Deep Dive into Deepfakes—Safeguarding Our Digital Identity

Yi Yan
INTRODUCTION

What would you do? You wake up one morning to the sound of your phone’s notification on your social media app. You check the notification and notice that you were tagged in a video. The video depicts you in compromising positions performing sexual acts with another individual. You think to yourself, this is impossible. The events in the video never happened. Who would do this sort of thing? Why were you targeted? You notice that the video already has tens of comments and has been reposted several hundred times. What can you do?

In 2017, a Reddit user posted doctored pornographic clips with face-swapped videos of celebrities and coined the phrase “deepfake.”1 Deepfake technology brings new prospects to video and image manipulation by using artificial intelligence algorithms to fabricate hyper-realistic photos, audio, and videos.2 In early 2019, a report showed that there were 7,964 deepfake videos online.3 In merely nine months, that number grew to 14,678.4 While the exact number of existing deepfake videos is unclear, it is almost certainly in the millions.5 The issues of deepfake technology extend across multiple fields and nations across the world.6 Deepfake technology can be used to digitally resurrect

3. Rob Toews, Deepfakes are going to wreak havoc on society. We are not prepared., FORBES (May 25, 2020), https://www.forbes.com/sites/robertoews/2020/05/25/deepfakes-are-going-to-wreak-havoc-on-society-we-are-not-prepared/?sh=15016bce7494.
4. Id.
the beloved deceased to memorialize their presence.\textsuperscript{7} Alternatively, it can be used maliciously in cases of non-consensual face-swapping pornography.\textsuperscript{8} In an age where people are susceptible to misinformation and distrust, deepfake technology can potentially amplify these problems.\textsuperscript{9} Without proper safeguards, legal reforms, and countermeasures, deepfake technology can cause significant damage to an individual's digital image as well as a nation's societal health.

In the United States (US), the federal government addressed these concerns by including provisions in the National Defense Authorization Act (NDAA), which prescribed the Department of Homeland Security (DHS) the task of compiling reports on deepfakes for five years.\textsuperscript{10} The report's scope entails the potential harm from deepfake technology, such as foreign influence in campaign elections, fraud, and targeted efforts against specific groups of people.\textsuperscript{11} Additionally, the act directs the DHS to investigate deepfake creation technology, methods of detection, and mitigation solutions.\textsuperscript{12} In 2020, former President Donald Trump signed the Identifying Outputs of Generative Adversarial Networks Act, which compels the National Science Foundation (NSF) to work in tandem with companies in the private sector, such as Facebook, Microsoft, and Google, to develop technology that can identify deepfakes.\textsuperscript{13} At the state level, only Texas, Virginia, California, and New York have enacted laws—in varying degrees and limited circumstances—to combat deepfake creation.\textsuperscript{14} Similar to the US, the Cyberspace Administration of China (CAC) and the Central Cyberspace Affairs Commission have opened channels to cooperate with social media platforms


\textsuperscript{8} Id.


\textsuperscript{11} Id.

\textsuperscript{12} Id.

\textsuperscript{13} Id.

\textsuperscript{14} Id.
to investigate countermeasures to deepfake technologies and voice manipulation software.\textsuperscript{15}

This Note seeks to examine and compare existing legislations of sovereign nations to establish the best practice, remedies, and potential solutions to the issues surrounding deepfake technology. This Note argues that, despite the adopted legislation of individual countries, a divided effort between sovereign nations to combat deepfake technology would be insufficient in regulating deepfake creation and usage due to the nature of how the falsified content is disseminated through the internet.

To understand the contemporary matters around deepfake technology and critique existing, as well as proposed legislation of sovereign nations around the world, this Note will proceed as follows. Part I will examine the history and creation of deepfake technology. Part II will dive into the current use and how it compounds the spread of online misinformation. Part III will explore the affected areas of the law. Part IV will analyze and compare the different approaches of sovereign nations in tackling the common issue. Finally, Part V will suggest potential solutions and legislation regarding deepfake limitations and usage by applying the analysis in Part III and Part IV.

I. THE CONCEPTION OF DEEPFAKE TECHNOLOGY

According to the International Data Corporation, worldwide spending on artificial intelligence (AI) is expected to reach $110 billion in 2024 from $50.1 billion in 2020.\textsuperscript{16} In an age of unprecedented volumes of data, AIs can be programmed to learn from an enormous amount of data sets to perform tasks and calculations that are beyond human capabilities.\textsuperscript{17} The applicable

\textsuperscript{15} Overview, Freedom House (2021), (Cyberspace Administration of China sought the assistance of social media platforms such as ByteDance, Tencent, Alibaba, Xiaomi, and Kuaishou to identify elusive activist using deepfake and voice-changing software to conceal their identity).


potentials of AI span across different industries, such as banking, securities, investment, retail, education, and media, to name a few.\footnote{18} The broad field of AI can be broken down into six subfields: (1) Machine Learning; (2) Deep Learning; (3) Cognitive Computing; (4) Computer Vision; (5) Neural Network; and (6) Natural Language Processing.\footnote{19}

Deepfakes are classified under the sub-category of pattern recognition within the umbrella of machine learning.\footnote{20} Deepfakes are created using a method called Generative Adversarial Network.\footnote{21} The technique was introduced in a 2014 paper\footnote{22} by researchers at the University of Montreal.\footnote{23} Their research proposed a new framework comprised of two neural networks.\footnote{24} The first is a generative model which captures the data distribution, and the second is a discriminative model that learns to distinguish whether or not a data set is from the model distribution or the data distribution.\footnote{25} These neural networks are pitted against each other through an adversarial process to separate fake data from real data.\footnote{26} After many iterations of generation and detection, the second model will no longer be able to tell apart the fake and real data, and the content would likely be able to deceive human viewers.\footnote{27} To put it simply, the authors of the paper analogized the adversarial process to a team of counterfeiters.

\footnote{18} Id.
\footnote{19} What is Artificial Intelligence: Definition & Sub-Fields of AI, SOFTWARE TESTING HELP (Dec. 5, 2022), https://www.softwaretestinghelp.com/what-is-artificial-intelligence/#:~:text=It%20is%20the%20process%20of%20learning%20by%20processing%20and%20analyzing%20input%20data%20to%20output.
\footnote{21} Id.
\footnote{22} Ian J. Goodfellow et al., Generative Adversarial Nets, UNIV. OF MONTREAL 1, 1 (Jun. 10, 2014).
\footnote{24} Goodfellow, supra note 22.
\footnote{25} Id.
\footnote{26} Id. at 1.
\footnote{27} TECHSLANG, supra note 20.
producing fake currency, who are representative of the generative model, and the police trying to detect fake currency, who are representative of the discriminative model. As the police try to separate the counterfeit currencies from the real currencies, the counterfeiters become more and more proficient at creating counterfeits that are imperceptible to the police. Competition in this game drives both teams to improve their methods until the counterfeits are indistinguishable from the genuine articles.

While truly understanding the complexities of how AI methods and processes may take more than a layperson’s knowledge of the subject, the tools and instructions on creating deepfakes are readily available over the internet without specialized knowledge. The director of Emerging Technologies for the US General Services Administration’s (GSA) Office of Technology Policy answered questions regarding their concerns with deepfake during a webinar. The director, Alex Cohen, stated, “Deepfakes are really easy to make. There are studios and downloadable tools available that are easy to use and allow people to get started immediately.” In 2019, a Chinese deepfake app, Zao, sparked controversy over privacy concerns after rising in popularity. The app’s signature function was to superimpose the users’ face and swap it with popular film or TV characters so they could play as the characters portrayed in the media. Despite concerns over their privacy policy, which states “free, irrevocable, permanent, transferable, and re-licensable rights to all user-generated content,” the app remained the top free download in China. This shows that the vast majority of social media

28. Goodfellow, supra note 22.
29. Id. at 1.
30. Id.
32. Id.
33. Id.
35. Id.
36. Id.
users either do not understand or are unconcerned about the ramifications of how social media platforms use their data in media manipulation technology. Under the guise of a new social media function to attract users, deepfake technology is widely available to the public. Notably, an iPhone app called Avatarify allows the user to become a puppeteer of another individual’s face. Other apps include Wombo, which converts a photo for humorous lip-sync music videos, and MyHeritage, an app that reanimates old still photos to construct a short and convincing video of the subjects smiling. As the use of synthetic media becomes increasingly more abundant and popular, some companies are incentivized to create more sophisticated software, and other companies will struggle to disseminate between falsified and real media. Finally, due to the open-source mindset of the machine-learning community, whenever there is a breakthrough in generative models, the research and technology are available and free for any individual to potentially weaponize for nefarious reasons.

II. CURRENT USAGE AND CONSEQUENCES OF SYNTHETIC MEDIA

The dangers of deepfake technology were not prevalent until a Reddit user started to release doctored videos of the faces of celebrities and overlayed their faces onto the bodies of adult film actresses. Since then, non-consensual deepfake pornography has spread to other platforms and dedicated deepfake websites.

---

37. Geoffrey A. Fowler, *Anyone with an iPhone can now make deepfakes. We aren’t ready for what happens next.*, WASH. POST (Mar. 25, 2021, 8:00AM), https://www.washingtonpost.com/technology/2021/03/25/deepfake-video-apps/ (Avatarify has been downloaded more than 6 million times since February alone).
38. Id.
39. Id. (Wombo generated 100 million clips in its first two weeks).
40. Id. (MyHeritage reanimated more than 65 million photos of people in March of 2021).
41. Id.
42. France-Presse, * supra note 34* (In June of 2019, Facebook’s Chief Executive, Mark Zuckerberg, said the social network was struggling to find ways to deal with deepfake videos, saying they may constitute “a completely different category” of misinformation than anything faced before).
43. Toews, * supra note 3*.
44. Ian Sample, *What are deepfakes-and how can you spot them?,* THE GUARDIAN (Jan. 13, 2020, 5:00PM), https://www.theguardian.com/technology/2020/jan/13/what-are-deepfakes-and-how-can-you-spot-them.
45. Id.
A report by the AI firm Deeptrace indicated that fifteen thousand deepfake videos were online in September of 2019. A disconcerting 96 percent of deepfake videos online were non-consensual pornography that disproportionately victimizes women. While these videos are often given an indicator that explicitly states that it has been digitally altered, the impact on the depicted individual’s social status, employment prospects, personal relationships, and digital image may be irreparably damaged.

A. Deepfakes effects on private individuals

In a case that garnered international attention for criminal charges against a suburban Pennsylvania mother Raffaela Spone was accused of creating a deepfake video to tarnish the reputation of her daughter’s cheerleading competitor. Initially, prosecutors brought a claim contending that Spone “spearheaded an advanced deepfake video scheme.” After their investigation, the police were unable to confirm whether or not the video evidence in question were truly falsified using synthetic media replacement technology. As a result of this action, Spone was vilified and received death threats online, where she was commonly referred to as “deepfake mom.”

In a segment of the Washington Post, three digital-forensic experts were brought in and shown the video evidence in question. Ultimately, they determined that it was “highly unlikely”

46. Id.
47. Id. (99% of deepfake pornography mapped faces from female celebrities onto porn stars).
50. Id.
51. Id. (While investigators originally believed at least one video showed evidence of the use of so-called Deep Fake face replacement technology, police are at this point unable to confirm the video evidence was falsified).
52. Id. (Her attorney Robert J. Birch said Raffaela Spone was subject to “ridicule, embarrassment and harassment” at home…her reputation right now is less than mud).
53. Id.
for the video to be synthetically manipulated and appeared to be “blatantly authentic” since it had none of the indications or give away of a deepfake video.54 Experts were not able to draw definitive answers due to the poor video quality and lack of evidence.55 Skeptics such as law professors Danielle Citron and Robert Chesney introduced the idea of “Liar’s Dividend,” which describes the concept of using deepfake as a “scapegoat to avoid accountability for things that are in fact true.”56 This case shows that the use of deepfakes can have consequences on two fronts. It can create doubt in authentic evidence.57 Conversely, it can be used to pass fraudulent evidence as authentic.58 As the synthetic media techniques gain more notoriety, it may be worth discussing the validity of media evidence used in the courtroom when authenticity is brought into question.59

B. Deepfakes in entertainment

In April 2020, State Farm launched a thirty-second advertisement video that seemingly showed a segment in 1998 of ESPN analysts making disturbingly precise predictions about events in 2020.60 The clip stupefied audiences and elicited feelings of surprise, amusement, and delight.61 Apart from the initial reaction, they should have also felt deep concern that this seemingly innocent use of synthetic media for entertainment purposes could also be used by those with bad intentions to manipulate public opinion. During the webinar with Alex Cohen, the Director of Emerging Technologies for the GSA, stated, “I think the biggest surprise is that deepfakes have not been used yet for more destructive purpose. Most of them are simply politicians or celebrities saying and doing stupid things rather than anything

54. Id.
55. Harwell, supra note 49.
57. Id. at 1785.
58. Id.
61. Toews, supra note 3.
In an effort to raise concerns about the potential dangers of deepfake technology in politics and propaganda, the comedian, Jordon Peele, played in a deepfake clip of former President Barack Obama and said, “President Trump is a total and complete dipshit.” Peele’s message was to inform the general public to stay vigilant in the new era of technology where “our enemies can make anyone say anything at any point in time.” If a comedian can produce convincing, manipulated media with the help of media companies, it is scary to imagine what motivated foreign intelligent entities can do with the deepfake technology.

C. Deepfakes impact on businesses

Deepfakes are also capable of mimicking the tone, cadence, and accent of a voice. In early 2019, a UK-based energy firm wired €220,000.00 to a business in Hungary after listening to the instructions of their firm’s leader over the phone. Only it wasn’t the business leader, but thieves that used AI technology to replicate the German business leader’s voice. The CEO of the company stated that he recognized the “slight German accent and the melody” of his chief’s voice. The transferred money was redirected to Mexico and into several different accounts. Businesses need more awareness of the potential threats, risk level, exposure, and vulnerabilities of their operations to synthetic media manipulations to prevent heavy business losses.

---

63. BuzzFeedVideo, You Won’t Believe What Obama Says In This Video!, YOUTUBE (Apr. 17, 2018), https://www.youtube.com/watch?v=cQ54GDm1eL0&t=40s&ab_channel=BuzzFeedVideo.
64. Id.
67. Id.
68. Id.
69. Id.
70. Id. (“What’s for certain, however, is that technology for this type of crime does exist, and it’s only a matter of when the next attack will happen and who will be the target”).
III. Affected Areas of the Law

In the modern era, the internet provides a wealth of information restricted only by an individual's curiosity and their appetite for knowledge. Do you need a new recipe to impress a significant other? Google it. Do you need a crash course on subatomic particles to prepare for a physics final? A ten-minute YouTube video can break it down for you. Are you preparing for an interview with a prospective employer? Look up their LinkedIn profile to look for talking points. The people we meet and the friends we make are no longer limited to the proximity of our physical encounters. On this new highway that we call the world wide web, piloting computers and cellphones as our metaphorical vehicles, we are connected with people across the globe through the internet, social media, and online forums. In our hasty desire to participate in the socially driven, interconnected world, we click on the “create an account” button and skip the terms and conditions, unaware of the ramifications. The issue arises when we project our digital identity to businesses that profit from the big data generated from our voluntary participation.71 The principles of many data protection laws can be simplified into seven points: (1) lawfulness, fairness, and transparency; (2) purpose limitation; (3) data minimization; (4) accuracy; (5) storage limitation; (6) integrity and confidentiality; and (7) accountability.72

A. Deepfake Issues Arising in Personal Data Protection

As the world becomes increasingly engaged with the internet and social media, stringent data protection laws are warranted to ensure compliance. The purpose of data protection regulations is to protect the personal data of the people and regulate businesses to determine proper practice in their control over the information.73 In consideration of the issues arising from deepfake technology, we must contemplate whether such synthetic media...
itself is considered personal data. According to the General Data Protection Regulation (GDPR), the European Union (EU) defines personal data to mean “any information relating to an identified or identifiable natural person (‘data subject’).” According to the GDPR, something as personal data, it must involve three components: (i) information, (ii) an identifiable natural person; (iii) to which such information is related. In the same provision of the GDPR, the article defines an identifiable natural person as:

An identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier, or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural, or social identity of that natural person.

Under the GDPR, the definition of an identifiable natural person is broad.

Most of the deepfake content available involves superimposing the facial structure and likeness of a targeted individual onto the bodies of another individual in existing videos. More sophisticated deepfake content mimics not only the targeted individuals’ face but also their voice and body movements. In the first instance, it is easy to see that questions of personal data come in two forms: (1) the individual the body belongs to, and (2) the targeted individual that the deepfake seeks to imitate. It should be immediately apparent that the bodies in existing videos are personal data of the data subject that the video is displaying. In other instances, deepfake technology utilizes the available software and data, such as the images, videos, and voice recordings of the targeted subject to create a digital

76. European Union Regulation, supra note 74, at art. 4.
78. Yildirim, supra note 75.
79. Id.
80. Id.
doppelganger. 81 In light of these new software and technology, an analysis of whether recreations of target subjects’ likeness are considered personal data within the definition of the GDPR. 82

For data protection laws, such as the GDPR, to provide coverage for data subjects, the deepfake content must be considered personal data as defined in Article 4. 83 Scholars and legal practitioners would argue that the “use of the individual’s name or mimics without the data subject’s consent might constitute a personal data violation.” 84 The major issue associated with the creation of deepfake content is not simply the usage of an individual’s name but the use of the likeness of another individual that is easily recognizable by the population. 85 It is easy to imagine that a regular individual’s image or prestige can also be damaged if malicious deepfake content is dispensed to the internet and found by their local population.

If deepfake content is established as personal data subject to the protection of the GDPR, the law grants remedies such as the right to request the erasure of the deepfake content and request compensation if the relevant natural person suffered damages as a result of “unlawful processing of personal data.” 86 Article 4(2) of the GDPR defines “processing” as:

Any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organization, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure, or destruction. 87

The GDPR’s overinclusion of terms to define processing signifies the legislative intent for broad interpretation.

Although the issue of whether deepfake content is defined as personal data within the contemplation of the GDPR is unresolved, the creation of deepfake content is unambiguously

81. Id.
82. Id.
83. European Union Regulation, supra note 74, at 33.
84. Yildirim, supra note 75.
85. Id.
86. Yildirim, supra note 75.
87. European Union Regulation, supra note 74.
considered processing, as broadly defined by the GDPR. As discussed, synthetic media, such as deepfake technology, uses personal data such as original pictures, voices, and videos of the targeted data subject as a foundation to generate deepfake materials. The act of processing personal data is lawful so long as it is performed in the context accepted by the GDPR. Absent consent from the data subject, considering the common usage and purpose of deepfake content, it is unlikely that the process of creating deepfake material would fall under the six conditions expressed in Article 6 of the GDPR.

B. Deepfake Issues Arising in Human Rights

Although synthetic media pose a threat to our society by exacerbating the spread of misinformation, legislation that outright bans deepfake content may violate the concept of freedom of expression, which many would argue is a fundamental right. Under Article 19 of the Universal Declaration of Human Rights (UDHR), it is widely recognized by members of the United Nations that freedom of expression is a fundamental right that

88. Yildirim, supra note 75.
90. European Union Regulation, supra note 74 (These grounds are established in Article 6 of the GDPR, which provides in relevant part, processing shall be lawful only if and to the extent that at least one of the following applies: (a) the data subject had given consent to the processing of his or her personal data for one or more specific purposes; (b) processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract; (c) processing is necessary for compliance with a legal obligation to which the controller is subject; (d) processing is necessary in order to protect the vital interests of the data subject or of another natural person; (e) processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller; (f) processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child).
91. Yildirim, supra note 75.
should be universally protected. The UDHR provides that “everyone has the right to freedom of opinion and expression. This right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers.” This contention is further expressed in the International Covenant on Civil and Political Rights, which states that “everyone shall have the right to freedom of expression; this right shall include freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice.” The United States Constitution specifically protects free speech as a first amendment right. US case law distinguishes classifications of speech that are granted little to no protection. Some classifications include (1) obscenity, (2) defamation, (3) fraud, (4) incitement, (5) fighting words, (6) true threats, (7) speech integral to criminal conduct, and (8) child pornography, to name a few.

According to the Supreme Court of the United States (the Court), for material to be obscene, which would bar the material from First Amendment protection, the material must “appeal to the prurient interest in sex,” depict sexual conduct in a clear and unequivocal offensive way, and lack “serious literary, artistic, political, or scientific value.” While the Court has not offered guidance on the constitutionality of deepfake content and technology, considering that 96 percent of deepfake videos are pornographic, the Court may find the content too obscene for First Amendment protection.

---

94. Id.
95. Id.
97. Reid, supra note 92, at 211.
99. Id.
101. Aja Romano, Deepfakes are a real political threat. For now, though, they’re mainly used to degrade women., Vox (Oct. 7, 2019, 7:00 PM), https://www.vox.com/2019/10/7/20902215/deepfakes-usage-youtube-2019-deeptrace-research-report.
For claims under defamation, the Court recognized that there must be a balance between false statements about an individual and hindering free speech due to the penalties incurred from making such statements. As a result, the Court places the burden on the party alleging defamation to demonstrate the intent of the speaker to arise to a certain level or to prove certain injuries. Another form of unprotected speech that the Court may find applicable to the creation of deepfakes is fraud. Fraud is also unlikely to prevail because the Court has historically adopted the concept of the “marketplace of ideas,” where an “uninhibited, robust, and wide-open debate” is granted significant weight in First Amendment protection. In relevant precedents, the Court—in a plurality opinion—has recognized that “some false statements are inevitable if there is to be an open and vigorous expression of views in public and private conversations.”

An article written by Suyoung Baek, “Free Speech In The Digital Age: Deepfakes And The Marketplace Of Ideas,” analyzes issues that may arise with the First Amendment in drafting legislation to regulate deepfake content and creation. Baek explains that the unique characteristics of deepfakes warrant a “different, less conventional approach to understanding the constitutional value of false speech, especially harmful false speech, in the digital age.” Baek’s arguments can be summarized in three points: (1) deepfakes unprecedented capability to spread altered content to a significant amount of individuals are sufficient grounds for a departure of the protections granted to false speech by the Court; (2) deepfakes’ threaten the democratic process through the exposure of misinformation and disinformation to manipulate public opinion; and (3) deepfake technology presents unprecedented forms of exploitation, intimidation, and personal sabotage. Baek further discusses the application of antiquated frameworks, such as false speech in constitutional

106. Baek, supra note 104.
107. Id.
108. Id.
challenges, and the shortcomings of such an analysis as applied to the exponential growth of artificial intelligence and deepfake technology. Ultimately, when the Court inevitably faces the constitutional challenges associated with artificial intelligence technology, a new framework is likely to arise to balance the fundamental right with the risk of deepfake usage.

C. Deepfake Issues Arising in Right to Privacy

Since states and nations are slow to adopt legislation to provide remedies for individuals injured by deepfake content, we must look to traditional remedies of the law within the scope of tort law. In addition to the possibility of a defamation claim, the tortfeasor—in this case, a creator of deepfake content—may also be liable for a claim under intentional infliction of emotional distress (IIED).

Under this privacy tort action, a plaintiff can recover from the defendant for the trauma they suffered because of their behavior that blatantly oversteps all reasonable bounds of decency. IIED is notorious for having a difficult standard to meet, in addition to the potential defenses to IIED. In most cases, the hardest element to establish for a claim under IIED is the tortfeasor’s intent to cause emotional distress. While the creation of the altered media is likely to be seen as a deliberate act, the court is unlikely to find intent to cause emotional distress. Although not an absolute defense to IIED, parody is generally protected regardless of how distasteful or distressing it may be to the subject. While it is unclear whether a court would connect a deepfake content creator’s act of altering and manipulation as an intent to cause emotional distress, IIED remains part of the arsenal that could potentially grant remedies to victims of deepfake material.

In some jurisdictions, another privacy action that a plaintiff may consider is a right of action under the “false light” theory. Under this theory, the victim can sue the individual that

109. Id. at 14.
111. Id.
112. Id.
114. Gerstner, supra note 110.
115. Id. at 95.
“communicated it onward to others.”116 Potential defendants could involve ordinary individuals, websites, television channels, or newspapers.117 For a claim under the false light theory to be actionable, the material in question must be “communicated” or “shared” with a third person.118 Additionally, the burden is on the plaintiff to establish that the manipulated content does not represent the plaintiff misleads others and invokes a sense of embarrassment or offensive by the average individual.119

IV. CURRENT APPROACHES TO DEEPFAKE TECHNOLOGY

With the rising concerns of deepfake technology affecting a broad spectrum of areas such as election interference,120 private individuals,121 misinformation and disinformation concerns,122 and national security concerns,123 those in the legal profession are scrambling to apply traditional methods to an emerging and unprecedented technology. Given the number of resources and funding from the public124 to the private125 sector to research and develop the next innovation, it is unsurprising how the law lags

117. Id.
118. Id.
119. Id.
120. See Nicholas O’Donnell, Have We No Decency? Section 230 and the Liability of Social Media Companies for Deepfake Videos, 2021 U. ILL. L. REV. 701, 708 (2021). (Discussion of the potential usage for election interference in American elections from state actors—such as Russia, China, or Iran—and non-state actors—such as Oligarchs, Corporations, and Activist—could manipulate damaging false information to sway public opinion).
to keep up with expeditiously evolving technology.126 This part seeks to provide a birds-eye view of the state of the legislation relevant to regulating deepfake content and creation. Starting at the federal and state level of the United States, we will then move on to the approaches of foreign nations such as Canada, the European Union, China, and Taiwan.

A. United States federal laws on deepfake content

Due to the scant nature of regulations involving the nascent topic of deepfake technology, prosecutors are forced to lean on existing federal laws—such as provisions under Title 18 of the U.S. Code—to hold culpable individuals accountable for their nefarious activities.127 These provisions are specifically tailored to prohibit communications involving ransom, extortion, and kidnapping threats, cyber-harassment, cyberstalking, unauthorized access to a computer, and stalking as interstate domestic violence.128 The proposed legislation, such as the Ending Non-consensual Online User Graphic Harassment Act of 2017 (the ENOUGH Act)129 was introduced as a means to combat “nonconsensual pornography,”130 but ultimately expired by the end of the 115th Congress.131 Similarly, the Malicious Deep Fake Prohibition Act of 2018132 was introduced to prohibit the distribution of deepfake with the intent of facilitating criminal or tortious conduct through interstate commerce.133 Unfortunately, the bill was introduced just prior to the 2018 government shutdown and expired by the end of 2018 with no cosponsors.134

128. *Id*.
130. Cyber Civ. Rights Org., https://www.cybercivilrights.org/faqs/ [Last visited Feb. 16, 2013, 1:12 PM] (Commonly known as revenge porn, nonconsensual pornography refers to the distribution of sexual or pornographic images of individuals without their consent, which includes images taken without consent or take with consent but later distributed without consent of those in the images).
133. *Id*.
134. Delfino, *supra* note 121.
On December 20, 2019, as part of the NDAA for Fiscal Year 2020, deepfake legislation was signed into law.\footnote{National Defense Authorization Act for Fiscal Year 2020, Pub. L. No. 116-92 (2019). [herein NDAA].} Section 5709 instructs the Director of National Intelligence (DNI) to produce comprehensive reports for the Congressional Intelligence Committees detailing the foreign weaponization of deepfakes on an annual basis and notify Congress of foreign activities relating to disinformation targeting US elections.\footnote{Gat S. 5709, 116th Cong. (2019).} The legislation specified that the report from the DNI is to include an assessment of the technical capabilities of foreign governments—such as the People’s Republic of China and the Russian Federation as well as their private sector, academic or nongovernmental affiliates—regarding the production and detection of machine-manipulated media.\footnote{Gat.} It also requires updated assessments of the potential usage of deepfakes that would harm national security interest with respect to (i) overseas or domestic dissemination of misinformation; (ii) attempted discrediting of political opponents or disfavored populations; and (iii) intelligence or influence operations directed against the United States and their allies that are subjected to Chinese or Russian interference.\footnote{Gat.} Other considerations include assessments of technologies to counter deepfakes; an analysis of the Intelligence Communities\footnote{6HH 0(0%(56 2) 7+( ,& ODNI, https://www.dni.gov/index.php/what-we-do/members-of-the-ic (last visited Feb. 16, 2023).} (IC) that should have the main responsibility of monitoring and developing such technologies and their current capabilities; and a list of recommendations to determine whether the IC needs additional resources to address the potential deepfake threat.\footnote{NDAA, supra note 135, at S. 5709 (“Director of National Intelligence...shall submit to congressional intelligence committees a report on the potential national security impacts of machine-manipulated media..., actual or potential use of machine-manipulated media by foreign governments to spread disinformation..., assessment of technical capabilities of foreign governments..., and updated recommendations to address national security threats posed by machine-manipulated media...”).}

The deepfake legislation also encourages and expedites the development of technology to counter deepfake issues.\footnote{Id. at S. 5724.}
5724 of the NDAA provides the establishment of the deepfakes prize competition.\textsuperscript{142} The DNI is instructed to carry out a program to award prizes, with a cap of five million dollars, to “stimulate the research, development, or commercialization of technology to automatically detect machine-manipulated media.”\textsuperscript{143} Since the addition of deepfake legislation to the NDAA in 2019, the National Security Commission on Artificial Intelligence (NSCAI) was established as an independent federal entity to conduct a review and provide recommendations on behalf of the DNI.\textsuperscript{144} On March 19, 2021, the NSCAI released an alarming report detailing the shortcomings and how ill-equipped the US currently is in its ability to “defend or compete in the AI era.”\textsuperscript{145} Much like the purpose of the NDAA, the focus of the report was tailored towards a concern for national security, as opposed to the regulation of the creation and dissemination of deepfake content.\textsuperscript{146} While the NDAA and the report from the NSCAI do scrutinize the vulnerabilities of the US to machine-manipulated media and AI technology, it does little to offer legislative construction.

Another bill adopted by the House is the Identifying Outputs of Generative Adversarial Networks (IOGAN) Act.\textsuperscript{147} With the

\begin{footnotesize}
\begin{enumerate}
\item[142.] Id.
\item[143.] Id.
\item[145.] NSCAI, Final Report 2021, https://www.nscai.gov/wp-content/uploads/2021/03/Final-Report-Digital-1.pdf (NSCAI states that unless drastic measures are taken, foreign states like China will surpass the U.S. as the world’s AI leader within a decade. The report further details the recommended steps to be taken in order for the U.S. to compete in the new “moon race.” The Executive summary of the 700+ page report recommendations can be broken down into two prongs: (1) Defending America in the AI Era; and (2) Winning the Technology Competition. For the former, the NSCAI instruct the President and Congress to establish a dedicated agency for AI and digital information; establish a foundation for AI integration; achieve a state of AI military readiness; reimagine and restructure the Intelligence community; scale up digital talents in government; and integrate AI usage into National Security. The later prong focuses on cultivating domestic talent and accelerating AI innovation domestically; revamping intellectual property policies; and working with international allies and partners to advance global adoption of digital infrastructure and technologies).
\item[146.] Id.
\item[147.] H.R. 4355, 116th Cong. (2019).
\end{enumerate}
\end{footnotesize}
same goal as Section 5724 of the NDAA, the IOGAN seeks to stimulate competition in the field of machine-manipulated media detection. The Director of NSF is directed to conduct outreach in the private, public, and academic fields in order to accelerate the development of feasible technological tools to combat deepfake content and creation. Similar to the NDAA and report from the NSCAI, the IOGAN Act focuses on accelerating the development of AI technology. Based on the intent of existing deepfake legislation, it is evident that the US is focusing on innovation. In contrast, private individuals who are victims of deepfake technology are left to decipher how to recover from their injuries using antiquated and impractical theories of relief. Although the focus on accelerating efforts to develop viable tools for the detection of machine manipulated media could benefit private individuals, it merely offers a safeguard, but no recourse for targeted individuals of deepfakes.

B. State laws on deepfake content

As the state with the largest population, and reports of cybercrime, California was also one of the first states to adopt laws that regulated deceptive media. Section 1708.86 of the California Civil Code provides a private right of action against individuals that intentionally disseminates sexually explicit material—where the individual knows or have reason to know that the depicted individual in the sexually explicit material did not consent to the creation or disclosure—regardless of whether the individual created the material or not. The statute defines depicted individual as “an individual depicted in a realistic

---

148. Id.
149. Id.
150. Id.
151. Id.
152. Baek, supra note 104.
153. Id.
157. CAL. CIVIL CODE § 1708.86 (West 2020).
digitized performance in which the individual did not actually perform” and further explains that “digitized” includes “depicting the nude body parts of another human being as being those of the individual or imposing digitally created nude body parts onto the individual.”\textsuperscript{158} The statute also attempts to evade conflicts with First Amendment protections by providing exemption from liability where the material is a matter of legitimate public concern; political or newsworthy work; or otherwise protected by California or US Constitution.\textsuperscript{159}

Additionally, California amended the election codes to prohibit the distribution of “materially deceptive audio or visual media” of a candidate on the ballot with intentions of damaging the candidate or swaying public opinion within sixty days of an election.\textsuperscript{160} While cases have been scarce for the courts to decide how the new California legislations withstand free speech considerations, we can expect judicial insights in the years to come as more deepfake content is promulgated.\textsuperscript{161}

Alternatively, Texas enacted legislation to prohibit the use of deepfake content to influence a candidate’s reputation.\textsuperscript{162} Section 255.004(d) of the Texas Election Code categorizes an offense under the statute as a Class A misdemeanor stating: “A person commits an offense if the person, with intent to injure a candidate or influence the result of an election: (1) creates a deepfake video; and (2) causes the deepfake video to be published or distributed within 30 days of an election.”\textsuperscript{163} The statute also defined deepfake video as “a video, created with the intent to deceive, that appears to depict a real person performing an action that did not occur in reality.”\textsuperscript{164} Like the California law, Texas places a thirty day prior instead of sixty days period from the election day.\textsuperscript{165} Based on the language of the law, the legislative intent is to safeguard the election process from misinformation.\textsuperscript{166} Private victims of deepfake content must seek

\textsuperscript{158} Id.
\textsuperscript{159} Id.
\textsuperscript{160} Sierra, supra note 156.
\textsuperscript{161} Ajder, supra note 77, at 5.
\textsuperscript{162} V.T.C.A., Election Code §255.004 (West 2021).
\textsuperscript{163} Id.
\textsuperscript{164} Id.
\textsuperscript{165} Id.
remedies with a patchwork theory of cyberbullying, extortion, harassment, and revenge porn. Virginia extended its nonconsensual-pornography statute to include altered media by redefining “another person.” New York expanded the right to publicity by providing limited protection against the commercial exploitation of deceased performer’s “likenesses.” Some scholars argue that deepfake pornographic videos should receive the same treatment as prohibitions on nonconsensual pornography under the First Amendment.

C. International communities’ approach to deepfake laws

On the international scale, the regulation of synthetic or manipulated media remains underdeveloped or in its early stages. Like the US, Canada must also rely on tort, criminal, and election offenses that predate digital technologies for deterrence and redress. The European Commission took on a balancing approach by publishing a proposal to harmonize the usage of AI technology with the values and fundamental rights of EU citizens. Most notably, the proposal allows for permissive use of deepfake technologies but sets forth minimum requirements


168. VA. CODE ANN. § 18.2-386.2 (West 2019).


170. Matthew B. Kugler & Carly Pace, Deepfake Privacy: Attitudes and Regulation, 116 NW. L. REV. 611, 673 (2021) (arguing that prohibitions on deepfake pornographic videos should receive the same treatment under the First Amendment as prohibitions on traditional nonconsensual pornography rather than being dealt with under the less-protective law of defamation).

171. See Elizabeth F. Judge & Amir M. Korhani, Deepfakes, Counterfeits, and Personality, 59 ALBERTA L. REV. 1, 5 (2021) (discussing how laws protecting personality interests fill the gap created by emerging digital technology such as deepfake).

172. European Commission, Proposal for a Regulation Laying Down Harmonized Rules on Artificial Intelligence (Artificial Intelligence Act) (2021) (taking on a risk-based approach by categorizing AI systems into (1) unacceptable risk, (2) high risk, (3) limited risk, and (4) minimal risk. Depending on how the AI systems are categorized, the results range from an outright ban to non-regulation).
such as transparency obligations.\textsuperscript{173} A critical missing component is that the proposal does not set forth definitive guidelines for altered media disclosures, which gives the content creator the discretion on the method of disclosure.\textsuperscript{174} Additionally, labeling the content as manipulated media absolves the malicious actors of their conduct but does not account for the intimidation, bullying, and reputational damage to the targeted victims.

The People’s Republic of China followed a similar approach by passing a law that placed the burden of identifying and marking or removing unlabeled content on platform operations and app providers.\textsuperscript{175} This law, which came into effect on January 1, 2020, forbids the production and spread of fake news and must therefore be removed upon identification.\textsuperscript{176} China’s approach to the issue of online anonymity was to pass the Cybersecurity Law that contained a provision—which was highly scrutinized abroad—requiring real-identity authentication, such as a valid mobile phone number or government-issued IDs for the registration of online services.\textsuperscript{177} As an alternative to automation, which is the use of AI’s to detect manipulated media, Taiwan takes on a manual approach coined “nerd immunity” by deploying hundreds of professional fact-checkers and galvanizing the general population to recognize false news and encourage checking the truth of online content.\textsuperscript{178} Sovereign states should consider automatic and manual approaches in addition to formulating legislation to combat disinformation and misinformation that can result from deepfake content.

\textsuperscript{173}  Id. at art. 52(3) (Creators of deepfakes are obligated to provide disclosures to viewers that the content they are about to view is altered media).

\textsuperscript{174}  Id.

\textsuperscript{175}  See China Netcom, Notice on Printing and Distributing the “Regulations on the Administration of Online Audio and Video Information Services, CYBERSPACE ADMIN. CHINA (Nov. 29, 2019, 2:00 PM), http://www.cac.gov.cn/2019-11/29/c_1576561820967678.htm.

\textsuperscript{176}  Lavender Au, China targets ‘deepfake’ content with new regulation, TECHNODE (Dec. 3, 2019), https://technode.com/2019/12/03/china-targets-deepfake-content-with-new-regulation/.

\textsuperscript{177}  See Eliza Gkritsi, Dust has yet to settle two years after China’s landmark cybersecurity law, TECHNOLOGY (Jun. 10, 2019), https://technode.com/2019/06/10/dust-has-yet-to-settle-two-years-after-chinas-landmark-cybersecurity-law/.

D. Potential regulatory gaps of the existing legal framework

Reliance on legal frameworks that preceded the digital era is likely to leave gaps for malicious actors to circumvent the system to avoid liability.\textsuperscript{179} To explore potential vulnerabilities of existing protections, the European Parliamentary Research Service released a comprehensive study in 2021 that accessed the sufficiency of European policies against synthetic media.\textsuperscript{180} The study breaks down the implications of deepfake technology against the existing European legal framework into three fictional scenarios: (1) deepfake pornography, (2) false political statements, and (3) manipulated court evidence.\textsuperscript{181} The study analyzes the thought experiment by identifying the actors into (1) the perpetrators, (2) the victim, (3) the technology provider, (4) the original performers, (5) the original author, the platform, and (6) the platform users.\textsuperscript{182}

1. Deepfake pornography

The following fictional scenario illustrates how deepfake technology could be weaponized to discredit professionals and harm private individuals who are targets of spiteful actors.

Imagine that you are a female investigative journalist who frequently writes critical commentary on political and socio-economic issues in your country. Recently, you wrote a piece on a corruption scandal that involved several politicians. You are used to being stifled by adherents of the politicians you write about, who often post cruel gossip and lies about you on social media. One day, you receive a message from a friend saying that there is a pornographic video of you circulating online. You are certain that this is impossible, but when you see the video, you are shocked. It is your face copied on someone else’s body. It’s a deepfake video, intended to harm and discredit you and your work.\textsuperscript{183}

Aside from instances of clear content for purposes of entertainment, satire, social, or political commentary, efforts to target


\textsuperscript{180} European Parliamentary Research Service (hereinafter EPRS), \textit{Tackling deepfakes in European Policy}, PE 690.039 STOA (2021).

\textsuperscript{181} \textit{Id.} at 48.

\textsuperscript{182} \textit{Id.} at 48-54.

\textsuperscript{183} \textit{Id.} at 48.
journalists, commentators, political figures, and those in positions of authority with manipulated media indistinguishable from actual media diminishes the functions of a democratic society.184

Notwithstanding a victim’s right to a claim, enforcement of any law remains encumbered by the borderless nature of the internet.185 Sophisticated perpetrators that go to great lengths to conceal their identities online are difficult for law enforcement to identify, and much more so for victims lacking resources.186 As a result, victims are left with limited options such as contacting the platform to request removal of the content.187 Such request is unlikely to be helpful because once media is circulating through the internet, it becomes increasingly more difficult to remove as time passes.188

2. False political statement

The second fictional scenario hypothesizes how foreign intelligence services can utilize deepfake technology to influence other sovereign states.

Deepfakes offer malicious actors new opportunities for manipulating public opinion. Imagine a situation in which organized actors, such as foreign intelligence services, are aiming to undermine trust in European politics. To this end, they produced a deepfake video that shows several European health ministers in a confidential conversation, saying that they are deliberately withholding vaccine supplies. They distribute the deepfake via different platforms using ‘social bots’- algorithms that autonomously produce content and imitate human behavior. The video spreads quickly as it gets picked up by other social media users who believe the video is authentic and redistributes it.189

While political figures in the public eye are regularly subjected to criticism and opposition, falsified materials can aggravate tensions that could potentially put their professional reputation

184. Id. at 49.
185. Id. at 51.
186. EPRS, supra note 180, at 51.
187. Court, supra note 48.
189. EPRS, supra note 180, at 51.
and lives at risk. Deepfake technologies are powerful tools that can be used to instigate discourse between different nations. A sovereign nation confronted with a cross-border cyberattack from a foreign intelligence service can complicate the identification of the organization responsible for the attack without international cooperation. While allied countries are presumed to be willing to assist in the exchange of information, there is no concrete obligation that arises from international law pertaining to deepfakes specifically.

3. Manipulated court evidence

The final scenario exhibits how deepfake technology can complicate the validity of admitted evidence.

Imagine you are a popular politician, swiftly climbing the country’s political ladder. This upsets those who are in power, as it seems you will gain a huge victory in the upcoming elections. Then, to your shock and disbelief, an audiotape emerges. The public can hear your voice in a telephone conversation speaking to an unknown person, discussing the possibility of taking bribes. Shortly afterward, the police arrest you. Election day passes by while you are in prison in anticipation of a criminal trial for corruption. The telephone conversation is the result of an AI-generated synthetic deepfake, devised by your political opponents. It is your voice, but you never said those things. How will you convince the judge that the audiotape is forgery?

This scenario highlights two major issues associated with falsified information, irreversible consequences, and undetectable falsified evidence.

The gathering of evidence is an integral part of the civil and criminal process for the judiciary to render a fair and impartial decision. While attempts to mislead the judiciary with manipulated evidence is not a modern development, the existence of easily accessible technology such as deepfake—which is unprecedently deceptive—is likely to be raised as grounds for contesting

190. Id.
191. Id. at 29.
192. EPRS, supra note 180, at 53 (discussing how existing regulations are insufficient to account for situations where a foreign state is a perpetrator that issued the deepfake since their identities are usually not found).
193. Id. at 53.
194. Id.
the authenticity of the evidence. Legal scholars and courts have heavily debated on whether there should be a presumption of the authenticity of evidence or whether eyewitness or expert testimony is necessary. The need to revisit the rules of evidence is buttressed by the novel degree of duplicity in deepfake technology. Alternatively, some proponents claim that current rules for authentication are adequate as-is since the question of authenticity can be raised to strike (in civil proceedings) or exclude (in criminal proceedings). Based on case law, scholars contend the courts are confident that existing procedures are sufficient to exclude manipulated media from the court system.

V. INTEGRATING DEEPFAKE REGULATIONS INTO INTERNATIONAL LAW

The examination of current usage of deepfakes, theoretical scenarios, and existing regulations overwhelmingly indicates that fragmented efforts to regulate borderless issues—like deepfake technology—are ineffective without addressing issues of extra-territorial jurisdiction, and international cooperation. This Part, therefore, proposes the adoption of language from the GDPR into the Convention on Cybercrime (Budapest Convention). The Budapest Convention was codified as an effort to harmonize domestic criminal law governing cyberspace with the community of nations to promote a mutual obligation to assist in the sharing of information and streamline efforts of investigation. It does so


199. Id.

substantively, in part, by requiring signatories to enact legislation to criminalize offenses specified by the Convention.201 It also details the procedural framework oriented towards an international network to establish a symbiotic relationship for gathering information,202 extradition,203 and jurisdiction.204 In formulating the substantive and procedural structure of the legislation, signatories must be mindful of international human rights treaties and conventions, such as the International Covenant on Civil and Political Rights.205

Substantively, the Budapest Convention requires participating nations to criminalize: (1) “[o]ffenses against the confidentiality, integrity, and availability of computer data and system,”206 (2) “computer-related offenses,”207 (3) “content-related offenses,”208 and (4) “criminal copyright infringement,”209 among other offenses.210 Under Title 3 of the Budapest Convention, content-related offenses currently only criminalize offenses related to child pornography.211 The Convention is silent on criminalizing non-consensual/revenge porn and manipulated media content. As many legal scholars have discussed in recent years, as is the case for this note, unregulated access, and usage of deepfake technology can cause irreparable harm to private individuals’ image, threaten national security, and exacerbate issues of misinformation.212 Therefore, this Note proposes the following amendment to Title 3 of the Budapest Convention:

**Offenses related to synthetic media**

1. Each Party shall adopt such legislative and other measures as may be necessary to establish as criminal

---

201. *Id.* at 3-7.
202. *Id.* arts. 25-27, 31-32.
203. *Id.* at art. 24.
204. *Id.* at art. 22.
205. ICCPR, *supra* note 96.
207. *Id.* at 5.
208. *Id.*
209. *Id.* at 6.
210. *Id.*
211. *Id.* at 5.
offenses under its domestic law, when committed intentionally and without right the following conduct:

a. Producing synthetic media, without the consent of identifiable natural persons and without legitimate interest, for the purpose of its distribution through a computer system;

b. Offering or making available synthetic media, without the consent of identifiable natural persons and without legitimate interest, through a computer system;

c. Distributing or transmitting synthetic media, without the consent of identifiable natural persons and without legitimate interest, through a computer system.

2. For the purpose of paragraph 1 above, the term “synthetic media” shall include media:

a. the existence of which required the processing of personal data;

b. designed to mislead or deceive the viewers into believing the occurrence of an event when it has not actually occurred.

3. For the purpose of paragraph 2 above, the term “personal data” shall include:

a. any information relating to an identified or identifiable natural person.

4. For the purpose of paragraph 2 above, the term “processing” shall include:

a. Any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organization, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure, or destruction.

5. For the purpose of paragraph 1 and paragraph 3 above, the term “identifiable natural persons” shall include:

a. one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier, or to one or more factors specific
to the physical, physiological, genetic, mental, economic, cultural, or social identity of that natural person.

6. For the purpose of paragraph 1 above, the term “legitimate interests” shall include:
   a. Entertainment, satire, social or political commentary, and other legitimate interests determined by the signatory Party.

The proposed amendment mirrors the language of the GDPR discussed in Part III of the note. Section (1) of the proposed amendment sets out the conduct that signatories nations must criminalize. It provides that the conduct of creating synthetic media is not subject to criminalization unless the perpetrator acted without the consent of the identifiable natural person and without legitimate interest. The exception of legitimate interest leaves room for permissive usage of deepfake technology. Section (6) of the proposed amendment established the floor of legitimate interest where signatory nations can build upon based on their independent interest. Taken from the GDPR, the definition of “identifiable natural person” is sufficiently broad to encompass the person in the original video (the body), and the person in the fabricated video (depicted individual/face). Consent of both individuals would be required for the creation of deepfake content. Similarly, the definition of “processing” taken from the GDPR should be sufficiently broad to cover the mechanisms to create deepfake content. A substantive integration of the language in the GDPR into Title 3 of the Budapest Convention conforms with its interest to deter actions directed against confidentiality, integrity, misuse of systems, networks, and data.

Procedurally, the Budapest Convention requires signatory nations to enact legislation granting their governments’ enforcement departments the power and procedures to investigate the offenses criminalized in the convention. Substantive laws criminalizing conduct without adequate enforcement leads to

214. Id.
216. Budapest Convention, supra note 200, Preamble.
217. Id. at art. 14.
superficial efforts. To effectively deter criminalized conduct, nations drafting legislation must balance public interest and the safeguards provided under domestic and international law. In addition to the common provisions, parties to the Convention are required to enact legislation for (1) expedited preservation of stored computer data, (2) production order, (3) search and seizure of stored computer data, and (4) real-time collection of computer data. Effectively, the inclusion of the proposed offenses to synthetic media would render the attachment of the procedural requirements of the Budapest Convention.

Chapter III of the Budapest Convention provides the obligations of parties regarding international cooperation. Under Article 25 of the Budapest Convention, it states that “[P]arties shall afford one another mutual assistance to the widest extent possible for the purpose of investigations or proceedings concerning criminal offenses related to computer systems and data, or for the collection of evidence in electronic form of a criminal offense.” This section of the convention reinforces the significance of the exchange of information by setting the scope and procedures of mutual assistance. The inclusion of a provision for the criminalization of offenses related to synthetic media resolves some of the major challenges of deepfake technology.

First, the misuse of deepfake—particularly in the digital space of pornography—would be criminalized at the national level for signatories of the Budapest Convention. Second, the international cooperation for the exchange of data would assist in the identification of foreign intelligent services disseminating deepfake content in matters of national security. Third, victims are

218. Id. at art. 15.
219. Id. at 8-9.
220. Id. at 9.
221. Id. at 9-10.
222. Budapest Convention, supra note 200, at 10-11.
223. Id. at 12.
224. Id. at art. 25.
225. Id.
226. Delfino, supra note 121 (discussing how there are currently no federal laws that criminalize the creation or distribution of pornographic deepfakes and proposes a national response is required due to the potential harm of deepfake technology).
227. Budapest Convention, supra note 200, at art. 45 (In circumstances where a signatory to the Budapest Convention is the perpetrator, thus
in a better position since there would be identifiable laws for a cause of action instead of relying on traditional tort and criminal law for remedy and redress.\textsuperscript{228} Some areas that remain unresolved are the identification of perpetrators; instances where the victim and perpetrator reside in different jurisdictions; and international conventions naturally require sovereign nations to opt into the multilateral agreement.

Due to the borderless nature of cyberspace, anonymous perpetrators are likely to be able to remain elusive.\textsuperscript{229} On this front, governmental intervention into researching and developing methods of tracking online activity would be instrumental. Adopting China’s approach to require identifiable information the registration of online services may be unfavorable given how the approach was scrutinized by the international community.\textsuperscript{230} Assuming the victim can identify the perpetrator, and obtain a binding judgment, the victim can seek redress, if provided by legislation, when both parties are situated in the same jurisdiction. Issues arise when the parties involved are situated in different jurisdictions and the judgments are rendered unenforceable. In such instances, unless there are international laws that recognize the judgment, victims are left to rely on principles of comity.\textsuperscript{231} Finally, since sovereign nations cannot be compelled to follow the obligations under the Budapest Convention unless they are signatories, its effectiveness relies entirely on the number of nations that opt into the convention.\textsuperscript{232} Most notably, Russian and China are not signatories of the Budapest Convention because they believe its provisions were intrusive.\textsuperscript{233}

\begin{footnotesize}
\begin{enumerate}
\item \textsuperscript{228} Delfino, \textit{supra} note 121.
\item \textsuperscript{229} Tsu, \textit{supra} note 178.
\item \textsuperscript{230} Gkritsi, \textit{supra} note 177.
\item \textsuperscript{231} Sung Hwan Co., Ltd. v. Rite Aid Corp., 7 N.Y.3d 78 (Ct. App. N.Y. 2006) (plaintiff sought enforcement of a foreign money judgment).
\item \textsuperscript{232} Benjamin Knight, \textit{How is International Law Enforced?}, ALL-AM. MODEL UN (2020), https://www.allamericanmun.com/how-is-international-law-enforced/.
\end{enumerate}
\end{footnotesize}
CONCLUSION

In light of deepfake technology’s extraordinary potential for creating deceptive content, efforts to curb the potential harms to private individuals, societal perceptions of information, and national security, are warranted. As deepfake technology improves, the need for superior detection technology is required. Individual governments should continue to sponsor and incentivize technology companies to produce increasingly sophisticated detection methods to combat deepfakes.234 Concurrently, informing the masses about the existence of deepfakes and training the people to spot indicators of falsified media is useful in detecting misinformation in general.235 On the legislative front, Delfino’s call for national criminalization of the creation and distribution of deepfake pornographic is justified by how it categorically targets mainly women.236 The adoption of the proposed offenses of synthetic media into the Budapest Convention should be sufficient to achieve that end.

Yi Yan*

235. Smith, supra note 178.
236. Delfino, supra note 121.

* B.B.A, CUNY Baruch College (2018); J.D., Brooklyn Law School (Expected 2023); Note Editor, Brooklyn Journal of International Law (2022-2023).

I would like to thank the staff and e-board of the Brooklyn Journal of International Law for their time and hard work. This Note is not just the product of my work but the culmination of the Journal staff that commented and edited the Note which led to its publication. I would like to thank the Managing Editor Samantha Past and Editor-in-Chief Lucie Couillard Sosa for their accommodations and walking me through the publication process. I would like to thank my closest companions Sagar Yadav, Elvis Landi, Rahul Ghai, Alvaro Hernandez, Sho Miyazaki, and Juan-Carlos Pleitez for existing because their presence brought joy into my life. I would like to thank my parents for laying the foundations of my success. I would also like to thank my ancestors, cannot forget the ancestors. All errors or omissions are my own.