







Individual Legal Protection in the Deepfake Technology Era

Zec Kie Tan ¹, Shao Zheng Chong ², Chee Ying Kuek ³
and Eng Siang Tay ⁴

¹ Student Researcher, Faculty of Law, Multimedia University, Jalan Ayer Keroh Lama, 75450 Melaka, Malaysia

² Student Researcher, Faculty of Law, Multimedia University, Jalan Ayer Keroh Lama, 75450 Melaka, Malaysia

³ Senior Lecturer, Faculty of Law, University of Malaya, 50603 Kuala Lumpur, Federal Territory of Kuala Lumpur, Malaysia

⁴ Senior Lecturer, Faculty of Law, Multimedia University, Jalan Ayer Keroh Lama, 75450 Melaka, Malaysia estay@mmu.edu.my

Abstract. The use of deepfake technology, which is video or image manipulation technology by superimposing the face of a person to the body of another, has become common in the modern days. Where it may be used for proper purposes such as in the entertainment industry, it is usually being abused by the users. Ninety-six percent of the deepfake videos and images are non-consensual pornography, while the others are used for the purpose of fraud, impersonation or misinformation. This caused the victims to suffer from economic loss, reputational loss and emotional loss. However, though there are available remedies for the victims to claim damages from the loss they have suffered, the conduct of misusing deepfake technology itself is not being governed in Malaysia. Hence, this paper seeks to identify the gaps in the protection over the victims of deepfake technology and recommend solutions to the legal issues raised, by referring to the US DEEP FAKES Accountability Act and the UK Online Safety Bill.

Keywords: Deepfake Technology, Non-consensual deepfake pornography, DEEP FAKES Accountability Act

1. Introduction

The use of the Internet started to bloom in the 1990s, which became the starting point of online interaction [1]. While technology continues to advance, artificial intelligence is being made available. Generally, artificial intelligence refers to the ability of the machines to simulate human intelligence, through machine learning. Deepfake technology is a form of artificial intelligence which is being widely used.

1.1 Definition of Deepfake Technology

© The Author(s) 2023

Y. C. Adam and S. A. B. Samsudin (eds.), *Proceedings of the 3rd International Conference on Law and Digitalization 2023 (ICLD 2023)*, Advances in Social Science, Education and Humanities Research 791, https://doi.org/10.2991/978-2-38476-154-8_7

Literally, the term “Deepfake” consists of the words “deep” and “fake”, which indicates that it is a combination of deep learning and fake production [2]. Deepfake technology is defined as a tool that aids in the alteration and manipulation of images and videos [3]. More specifically, it is a video manipulation technology, which allows the swapping of a person’s face in a video clip or image, to the face of another person [4].

An important characteristic of deepfake technology is that the manipulation is not noticeable or hard to distinguish from the authentic material. This is evident from another definition of deepfake technology, which is a technology that allows users to switch the face of an actor in a video with the face of another actor, in a photorealistic manner. It is being described as a technological advancement that produces hyper-realistic videos, leaving little trace of manipulation [5].

1.2 The Use of Deepfake Technology

The technology was created in 2014 by Ian Goodfellow [6]. It later became popular in 2017, when a group of Reddit users created synthetic celebrity pornography videos by AI technology [7]. Though deepfake pornography was banned on Reddit, the use of deepfake technology became common online, as evident in a statistic, the amount of deepfake video has increased by 550% from 2018 to 2022 [8]. Furthermore, the creation of deepfake videos is convenient, as there are many deepfake software available, such as Face Swap, DeepFaceLab, FaceApp, Wombo and more. Some social media applications, such as Snapchat and Tiktok also integrated the deepfake technology, as one of their features [9].

Though the use of deepfake technology can benefit certain industries such as film making, video games, fashion and e-commerce, it is by and large used for improper purposes, as the deepfake videos created are generally made without the consent of the person to whom the face was swapped. Therefore, any individual could become the victim of deepfake technology, where some of them may not even be aware about it. This could lead to many legal implications, including violation of individual rights, defamation and invasion of privacy.

The deepfake videos and applications are accessible in Malaysia. However, Malaysia does not provide any specific protection and limitation over its usage. This paper focuses on the discussion on the legislative protection for the individual victims of deepfake technology. The problem statement of this paper is the Malaysian laws do not provide sufficient protection over the victims of deepfake technology. This paper aims to identify the gaps in the protection over the victims of deepfake technology and recommend solutions to the legal issues raised.

The research methodology utilised is doctrinal legal research. The primary sources such as the legislation, case law and policies are used. While for secondary sources,

the journal articles, conference papers, news articles and more are accessed through Google Scholar, HeinOnline, LexisAdvance, Malaysian Current Law Journal, Westlaw, Emerald Insight and others.

2. Examining the Threats of Deepfake

Notwithstanding the beneficial side of the deepfake technology, it has been abused by unscrupulous persons resulting in fake news, misinformation, disinformation and deception that can threaten the society. The face swap involving the deepfake technology is usually done non-consensually, consequently depicting the victims in a negative way.

The creation of deepfake videos was done without the consent of the individual whom the face is being swapped, might be due to two reasons [10]. Firstly, the deepfake application is developed to be less demanding in their input for video creation. All that is required to complete a face swap is just a series of selfies with specific facial expressions and head postures [11]. Connecting to the second reason, the individuals' posting of selfies on social media could be easily found online by the video creators. The deepfake programs is also having the ability to explore different social medias through Google's Image Search, seeking the replacement of faces by automation [12]. These extraction of information and replacement of face is done without the consent of the individuals, which constitute an infringement of privacy and personal data [13].

Most of the non-consensual deepfake videos online are for pornography use. This could be evidenced from research, which showed that 96% of the deepfake videos posted online were categorised as pornographic video. More specifically, the victims of deepfake pornography videos are mainly the women, with rare occasions that men could be victimised [14]. Though the deepfake pornography do not depict the actual body of the victim, the fact that the face of the victim was being superimposed to a body performing a sexual act would cause an analogous harm to the person, as people would recognise the videos as genuine [15]. Worst still, such deepfake pornography is often highly viral and will spread rapidly through social media. Nonpornographic deepfake videos which are used to depict politicians and corporate leaders may potentially influence elections, tarnish the reputation of a corporation, undermine the confidence of shareholders, or manipulate the share price.

The non-consensual pornography would harm the victims in several ways, including reputational and emotional harm. In the case where these deepfakes pornography videos were not circulated online, it may be used against the victims through blackmailing. The individuals would also be receiving the deepfake videos or photos, which their face was being superimposed on. The victims would be required to provide money or value in exchange for the deletion or prevention of spreading of the deepfake media. For example, a mother of a high school cheerleader has created

deepfake videos, showing that her daughter's rival was drinking and smoking nakedly. For the destruction of the photos, the rival has to quit the cheerleading team [16].

Furthermore, the deepfake technology is also used for fraud through impersonation. Videos would be created by superimposing the face of a person doing something that he never did. With that misunderstanding, the video is used to defraud the person around him, causing them to suffer from losses. For example, a UK based subsidiary company in Germany was being defrauded by an impersonation of the CEO of their holding company, generated through deepfake technology. The "CEO" demanded a payment of 2 lac pounds to a supplier, where the money was later being transferred to a private account [17]. Other fraud techniques of deepfake include vishing and cloned voice.

3. The Legal Protection for Individuals Against the Deepfake Technology in Malaysia

Seeing the use of deepfake technology is getting more convenient and common in the modern days, it raises the concern that the individuals who are relatively weaker in power shall be given sufficient protection over the abuse of deepfake technology. Therefore, an examination on the protection and remedies available against Deepfake technology in Malaysia becomes necessary.

Though Malaysia has no specific legislation that deals with the use of deepfake technology, Malaysia does provide basic protection against the deepfake technology, by allowing the individuals to bring an action for the loss they suffered due to the misuse of technology.

3.1 The Legal Position of Deepfake Technology Abuse in Malaysia

In Malaysia, the only legislation that may possibly govern over the issue of deepfake is the Communications and Multimedia Act 1998 ("CMA"). Specifically, s 211(1) of the CMA provides that no one is allowed to provide online content, which is indecent, obscene, false or offensive in character, with the intention to annoy, abuse, threaten or harass any other person. Any person contravenes so is committing an offence that is punishable by a fine not more than RM50,000, or imprisonment for not more than one year or both, and a further fine of RM1,000 for continuing the offence after conviction, as provided in s 211(2) of the CMA.

This provision is legislated wide enough to cover most forms of misuse of deepfake technology, including the non-consensual pornography under indecency, impersonation under false information and more. However, the criminalisation is only limited to the final "providence" of the video through the online platform. The process of creating and preparing the video is not criminalised [18].

It is pertinent to note that deepfakes may also be abused by paedophiles to produce child sexual abuse material. This is sanctioned under ss 4 to 10 of the Sexual Offences Against Children Act 2017, including making, preparing to make, exchanging, publishing, and accessing child sexual abuse material. Deepfakes would fall within the definition of child sexual abuse material in s 4 if the output image or video appears to show that a child is engaging in sexually explicit conduct. It is worth noting that s 6 states that it is sufficient to criminalise the accused if it is satisfied that he is preparing to produce child sexual abuse material using deepfake technology.

While in terms of consequences caused to the victims, if the deepfake image or video is defamatory in nature to the victim, the victim is able to bring an action against the creator under tort law. In *Dato' Sri Dr Mohamad Salleh bin Ismail & Anor v Nurul Izzah bt Anwar & Anor* [19], the appellants sued the respondents for defamation in respect of the first respondent's statements in a press conference. The Federal Court followed the English case of *Charleston & Anor v News Group Newspapers Ltd* [20], which involved a superimpose of the plaintiffs' faces to near-naked bodies of models, engaged in pornographic poses. The House of Lords dismissed the plaintiffs' appeal as a libel claim could not be founded merely on a headline or photograph in isolation from the entire text. Whether an article was defamatory is determined by the reasonable reader's response to the whole publication. Using this as an analogy, if the deepfake image or video is *prima facie* defamatory but there is some additional information associated with the image or video, indicating that this is a superimposition, no remedy may be granted to the victim. This is because when the image or video and the words were read in entirety, an ordinary, reasonable and fair-minded reader would not have the impression that the victim actually did or said such a thing.

For the emotional distress suffered by the victim, it might be compensable. The general principle for awarding general damages in Malaysia could be seen in the case of *Sembaga Valli a/p KR Ponnusamy v Datuk Bandar Kuala Lumpur & Ors* [21]. If a person is injured by another's wrong, general damages may be granted for non-pecuniary loss including mental distress. However, there is no standard rule in measuring the damage caused. Hence, the court is given discretion to determine a fair and reasonable amount, based upon evidence tendered. As there is yet to be relevant case law in Malaysia, the position in granting the remedy for the victims of deepfake technology remains uncertain.

Furthermore, Malaysia is also criminalising the act of blackmailing and extortion. Section 384 of the Penal Code ("PC") provides that extortion is an offence, punishable by imprisonment up to 10 years, or with fine, or with whipping, or any two. While blackmailing by threatening to publish deepfake generated videos in exchange of money or valuables falls squarely within the definition of extortion under s 383 of the PC.

3.2 Legal Challenges Faced in Malaysia

Generally, there are some forms of legal protection given for the misuse of deepfake technology in Malaysia. However, these legal protections are weak and flawed. Firstly, there is no legal framework which is established specifically in regulating the use of deepfake technology itself, despite its popularity in modern society. This could be evidenced from the fact that there is no legislation for such a purpose, nor is there any effective discussion on the passing of law over the potential misuse of deepfake technology in Malaysia. In a broader sense, there is not even legislation or policy which provides protection specifically for the use of artificial intelligence.

Another legal challenge faced in Malaysia is that the laws in Malaysia only provide the protection when the victim has suffered from some loss. The misuse of deepfake technology is not criminalised. No liability is imposed on the improper use of deepfake technology. Therefore, no preventive measure could be taken to protect the potential victims before the videos or images are disclosed to the public online.

4. Other Jurisdictions

4.1 United States

In the United States, a bill was proposed in 2019 and again in 2021, namely the DEEP FAKES Accountability Act, which sought to specifically address the issues arising from the new technological threat of deep fakes. In both occasions, the bill did not receive a vote and the bill was not enacted. [22], [23]. Nevertheless, it was reported that there is a plan to reintroduce the bill in 2023 [24]. Generally, the bill specified the transparency requirements where disclosure of deepfake is mandatory. It required any advanced technological false personation record containing an audio and visual element to include verbal and written statements that identify the record's altered audio and visual elements and the extent of alterations. Failure to comply with the audio and visual disclosure or the act of removing the disclosure would subject the offender to both criminal and civil penalties. The spectrum of the criminal penalty was wide enough to cover different perspectives on the law. It is pertinent to note that s 2(f)(1) of this bill expressly criminalises people who use the advanced technological false personation record to commit the stated acts. This means that the law does not criminalise one who uses the deep fake technology, but those who use the products of the deep fake technology to commit the acts. From a criminal perspective, one is not criminally liable merely because of his failure to disclose. Instead, he is only liable if he has the intention to use deep fake products with sexual content to humiliate others, the intention to cause violence, commit fraud, or the intention to influence an election. In contrast, from a civil perspective, one is subjected to a civil penalty of up to \$150,000 per record and appropriate injunctive relief as long as he fails to disclose.

However, it seems that the bill criminalises the failure of disclosure rather than the act itself to cause harm. Hence, if the person does disclose, but with the intention to cause harm, will he be criminally liable? In *United States v. Tatum*, 2023 U.S. Dist. LEXIS 75482, the defendant was found to have generated deepfake nude images of

young girls, ex-girlfriends, and other acquaintances of the defendant. The defendant was charged with possessing and transporting child pornography and producing sexually explicit content of children. Notably, the defendant was only charged with offences against minors, but offences against majors, such as the defendant's ex-girlfriends and acquaintances, are not mentioned in the case. Instructively, the Supreme Court in *United States v. Alvarez*, 567 U.S. 709 decided that falsity alone is not sufficient to exclude expression from the protection of the First Amendment.

Furthermore, s 2(g) of the bill confers private entities the right of action to claim damages, and injunctive relief against one who fails to disclose or alter to remove the disclosure. The damages are categorised into four distinct levels, from the level of failure of disclosure to the level where the deepfake contains explicit sexual content to humiliate the victim. Section 2(h) further provides privacy protections to the victims of deepfake. If the action is brought by the federal authorities, reasonable measures should be taken to protect their privacy and minimise additional public viewings of the deep fake records. As for private actions, the parties could be permitted to file their petition under seal.

Section 7 of the bill further provides for the detection of deep fakes. The law establishes a Deep Fakes Task Force to provide support to the government in researching relevant technologies to tackle the issues arising from deepfakes. The government should also provide access to the technologies to the relevant private sector to foster collaboration.

Meanwhile, it is pertinent to note that s 230 of the Communications Decency Act provides immunity to online platforms to be exempted from any liability arising from user-generated content, even if they deliberately posted the content.

4.2 United Kingdom

There is currently no deepfake legislation in the United Kingdom, but the UK is currently enacting an Online Safety Bill. Sections 6(2) and 20(2) of the Bill subject the providers of user-to-user services and providers of search services to several duties of care, including the duty to conduct illegal risk assessments, the duty to ensure the users are safe from illegal content, and duties about complaints procedures. Furthermore, ss 33 and 34 stipulate that the regulated user-to-user service and regulated search service have the duty to prevent users from viewing fraudulent advertisements and swiftly take down such content after knowing the presence of the fraudulent advertisements.

5. Recommendations and Solutions

5.1 Technical Analysis

The relevant authorities have to invest in and develop their identification techniques through further research and training their data sets to establish more reliable detection tools. To prevent the detection tools from being outdated, the detection tools must also be constantly trained with updated data sets, as the deepfake technology is improving swiftly. Detection tools that are able to automatically detect the deepfake in media are required to ensure that the law can be enforced as soon as the deepfake is disseminated [25].

In light of the above, it is recommended that the relevant authorities collaborate with social media corporations to devise strategies aimed at countering the abuse of deepfake technology. Moreover, collaboration with other countries is recommended because cyberspace is not geographically situated within any specific country, yet deepfake technology can be easily disseminated around the world. This means that harmonisation of international laws is crucial to foster the effective implementation of legal regulations.

5.2 Legal Analysis

Despite the fact that Article 10(1)(a) of the Federal Constitution guarantees the right to freedom of speech and expression, it is subject to limitation under Article 10(2)(a). The Parliament assumes a significant function in identifying the boundaries for illegal deepfakes and imposing prohibitions on them. This gives effect to Articles 5(1) and 10(2)(a) to protect the affected person's privacy and reputation from illegal deepfakes and defamation.

Relevant laws to combat the abuse of deepfake have to be enacted due to the lacuna in the law in Malaysia. The law has to first recognise the illegality of creating and disseminating deepfakes without the doctored person's consent by inserting the relevant provisions into the CMA. Sanctions should be divided into different levels based on the degree of harm caused by the deepfake. As for civil remedies, the law should require injunctive relief to be granted in order to stop the dissemination and mitigate the damage. The creator of the deepfake in question should be liable if he has knowledge that the deepfake is against the law, such as if it contains elements of obscenities, and/or if it was created without the consent of the doctored person.

Furthermore, relevant policies, procedures, and guidelines should be clearly specified for service providers and social media companies to follow. For example, the internal complaint-handling system, measures and protection against misuse, risk assessment, and mitigation of risks in Articles 20, 23, 34, and 35 of the European Digital Services Act could be taken into reference to be inserted into the CMA. These confer a duty of care on the service providers and online platforms and could hold them liable if they breach the duties.

As the technology of deepfake is increasingly advancing, the technology of voice cloning is also advancing and causing legal problems gradually. The combination of these two technologies increases the risk significantly and can hardly be identified by merely relying on the human eye. It is believed that to solve these problems at their root, it is important to build a solid foundation for privacy law in Malaysia. The only legislation addressing privacy law, the Personal Data Protection Act 2010, is limited to commercial transactions under s 2(1). In light of that, it is suggested that an Act for the protection of personal data against private entities be enacted due to the fact that most abuse of deepfake is caused by individuals rather than commercial entities. Under the proposed Act, “personal data” should be given a wider interpretation that includes facial recognition data and voice data, as these are the main contributors to the abuse of deepfake. Training or processing the data without consent should be prohibited and criminalised as well.

6. Conclusion

In conclusion, the trend of deepfake necessitates Malaysia to move forward due to its insufficiency to address the issue of deepfake. The proposed legal frameworks in other jurisdictions, such as the US and UK, provide some valuable references to which Malaysia can refer. Malaysia not only has to recognise the illegality of deepfake without the depicted person’s consent, but also has to take sufficient preventive measures and be prepared with remedies, such as injunctions, for the victims.

References

- [1] Rodrigo López, J. F. (2022). Tragic realism: How to regulate deepfakes in Colombia? *Latin American Law Review*, (8), 125–145. <https://doi.org/10.29263/lar08.2022.08>
- [2] Vizoso, Á., Vaz-Álvarez, M., & López-García, X. (2021). Fighting deepfakes: Media and internet giants’ converging and diverging strategies against hi-tech misinformation. *Media and Communication*, 9(1), 291–300. <https://doi.org/10.17645/mac.v9i1.3494>
- [3] Albahar, M., Almalki, J., (2019). Deepfakes: Threats and Countermeasures Systematic Review, *Journal of Theoretical and Applied Information Technology*, 97(22), 1817-3195.
- [4] Maras, M.-H., & Alexandrou, A. (2018). Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos. *The International Journal of Evidence & Proof*, 23(3), 255–262. <https://doi.org/10.1177/1365712718807226>
- [5] Westerlund, M. (2019). The emergence of Deepfake Technology: A Review. *Technology Innovation Management Review*, 9(11), 39–52. <https://doi.org/10.22215/timreview/1282>
- [6] Kugler, M. B., & Pace, C. (2021). Deepfake privacy: attitudes and regulation. *Northwestern University Law Review*, 116(3), 611-680.
- [7] Gamage, D., Ghasiya, P., Bonagiri, V., Whiting, M. E., & Sasahara, K. (2022). Are deepfakes concerning? analyzing conversations of deepfakes on reddit and exploring societal implications. CHI Conference on Human Factors in Computing Systems. <https://doi.org/10.1145/3491102.3517446>
- [8] Donegan, M. (2023, March 13). Demand for deepfake pornography is exploding. we aren’t ready for this assault on consent. *The Guardian*.

<https://www.theguardian.com/commentisfree/2023/mar/13/deepfake-pornography-explosion>

- [9] Kugler, M. B., & Pace, C. (2021). Deepfake privacy: attitudes and regulation. *Northwestern University Law Review*, 116(3), 611-680.
- [10] Gerstner, E. (2020). Face/off: “deepfake” face swaps and privacy laws. *Defense Counsel Journal*, 87(1), 1-14.
- [11] Kietzmann, J., Lee, L. W., McCarthy, I. P., & Kietzmann, T. C. (2020). Deepfakes: Trick or treat? *Business Horizons*, 63(2), 135-146. <https://doi.org/10.1016/j.bushor.2019.11.006>.
- [12] Albahar, M., Almalki, J., (2019). Deepfakes: Threats and Countermeasures Systematic Review, *Journal of Theoretical and Applied Information Technology*, 97(22), 1817-3195.
- [13] Lussier, N. (2022). Nonconsensual deepfakes: detecting and regulating this rising threat to privacy. *Idaho Law Review*, 58(2), 353-383.
- [14] Kugler, M. B., & Pace, C. (2021). Deepfake privacy: attitudes and regulation. *Northwestern University Law Review*, 116(3), 611-680.
- [15] Ryan, G., (2022). Cyberflashing and Deepfake Pornography. Paper 01/22. Northern Ireland Assembly. <http://www.niassembly.gov.uk/globalassets/documents/raise/publications/2017-2022/2022/justice/0122.pdf>.
- [16] Mother used deepfake to frame cheerleading rivals’. (2021, March 15). BBC News. <https://www.bbc.com/news/technology-56404038>
- [17] Thombre, M. (2021). Deconstructing Deepfake: Tracking Legal Implications and Challenges. *International Journal of Law Management & Humanities*, 4, 2267-2274.
- [18] Jalil, J. A., (2015) Combating Child Pornography in Digital Era: Is Malaysian Law Adequate to Meet the Digital Challenge? *Pertanika Journal of Social Science & Humanities*, 23(5), 137-152.
- [19] Dato’ Sri Dr Mohamad Salleh bin Ismail & Anor v Nurul Izzah bt Anwar & Anor [2021] 2 MLJ 577
- [20] Charleston & Anor v News Group Newspaper & Anor [1995] 2 All ER 313
- [21] Sembaga Valli K R Ponnusamy v Datuk Bandar Kuala Lumpur & Ors [2017] 1 LNS 500
- [22] H.R. 3230 (116th): Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019. <https://www.govtrack.us/congress/bills/116/hr3230>
- [23] H.R. 2395 (117th): DEEP FAKES Accountability Act. <https://www.govtrack.us/congress/bills/117/hr2395>
- [24] Hsu, T. (2023). As Deepfakes Flourish, Countries Struggle With Response. *The New York Times*. <https://www.nytimes.com/2023/01/22/business/media/deepfake-regulation-difficulty.html>
- [25] Delfino, R. (2022). Deepfakes on trial: A call to expand the trial judge’s gatekeeping role to protect legal proceedings from technological fakery. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4032094>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

