the target is properly managed and has a positive reputation before investing. While the results of legal and financial diligence are certainly important, the acquirer also wants assurance that the target's management is experienced and reputable, and that no controversies are likely to emerge that will impact the bottom line. This is true of both public and private companies, and is increasingly important in today's climate, where technology allows bad news to travel fast and consumers are increasingly concerned about ethics and social responsibility.

As such, a deepfake that seemingly shows a company's CEO making offensive remarks or employees engaging in inappropriate or unsafe behavior in the workplace could prevent such a merger or acquisition from continuing. Similarly, nefarious deepfakes could cause a public company's share price to drop or prevent the issuance of additional securities.

Privacy torts

In comparison to the context of individual actions, recovering damages for harms caused by a deepfake under existing Canadian privacy torts may be more difficult for a corporation. The causes of action of intentional infliction of emotional harm, breach of confidence and publication of private facts described above relate to the individual who has inadvertently become the subject of such a video, not a corporation who is subsequently affected.

Copyright

Corporations may attempt to seek recourse through a claim of breach of copyright. As set out in the *Copyright Act*, "[i]t is an infringement of copyright for any person to do, without the consent of the owner of the copyright, anything that by this Act only the owner of the copyright has the right to do."[11]

Due to the fact that a deepfake is a novel work (often using existing public images), it may be difficult to for a corporation to establish a copyright interest in the deepfake. That said, if the deepfake uses a corporate logo or other copyrighted image, it is possible that a breach of copyright has occurred.

Intentional interference with economic relations

Corporations may also seek recourse for economic harm caused by deepfakes through the tort of intentional interference with economic relations. This tort has three elements: a) the defendant must have intended to injure the plaintiff's economic interests, b) the interference must have been by illegal or unlawful means, c) the plaintiff must have suffered economic harm or loss as a result. While it may be possible to demonstrate an intention to harm the corporation's interests and the economic loss suffered as a result, a corporate victim of a deepfake may have difficulty establishing that the interference was "illegal" or "unlawful."

Who to sue?

Another challenge for victims of deepfakes is: who to sue? The anonymity of the Internet is such that it may not always be clear who created or distributed the deepfake. Therefore, instead of naming individual defendants, claimants in Canada may choose to pursue the websites and internet service providers who host and enable the dissemination of such content. The issue that will arise in such cases is whether these web-based companies have a legal obligation to identify, remove or prevent the spread of harmful deepfakes.

In the United States, the answer is clear (for now): internet service providers and web-sites cannot be held liable for failure to remove a video that others would view as offensive due to statutory immunity under the *Communication Decency Act* (CDA)[12]

The answer is less clear in Canada. There is no existing case law testing whether an internet service provider or content publisher is liable for NCDII. However, there are no statutory protections for internet service providers and web-sites in Canada akin to the CDA. In fact, the Supreme Court has even gone so far so to uphold injunctions that require Google to remove certain content from its search results.[13] Further, it is possible that failure of internet service providers and web-sites to prevent and remedy the publication of deepfakes could ground a claim in negligence.

Conclusion

While our current laws provide some recourse for victims of deepfakes, it is clear that additional causes of action and remedies may be necessary to appropriately capture the privacy and business concerns posed by such videos. This is especially true as new apps and software make it easier for all individuals—not just sophisticated hackers—to create and disseminate deepfakes.

---

[1] Section 29 of the *Copyright Act* recognizes parody and satire as fair dealing that does not infringe copyright. *Copyright Act*, R.S.C., 1985, c. C-42 .

[2] See: *Jane Doe 464533 v D(N)*, 2016 ONSC 541 (*D(N)*).

[3] *Prinzo v Baycrest Centre for Geriatric Care*, 60 OR (3d) 474 (2002) at para 43; *D(N), supra* note 4 at para 26.

[4] In *D(N)*, Justice Stinson stated that the NDCII is "more than an invasion of a right to informational privacy [and]… it is in many ways analogous to a sexual assault" (para 58).

[5] *Grant v Torstar Corp*. 2009 SCC 61 at para. 28.

[6] *Grant v Torstar Corp*. 2009 SCC 61.

[7] *D(N)*, *supra* note 4 para 21.

[8] *Jane Doe 72511 v Morgan*, 2018 ONSC 6607 (*Morgan*) at para 97; *D(N)*, *supra* note 4 at para 41.

[9] See for example *Peoples Bank & Trust Co. v. Globe Int'l, Inc*., 786 F. Supp. 791, 792 (D. Ark. 1992). See also William L. Prosser, "Privacy" (1960), 48 Cal. L. R. 383 at 389.

[10] *Aubry v. Éditions Vice-Versa Inc.* [1998] 1 SCR 591.

[11] *Copyright Act*, R.S.C., 1985, c. C-42 s. 27(1).

[12] *Communications Decency Act of 1996*, 57 U.S.C, s. 230, For example, in *Herrick v. Grinder*, hundreds of strangers began appearing at the plaintiff's residence, due to fake profiles filed by an ex-partner on the website Grindr. The victim sued Grindr for the attack, styling the case as a products liability claim to get around the CDA immunity. Last month, the Second Circuit court affirmed a lower court decision that Grindr was covered by CDA immunity because a website "'will not be held responsible unless it assisted in the development of what made the content unlawful."

[13] *Google Inc. v Equustek Solutions Inc.,* 2017 SCC 34.