The Closer You Look, The Less You See: Examining the Sufficiency of the Rules on Evidence in Light of the Development of Deepfakes and the Impacts Thereof

Justin Mark C. Chan[★]

★ '21 J.D., Ateneo de Manila University, School of Law. The Author is currently an Associate at Gatmaytan Yap Patacsil Gutierrez & Protacio (C&G Law). He was a Member of the Board of Editors of the *Ateneo Law Journal*. He served as the Associate Lead Editor of the Fourth Issue of the 64th Volume.

This Note is a revised and abridged version of the Author's Juris Doctor thesis in the Ateneo de Manila University School of Law (on file with the Professional Schools Library, Ateneo de Manila University).

Cite as 68 ATENEO L.J. 640 (2023).

I. INTRODUCTION

A. Background of the Study

Back in early 2018, the channel, BuzzFeedVideo, posted a video on YouTube depicting former United States (U.S.) President Barack Obama cursing at the then current U.S. President Donald Trump.[1] The video turned out to be a *fake* as comedian Jordan Peele later on revealed himself as the one uttering every single word.[2] Unfortunately, a lot of people, including the Author, initially thought that the video was real since it looked extremely legitimate and authentic. If people had not been previously informed that the video was a fake, it could have easily passed off as legitimate to the eyes of the majority. The video was produced to spread awareness of "deepfakes" as it urged everyone not to believe everything on the Internet, to rely more on trusted news sources, as well as to "stay woke."[3] This is not an isolated case. There are numerous high-profile examples of malicious use of deepfakes technology. These include an "altered video of House Speaker Nancy Pelosi, [which] was retweeted by [U.S.] President [Donald] Trump as real, that made it look like she was drunkenly stumbling over her words."[4] Facebook Chief Executive Officer Mark Zuckerberg was also a victim of deepfake technology. "Two British artists created a deepfake [video[,] depicting] ... Mark [ ] talking to CBS News about the 'truth of Facebook and who really owns the future.'"[5] These are just a few examples of the horrors to come should this technology be left unchecked.

Countless cases of deepfakes are being produced around the world by ordinary individuals; and applications are readily available to download on the App Store and on the Google Play Store, such as Doublicat and FaceSwap.[6]

---

1. BuzzFeedVideo, Video, *You Won't Believe What Obama Says in This Video!*, Apr. 17, 2018, YOUTUBE, *available at* https://www.youtube.com/watch?v=cQ54GDm1eL0 (last accessed Oct. 31, 2023) [https://perma.cc/T8FT-Y8NQ].

2. *Id.*

3. *Id.*

4. Bernard Marr, *The Best (and Scariest) Examples of AI-Enabled Deepfakes*, FORBES, July 29, 2019, *available at* https://www.forbes.com/sites/bernardmarr/2019/07/22/the-best-and-scariest-examples-of-ai-enabled-deepfakes (last accessed Oct. 31, 2023) [https://perma.cc/J26D-3UFN].

5. *Id.*

6. *See* Ivan Mehta, New Deepfake App Pastes Your Face onto GIFs in Seconds, *available at* https://thenextweb.com/news/new-deepfake-app-pastes-your-face-onto-gifs-in-seconds (last accessed Oct. 31, 2023) [https://perma.cc/CM9X-

These are just a couple of applications which enable the ordinary person to create deepfakes in just a few seconds and in just a few clicks.[7]

What exactly are "deepfakes?" The term came from a Reddit user and was initially used to create fake but hyper realistic pornographic videos of famous celebrities.[8] The Reddit user used "deepfakes" as his username and the word was "simply a portmanteau of 'deep learning' (the particular flavor of AI used for the task) and 'fakes[.]'"[9] This type of technology uses artificial intelligence combined with machine learning to create deepfake videos, images, and audio.[10] Artificial Intelligence and Machine Learning are two different things.[11] On one hand, "Artificial Intelligence can be defined as an area of computer science that has an emphasis on the creation of intelligent machines that can work and react like humans."[12] On the other hand, "Machine Learning can be defined as a subset of AI or can be termed as an application of Artificial Intelligence. In Machine Learning, machines have the ability to learn on their own without being explicitly programmed."[13]

ZQA9] & FaceSwap, Faceswap, *available at* https://faceswap.dev (last accessed Oct. 31, 2023) [https://perma.cc/HW3T-F7ZB].

7. *Id.*

8. Holly Kathleen Hall, *Deepfake Videos: When Seeing Isn't Believing*, 27 CATH. U. J.L. & TECH. 51, 57 (2018) (citing Kristen Dold, Face-Swapping Porn: How a Creepy Internet Trend Could Threaten Democracy, *available at* https://www.rollingstone.com/culture/culture-features/face-swapping-porn-how-a-creepy-internet-trend-could-threaten-democracy-629275 (last accessed Oct. 31, 2023) [https://perma.cc/S56T-ES63]).

9. James Vincent, Why We Need a Better Definition of 'Deepfake', *available at* https://www.theverge.com/2018/5/22/17380306/deepfake-definition-ai-manipulation-fake-news (last accessed Oct. 31, 2023) [https://perma.cc/KV4V-WYNJ].

10. Elizabeth Caldera, *"Reject the Evidence of Your Eyes and Ears": Deepfakes and the Law of Virtual Replicants*, 50 SETON HALL L. REV. 177, 178 (2019).

11. Amyra Sheldon, Artificial Intelligence vs Machine Learning: What's the Difference?, *available at* https://hackernoon.com/artificial-intelligence-vs-machine-learning-whats-the-difference-9e35u30a0 (last accessed Oct. 31, 2023) [https://perma.cc/N582-9MUT].

12. *Id.*

13. *Id.*

This is in contrast with mainstream videos being labeled as "cheapfakes."[14] While deepfakes are created through machine learning, cheapfakes are created by using conventional tools to manipulate audio and visual materials such as Photoshop.[15] Before the dawn of deepfakes, cheapfakes were being used for the same purpose as deepfakes. Fortunately, as these cheapfakes were not as realistic as deepfakes, the problem of malicious use never really became a real threat as opposed to deepfakes.

Ian Sample, a writer for The Guardian, provides a straightforward overview of the process behind creating deepfakes, to wit —

> It takes a few steps to make a [face swap] video. First, you run thousands of face shots of the two people through an AI algorithm called an encoder. The encoder finds and learns similarities between the two faces, and reduces them to their shared common features, compressing the images in the process. A second AI algorithm called a decoder is then taught to recover the faces from the compressed images. Because the faces are different, you train one decoder to recover the first person's face, and another decoder to recover the second person's face. To perform the face swap, you simply feed encoded images into the 'wrong' decoder. For example, a compressed image of person A's face is fed into the decoder trained on person B. The decoder then reconstructs the face of person B with the expressions and orientation of face A. For a convincing video, this has to be done on every frame.[16]

Another method of creating deepfakes is done by using generative adversarial networks (GANs) —

> [By using] ... [GANs], in which two machine learning (ML) models duke it out[,] [o]ne ML model trains on a data set and then creates video forgeries, while the other attempts to detect the forgeries. The forger creates fakes until the other ML model [cannot] detect the forgery. The larger the set of training data, the easier it is for the forger to create a believable deepfake. This is why videos of former presidents and Hollywood celebrities have been frequently

14. Britt Paris & Joan Donovan, Deepfakes and Cheapfakes: The Manipulation of Audio and Visual Evidence, *available at* https://datasociety.net/library/deepfakes-and-cheap-fakes (last accessed Oct. 31, 2023) [https://perma.cc/VW88-UEEM].

15. *Id.*

16. Ian Sample, *What are Deepfakes – and How Can You Spot Them?*, GUARDIAN, Jan. 13, 2020, *available at* https://www.theguardian.com/technology/2020/jan/13/what-are-deepfakes-and-how-can-you-spot-them (last accessed Oct. 31, 2023) [https://perma.cc/XF74-ZMUA].

> used in this early, first generation of deepfakes — [there is] a ton of publicly available video footage to train the forger.[17]

These deepfakes pose numerous problems to society. The most common problem with deepfakes is revenge porn, which "refers to the sharing of explicit or sexual, images or videos, without the consent of the person in the image."[18] This is done easily as anyone with "a computer and access to the [I]nternet can technically produce a 'deepfake' video[.]"[19] Another problem is that deepfakes are being used for political purposes — to sway the electorate's votes in another direction.[20] It may even go so far as to create an international crisis between nuclear-armed states.[21] In 2019, "the CEO of an unnamed [United Kingdom (U.K.)]-based energy firm believed he was on the phone with his boss, the chief executive of firm's the German parent company, when he followed the orders to immediately transfer €220,000 (approx. $243,000) to the bank account of a Hungarian supplier."[22]

---

17. J.M. Porup, How and Why Deepfake Videos Work — and What Is at Risk, *available at* https://www.csoonline.com/article/3293002/deepfake-videos-how-and-why-they-work.html (last accessed Oct. 31, 2023) [https://perma.cc/AHK4-AHTN].

18. Safeline, Revenge Porn – What It Means for the Victim and the Offender, *available at* https://www.safeline.org.uk/revenge-porn-what-it-means-for-the-victim-and-the-offender (last accessed Oct. 31, 2023) [https://perma.cc/AFR5-5B69].

19. Grace Shao, What 'Deepfakes' Are and How They May Be Dangerous, *available at* https://www.cnbc.com/2019/10/14/what-is-deepfake-and-how-it-might-be-dangerous.html (last accessed Oct. 31, 2023) [https://perma.cc/JT8M-FHWD].

20. Katherine Charlet & Danielle Citron, Campaigns Must Prepare for Deepfakes: This Is What Their Plan Should Look Like, *available at* https://carnegieendowment.org/2019/09/05/campaigns-must-prepare-for-deepfakes-this-is-what-their-plan-should-look-like-pub-79792 (last accessed Oct. 31, 2023) [https://perma.cc/4VDT-AN42].

21. Joe Littell, Don't Believe Your Eyes (or Ears): The Weaponization of Artificial Intelligence, Machine Learning, and Deepfakes, *available at* https://warontherocks.com/2019/10/dont-believe-your-eyes-or-ears-the-weaponization-of-artificial-intelligence-machine-learning-and-deepfakes (last accessed Oct. 31, 2023) [https://perma.cc/T5DH-LYUW].

22. Jesse Damiani, A Voice Deepfake Was Used to Scam a CEO Out of $243,000, *available at* https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000/?sh=419434792241 (last accessed Oct. 31, 2023) [https://perma.cc/MW2J-4GLD].

The implications extend even further. As the legal landscape continues to evolve, concerns arise about the harmful impact of deepfakes in court proceedings. This uncertainty could foster public distrust in the Judiciary due to the uncertain treatment of deepfakes.

A major problem is that deepfakes could be admitted as evidence in court.[23] Although there have not been many reports or news of deepfakes actually being introduced in legal proceedings, there have been multiple attempts to admit cheapfakes into evidence.[24] One notable example involves a child custody case in the U.K.[25] where the wife presented a doctored audio recording of the husband to persuade the court that he was violent and aggressive.[26] The husband's lawyer stated that they were fortunate enough to have had the opportunity to study the meta data on the recording to prove that such audio record was a cheapfake.[27] Unfortunately, this method, generally, does not apply to detecting deepfakes.[28] Deepfakes are typically detected by identifying inconsistencies in facial expressions, eye blinking patterns, and unnatural head movements.[29] The reliance on these unconventional and unreliable methods could potentially wreak havoc when these deepfakes are introduced in court as pieces of evidence.

Without adequate safeguards, deepfakes could potentially be admitted as evidence in court. Currently, the 2019 Revised Rules on Evidence allow the introduction of audio, video, and photographs into evidence so long as they are duly authenticated. According to the Rules, documentary evidence consists of "writings, recordings, photographs[,] or any material containing letters, words, sounds, numbers, figures, symbols, or their equivalent, or other modes of written expression offered as proof of their contents. Photographs

23. Matt Reynolds, Courts and Lawyers Struggle with Growing Prevalence of Deepfakes, *available at* https://www.abajournal.com/web/article/courts-and-lawyers-struggle-with-growing-prevalence-of-deepfakes (last accessed Oct. 31, 2023) [https://perma.cc/7E35-AJFX].

24. *Id.*

25. *Id.*

26. *Id.*

27. *Id.*

28. Jonathan Hui, Detect AI-Generated Images & Deepfakes (Part 4), *available at* https://medium.com/@jonathan_hui/detect-ai-generated-images-deepfakes-part-4-5f9ae1dfeb13 (last accessed Oct. 31, 2023) [https://perma.cc/M66M-86DX].

29. *Id.*

include still pictures, drawings, stored images, x-ray films, motion pictures[,] or videos."[30]

Despite the amendments to the Revised Rules on Evidence, it remains unclear whether the Rules on Electronic Evidence have been superseded, as A.M. No. 19-08-15-SC, the Supreme Court resolution which amended the Revised Rules on Evidence, contains no express or implied repealing clause.[31] The Rules on Electronic Evidence, however, does provide a provision which may hint that the Rules on Electronic Evidence should be the governing rule when it comes to electronic evidence despite being covered under the Revised Rules on Evidence.[32]

The authentication process prescribed by the Revised Rules on Evidence for documentary evidence is described as follows —

> Sec. 20. Proof of private documents. – Before any private document offered as authentic is received in evidence, its due execution and authenticity must be proved by any of the following means:
>
> (a) By anyone who saw the document executed or written;
>
> (b) By evidence of the genuineness of the signature or handwriting of the maker; or
>
> (c) By other evidence showing its due execution and authenticity.
>
> Any other private document need only be identified as that which it is claimed to be.[33]

Under the Rules on Electronic Evidence, these rules cover electronic data messages as defined in Rule 2 of the same rules.[34] As stated, "'electronic data message' refers to information generated, sent, received[,] or stored by electronic, optical[,] or similar means."[35] The Rules on Electronic Evidence prescribes the following as its authentication process —

> SECTION 1. Audio, video[,] and similar evidence. – Audio, photographic and video evidence of events, acts[,] or transactions shall be admissible

30. 2019 AMENDMENTS TO THE 1989 RULES ON EVIDENCE, rule 130, § 2.

31. Supreme Court, 2019 Proposed Amendments to the Revised Rules on Evidence, Administrative Matter No. 19-08-15-SC [SC A.M. No. 19-08-15-SC] (Oct. 8, 2019).

32. RULES ON ELECTRONIC EVIDENCE, A.M. No. 01-7-01-SC, rule 1, § 3 (July 17, 2001).

33. 2019 AMENDMENTS TO THE 1989 RULES ON EVIDENCE, rule 132, § 20.

34. RULES ON ELECTRONIC EVIDENCE, rule 1, § 1.

35. *Id.* rule 2, § 1 (g).

> provided is shall be shown, presented[,] or displayed to the court and shall be identified, explained[,] or authenticated by the person who made the recording or by some other person competent to testify on the accuracy thereof.[36]

These provisions are insufficient to address the problem being posed by the introduction of deepfakes into evidence, as well as the problem of authentic evidence being labeled as deepfakes since both rules do not prescribe specific procedures to address these novel problems. This leaves the judicial process with much uncertainty, especially with the assertions that deepfakes will possibly become undetectable in the long run.[37]

Fortunately, tech giant companies like Facebook, Google, Twitter, and Microsoft are making efforts to establish a reliable way to detect these deepfakes.[38] Unfortunately, this does not seem to be an easy task. As companies develop techniques to detect deepfakes, the people who create deepfakes will also develop techniques to by-pass the detection process.[39]

This Note aims to address the gap in the rules and establish a framework to properly authenticate electronic evidence which may be the subject of deepfakes, particularly audio, video, and photographic evidence.

B. Definition of Terms

Considering the technicalities involved in this Note, the following are the terms and their definitions that are used herein:

> (1) Artificial Intelligence – "an area of computer science that has an emphasis on the creation of intelligent machines that can work and react like humans."[40]

---

36. *Id.* rule 11, § 1.

37. Cade Metz, *Internet Companies Prepare to Fight the 'Deepfake' Future*, N.Y. TIMES, Nov. 19, 2019, *available at* https://www.nytimes.com/2019/11/24/technology/tech-companies-deepfakes.html (last accessed Oct. 31, 2023) [https://perma.cc/LUE9-QRLQ].

38. Katie Schoolov, How Facebook, Twitter and Google Are Working to Prevent Deepfakes from Fooling You, *available at* https://www.cnbc.com/2019/09/29/how-facebook-twitter-and-google-work-to-detect-and-prevent-deepfakes.html (last accessed Oct. 31, 2023) [https://perma.cc/V4L2-7QSV].

39. Hany Farid, Imposter Syndrome, *available at* https://web.archive.org/web/20220112203746/https://octavianreport.com/article/hany-farid-fight-threat-deepfakes/2.

40. Sheldon, *supra* note 11.

(2) Cheapfakes – Audio-visual manipulations which use conventional techniques like speeding, slowing, cutting, re-staging, or re-contextualizing footage to change the meaning and interpretation of media.[41]

(3) Deepfakes – "a combination of 'deep learning' and 'fake.' Most often, deepfakes refer to videos, images, audio[,] or text created with artificial intelligence (AI) technologies such as Generative Adversarial Networks (GANs) or Recurrent Neural Networks (RNNs). [The] content synthesis technologies enable media representations of non-existent subjects as well as subjects doing or saying thing [they have] never done or said."[42]

(4) Generative Adversarial Networks – "algorithmic architectures that use two neural networks, pitting one against the other (thus the 'adversarial') in order to generate new, synthetic instances of data that can pass for real data. They are used widely in image generation, video generation[,] and voice generation."[43]

(5) High Technology – "any technology requiring the most sophisticated scientific equipment and advanced engineering techniques, as microelectronics, data processing, genetic engineering, or telecommunications[.]"[44]

(6) Machine Learning – "a subset of AI or can be termed as an application of Artificial Intelligence. In Machine

41. *Id.*

42. DeepTrust Alliance, Deepfake, Cheapfake: The Internet's Next Earthquake (FixFake Symposium Proceedings, Part 1), at 3, *available at* https://static1.squarespace.com/static/5d894b6dcd6a2255c38759fe/t/5e44d9257a6edf3b61208568/1581570371567/DeepTrust+Report+1 (last accessed Oct. 31, 2023) [https://perma.cc/E9QP-XW7N].

43. Pathmind, A Beginner's Guide to Generative Adversarial Networks (GANs), *available at* https://pathmind.com/wiki/generative-adversarial-network-gan (last accessed Oct. 31, 2023) [https://perma.cc/L9Z4-F9B3].

44. Dictionary.com, High Technology, *available at* https://www.dictionary.com/browse/high-technology (last accessed Oct. 31, 2023) [https://perma.cc/HSZ4-6RHT].

> Learning, machines have the ability to learn on their own without being explicitly programmed."[45]

C. Significance of the Study

The courts and the counsels may use the analyses of this Note in determining the different ways to detect deepfakes and to prevent them from being admitted into evidence. The counsels who represent the victims of deepfakes and those whose clients' evidence are accused of being deepfakes will be greatly benefitted with the findings of this Note upon which they may be able to defend the interests of their clients properly.

The public in general will benefit from this Note as the problem of deepfakes is not confined to the legal profession. The findings of the Note will help with the detection of deepfakes as this will be a necessary by-product of the recommendations. This will also spread awareness of the dangers of deepfakes.

The findings of the Author will also be useful as a basis for future regulations with regard to deepfakes, as well as to future technologies which share similar features and specifications with deepfakes. Furthermore, this Note may prove to be beneficial to legislators who plan on enacting statutes to address the problem similar to how the legislators of foreign other jurisdictions have commenced.

II. DEEPFAKE LITERATURE

For this portion of the Note, the Author discusses the different approaches proposed by foreign authors. Unfortunately, due to the fact that the emergence of deepfakes is fairly recent, the amount of literature discussing its potential problems, particularly on the areas of litigation and evidence, are quite limited. Nevertheless, the Author still discusses the relevant related literature, both in favor of and against his conclusion.

A. Foreign Literature on Deepfakes

In recent years, cheapfakes have grown rampant around the world.[46] Although these cheapfakes are not nearly as convincing as deepfakes, they still have the

45. Sheldon, *supra* note 11.

46. Donovan & Paris, *supra* note 14.

potential to fool the public.[47] The rise of these high technologies, including deepfakes, could impact the way people view their day-to-day lives.[48]

1. Sufficiency of Existing Rules

Deepfakes, particularly in the context of evidence and in the field of litigation, can be countered in different ways.[49] Contrary to the opinion of the Author, some authors argue that the traditional rules used by courts around the world are sufficient to combat the threat posed by deepfakes.[50] The U.S.' "current legal system offers powerful tools for identifying and challenging potential deepfakes offered by an opposing party."[51] The use of traditional litigation tools, such as: discovery, taking advantage of the rules of authentication available to the litigants, cross-examination, and expert testimony can prove to be useful to reveal the truth — the main goal of litigation.[52] Although some of these tools are more focused on the courts of law located in the U.S., these may be used analogously in the Philippine context. With regard to the use of targeted discovery, it is opined that the initial attempt to sniff out deepfakes is through discovery.[53] Those who attempt to admit these suspicious evidence that may be compromised by deepfakes could be questioned as to:

(1) the time, place, and date the recording was made;

(2) the name and address of any individual depicted in or present at the time of recording;

(3) the name and address of any individual under whose direction and upon whose behalf the recording was created;

(4) the name and address of any other individual involved with the creation of the recording;

(5) the steps undertaken by the identified individuals to create the recording; and

47. *Id.*

48. Gary E. Marchant, *Emerging Technologies and the Courts*, 55 CT. REV. 146, 146 (2019).

49. Kathryn Lehman, et al., 5 Ways to Confront Potential Deepfake Evidence in Court, *available at* https://www.law360.com/articles/1181306/5-ways-to-confront-potential-deepfake-evidence-in-court (last accessed Oct. 31, 2023) [https://perma.cc/R2GS-68LR].

50. *Id.*

51. *Id.*

52. *Id.*

53. *Id.*

> (6) the name and address of any individual who has had possession or control of the recording (either the original or a copy) since it was created.[54]

Depending on how these questions are answered, it could lead to follow-up questions with regard to "production of metadata, interrogatories[,] or depositions of custodians or witnesses to the events in the video."[55]

Another approach is to apply the current rules on authentication.[56] Although the journals and articles on this topic are tailored towards the Federal Rules of Evidence[57] of the U.S., these can be used analogously in the Philippine context, except for some particular rules which are not found in the Rules of Court, such as the silent witness theory and the concept of self-authenticating evidence.[58] Under the Federal Rules of Evidence —

> Federal Rule 901 (a) states a general rule that the proponent of evidence 'must produce evidence sufficient to support a finding that the item is what the proponent claims it is.' Rule 902 provides that certain items of evidence are 'self-authenticating; they require no extrinsic evidence of authenticity in order to be admitted.'
>
> Two recent additions to Rule 902 address electronically stored information. Rule 902 (13) allows authentication of a record 'generated by an electronic process or system that produces an accurate result' if 'shown by the certification of a qualified person' that complies with certain requirements. Rule 902 (14) allows authentication of data 'copied from an electronic device, storage medium, or file, if authenticated by a process of digital identification, as shown by a certification of a qualified person' if authenticated 'by a process of digital identification, as shown by a certification of a qualified person' that meets those same requirements.
>
> In addition, some courts have allowed video or audio recordings to be authenticated through the 'silent witness' theory — '[u]nder this approach, the foundation focuses on the automatic operation of the recording device and does not consider a witness's observations of the recorded events because the recording speaks for itself.'[59]

---

54. *Id.*

55. Lehman, et al., *supra* note 49.

56. *Id.*

57. FEDERAL RULES OF EVIDENCE (U.S.).

58. Lehman, et al., *supra* note 49.

59. *Id.* (citing FEDERAL RULES OF EVIDENCE, art. IX, rules 901 (a) & 902 & Jonathan Mraunac, The Future of Authenticating Audio and Video Evidence, *available at* https://www.law360.com/articles/1067033/the-future-of-authenticating-

Therefore, an experienced litigator should have the tools necessary to combat the threat of deepfakes in a court of law.[60] The other way to attack deepfakes is through cross-examination.[61] The argument concerning cross-examination is stated to apprise the reader of the different options available to a litigator. However, it falls outside of the coverage of this Note, which is focused specifically on the admissibility of electronic evidence susceptible to deepfakes, rather than on the probative value of such evidence.

Another way to combat the threat of deepfakes in a court of law is through expert testimony.[62] It is also important to note that, because the technology behind deepfakes is still in its early stages, there is a lack of experts who may be able to determine deepfakes from legitimate electronic evidence.[63]

2. Insufficiency of Existing Rules

In contrast to the opinion presented earlier, several authorities have addressed the threat of deepfakes in a court of law and have recommended proactive measures to prevent judicial proceedings from becoming a sham.[64] Commonly, these deepfakes are being exploited by individuals attempting to use blackmail as a means of extorting resources.[65] Sometimes it is also used to destroy the reputations of public officials.[66] The potential harm posed by deepfakes could be limitless, particularly given the significant disruption they are already causing, despite the technology still being in its early stages.[67] It could potentially be the "greatest evidentiary threat to the courts[.]"[68] One author likened the difficulty of detecting deepfakes to a science fiction movie

audio-and-video-evidence (last accessed Oct. 31, 2023) [https://perma.cc/99TL-728R]).

60. Lehman, *supra* note 49.

61. *Id.*

62. *Id.*

63. *Id.*

64. *See* Hall, *supra* note 8; Caldera, *supra* note 10; Neil Fulton, *Fake News on Trial: The Jury Trial as a Guard Against Societal Entropy*, 52 TEX. TECH. L. REV. 743 (2020); Marchant, *supra* note 48; Jeff Ward, *10 Things Judges Should Know About AI*, 103 JUDICATURE 12 (2019); & Alexa Koenig, *Half the Truth Is Often a Great Lie: Deep Fakes, Open Source Information, and International Criminal Law*, 113 AJIL UNBOUND 250 (2019).

65. Hall, *supra* note 8, at 52.

66. *Id.*

67. Caldera, *supra* note 10, at 178.

68. Marchant, *supra* note 48, at 150-51.

— Blade Runner[69] — where one is to detect "whether a being is a human or a replicant [the 'Voight-Kampff' test is used.]"[70] One of the main problems in the film is the "diminishing boundary between man and machine, [which replicates the problem] surround[ing] deepfakes."[71] Although there are numerous tech companies, as well as the U.S. government attempting to develop tools to detect deepfakes, the effectivity of these tools could become questionable because deepfakes could become undetectable in the future.[72] Moreover, the mere existence of deepfakes could undermine video as evidence, thereby increasing the effectiveness and impact of deepfakes.[73] "If video cannot be trusted, having a corroborating video to debunk a deepfake would no longer be sufficient; the risk of the supposedly corroborating evidence also being a deepfake may be too high if there is no ability to determine if a video has been doctored."[74]

Generally, video evidence is considered as a "reliable source of information[.]"[75] However, deepfakes would change the reliability of not just video evidence, but also all kinds of evidence which are susceptible to deepfakes.[76] Deepfakes making their way to courts is certainly a probability as opposed to a possibility.[77] Ethical considerations could stop lawyers from presenting doctored evidence in court; however, lawyers may not always be entirely certain that the pieces of evidence provided by their clients are not doctored or are not deepfakes.[78]

a. What Is Being Done?

The big problem right now is that even experts could be tricked by these deepfakes.[79] Unfortunately, experts would even take a lot of time to determine

69. Caldera, *supra* note 10, at 180 (citing BLADE RUNNER (The Ladd Company 1982)).

70. *Id.*

71. *Id.*

72. *Id.*

73. Caldera, *supra* note 10, at 187-88.

74. *Id.* at 188.

75. *Id.*

76. *Id.*

77. *See* David Dorfman, *Decoding Deepfakes: How Do Lawyers Adapt When Seeing Isn't Always Believing?*, 80 OR. ST. BAR BULL. 18, 22 (2020).

78. *Id.*

79. Marchant, *supra* note 48, at 151.

whether a video is a deepfake or not.[80] Even worse is that multiple experts believe that deepfakes will become undetectable within a year or maybe even less.[81]

> If we can no longer believe what we see, the privileged position that photographs and videos have had in our litigation system will disappear. Not only will we not know that a fake video or photo has been fabricated, but it will be easy to claim that a real video or photo is fake.[82]

Hany Farid, a computer scientist in University of California, Berkeley, who is "considered by many to be the 'father of digital forensics,'"[83] stated "that today[,] there is no operationalized technique for reliably detecting deepfakes. Part of that is because deepfakes are a relatively new phenomenon and we and other people are in the early stages of developing those techniques."[84] Farid predicts that tools for detecting deepfakes will start to get developed over the course of the year, but this continues to be a "cat-and-mouse game" as detection mechanisms become more reliable, deepfakers will also adapt their methods to circumvent and evade the mechanisms designed to keep them at bay.[85]

Farid mentions two ways to detect deepfakes.[86] One is a "soft biometric." Farid says that

> '[t]he basic idea is this,' ... 'When somebody is speaking, there is a correlation between what they say and how they say it. For example, when I frown and pinch my brow, something is upsetting to me. If I say something funny, I tend to smile and maybe lift my head up a little bit. How our faces move, how our head moves[;] we are finding to be tightly correlated to what we are saying.'[87]

The other technique is called "controlled capture" technology.[88] Simply, authentication software could be pre-installed in cameras to facilitate the

80. *Id.*

81. *Id.* & Ward, *supra* note 64, at 17.

82. Marchant, *supra* note 48, at 151.

83. Gradient Flow, Issue #10: ML in Finance, Disinformation, AI Superpowers, *available at* https://gradientflow.com/issue-10-ml-in-finance-disinformation-ai-superpowers (last accessed Oct. 31, 2023) [https://perma.cc/9GH8-UDPH].

84. Farid, *supra* note 39.

85. *Id.*

86. *Id.*

87. *Id.*

88. *Id.*

verification of videos and photos. Given the susceptibility of such evidence to deepfakes, these files could include a digital watermark to confirm their legitimacy.[89] This technology could be adopted by leading smartphone companies around the world to give their customers the option to record securely or to not record securely.[90] Fortunately, there are already existing businesses which produce this type of technology commercially.[91]

The U.S. Congress is way ahead of the game by introducing legislation aimed at combating deepfakes. Although the DEEP FAKES Accountability Act[92] has not yet been passed into law, it is a step in the right direction to combat the harmful effects of deepfakes.[93] This law focuses more on punishing those who do not disclose that the material is a deepfake, as opposed to laying down rules to keep these deepfakes away from the courts.[94] In the U.S., their Congress is able to pass a law which prescribes the rules of procedure to be followed in court such as the Federal Rules of Evidence.[95]

In the Philippines, there have been manifestations by the Congress of their intent to suppress the proliferation of deepfakes. On one hand, in the Senate, there is a Senate resolution,[96] which was introduced by Senator Ralph G. Recto, pertaining to deepfakes. The resolution was introduced to conduct an inquiry on deepfakes to strengthen Philippines laws, particularly the laws on privacy.[97] The Senate recognized that the advancements in modern

89. *Id.*

90. Farid, *supra* note 39.

91. *Id.*

92. Defending Each and Every Person from False Appearances by Keeping Exploitation to Accountability Act of 2019 or the DEEP FAKES Accountability Act, H.R. No. 3230, 116th Cong. (2019) (U.S.).

93. *Id.*

94. *Id.* § 2.

95. Disini Law, Facebook Live, Aug. 6, 2020: 3:57 p.m., FACEBOOK, *available at* https://www.facebook.com/watch/live/?ref=watch_permalink&v=3100134093433131 (last accessed Oct. 31, 2023) [https://perma.cc/M8MD-TQJK].

96. Resolution Directing the Appropriate Senate Committee to Conduct an Inquiry, in Aid of Legislation, on the Proliferation of Artificial Intelligence-Synthesized Audiovisual Materials, Otherwise Known as Deepfakes, with the End in View of Strengthening Government Mechanisms to Implement Cybercrime and Data Privacy Laws, Safeguarding the Privacy, Information and Identity of All Filipinos and Protecting the Integrity of Philippine Social, Political, Economic and Financial Institutions, S. Res. No. 188, 18th Cong., 1st Reg. Sess. (2019).

97. *Id.*

technology, combined with the reach of mainstream social media platforms, could lead to severe consequences, particularly with the rise of deepfakes in areas like targeted harassment and manipulated pornography.[98]

On the other hand, the House of Representatives created a bill,[99] sponsored by Hon. Rozzano Rufino B. Biazon, which declares "the creation and disclosure of any deepfake material or a materially deceptive audio or video recording" as unlawful.[100] In contrast to the U.S. Congress' Bill, this Philippine House of Representatives' Bill on deepfakes is much shorter. The Philippine Bill only provides for the unlawful act, its exemptions, and the penalties.[101] The U.S. Bill on deepfakes is more comprehensive as it even provides for penalties on altering the disclosures — a feature certainly missing from the proposed Philippine Bill on deepfakes.[102] The U.S.' Bill also provides for some procedural rules.[103]

These are just some of the tools available to combat the threat of deepfakes. Unfortunately, these tools may be inadequate to prevent deepfakes from entering the dockets of the courts, especially if these are not coupled with rules tailor-fit towards the prevention of such spurious pieces of evidence from being admitted in court.

III. PROCEDURAL DUE PROCESS

A. Supreme Court

The third branch of government — the Judiciary — is lodged in "[the] Supreme Court and in such lower courts as may be established by law."[104] Judicial power is the "duty of the courts ... to settle actual controversies involving rights which are legally demandable and enforceable, and to determine whether or not there has been a grave abuse of discretion amounting to lack or excess of jurisdiction on the part of any branch or

98. *Id.* whereas cl. paras. 4-6, 9, & 12.

99. An Act Declaring as Unlawful the Creation and Disclosure of any Deepfake Material or a Materially Deceptive Audio or Video Recording of an Individual Without his or her Consent, H.B. No. 5406, 18th Cong., 1st Reg. Sess. (2019).

100. *Id.* explan. n.

101. *Id.* §§ 3, 4, & 5.

102. DEEP FAKES Accountability Act, § 2.

103. *Id.*

104. PHIL CONST. art. VIII, § 1.

instrumentality of the Government."[105] The 1987 Constitution of the Philippines also gives the Court, the following powers:

> (1) Exercise original jurisdiction over cases affecting ambassadors, other public ministers[,] and consuls, and over petitions for certiorari, prohibition, mandamus, quo warranto, and habeas corpus.
>
> (2) Review, revise, reverse, modify, or affirm on appeal or certiorari, as the law or the Rules of Court may provide, final judgments and orders of lower courts in:
>
> (a) All cases in which the constitutionality or validity of any treaty, international or executive agreement, law, presidential decree, proclamation, order, instruction, ordinance, or regulation is in question.
>
> (b) All cases involving the legality of any tax, impost, assessment, or toll, or any penalty imposed in relation thereto.
>
> (c) All cases in which the jurisdiction of any lower court is in issue.
>
> (d) All criminal cases in which the penalty imposed is *reclusion perpetua* or higher.
>
> (e) All cases in which only an error or question of law is involved.
>
> (3) Assign temporarily judges of lower courts to other stations as public interest may require. Such temporary assignment shall not exceed six months without the consent of the judge concerned.
>
> (4) Order a change of venue or place of trial to avoid a miscarriage of justice.
>
> (5) Promulgate rules concerning the protection and enforcement of constitutional rights, pleading, practice, and procedure in all courts, the admission to the practice of law, the Integrated Bar, and legal assistance to the underprivileged. Such rules shall provide a simplified and inexpensive procedure for the speedy disposition of cases, shall be uniform for all courts of the same grade, and shall not diminish, increase, or modify substantive rights. Rules of procedure of special courts and quasi-judicial bodies shall remain effective unless disapproved by the Supreme Court.
>
> (6) Appoint all officials and employees of the Judiciary in accordance with the Civil Service Law.[106]

Pursuant to this power, the Court has enacted various rules and regulations regarding the procedure to be observed in all courts to afford the

105. PHIL CONST. art. VIII, § 1.

106. PHIL CONST. art. VIII, § 5.

people a simplified and inexpensive procedure to enforce their rights.[107] These rules of procedures are ultimately contained in the Rules of Court, as well as in issuances by the Court of specific rules to address specific issues usually in the form of an Administrative Matter.[108]

B. Brief History of Electronic Evidence

Before the recent amendment of the Revised Rules on Evidence in 2019, the 1989 Revised Rules on Evidence did not cover electronic forms of evidence.[109] The 1989 Revised Rules on Evidence defined documentary evidence as "evidence [which] consists of writings or any material containing letters, words, numbers, figures, symbols[,] or other modes of written expressions offered as proof of their contents."[110] This left a gap in the law with regard to the introduction of electronic evidence, as the 1989 Revised Rules on Evidence did not specifically address these types of evidence in the Rules of Court.[111] In 2008, however, the Court remedied this gap in the Rules of Court when it enacted A.M. No. 01-7-01-SC, otherwise known as the Rules on Electronic Evidence.[112] Under the Rules on Electronic Evidence, an electronic data message is defined as "information generated, sent, received[,] or stored by electronic, optical[,] or similar means."[113] The same Rules also defines electronic evidence as

> information or the representation of information, data, figures, symbols[,] or other modes of written expression, described or however represented, by which a right is established or an obligation extinguished, or by which a fact may be proved and affirmed, which is received, recorded, transmitted, stored processed, retrieved[,] or produced electronically. It includes digitally signed documents and any print-out or output, readable by sight or other means, which accurately reflects the electronic data message or electronic document. For purposes of these Rules, the term 'electronic document' may be used interchangeably with 'electronic data message.'[114]

---

107. *See, e.g.*, RULES OF COURT; RULES ON ELECTRONIC EVIDENCE; JUDICIAL AFFIDAVIT RULE, A.M. No. 12-8-8-SC (Sep. 4, 2012); & THE RULE OF PROCEDURE FOR SMALL CLAIM CASES A.M. No. 08-8-7-SC (Oct. 1, 2008).

108. *Id.*

109. *See* 1989 REVISED RULES ON EVIDENCE, rule 130, § 2 (superseded in 2019).

110. *Id.*

111. *See* 1989 REVISED RULES ON EVIDENCE (superseded in 2008).

112. RULES ON ELECTRONIC EVIDENCE.

113. *Id.* rule 2, § 1 (g).

114. *Id.* § 1 (h).

In addition, the Rules on Electronic Evidence also introduced a way to authenticate different forms of electronic evidence.[115] The modes of authentication particularly relevant to this Note are those established for audio, photographic, and video evidence.[116] During this time and well before the Revised Rules on Evidence was amended in 2019, the manner of authentication merely required that the pieces of evidence be shown or presented to the court and identified by the person who made the recording or by some other person competent to testify on such pieces of electronic evidence.[117] Under the Rules on Electronic Evidence, the issuance specifically states that the rules apply to "all civil actions and proceedings, as well as quasi-judicial and administrative cases."[118] Due to this express statement of the scope of the rule, it is reasonable to assume that the Rules on Electronic Evidence do not apply to criminal actions and proceedings.

The case of *People v. Enojas*, however, illustrates otherwise. In this case,[119] a criminal action for murder was filed against Noel Enojas y Hingpit, along with several others.[120] Here, the police were able to apprehend Enojas, along with the other accused, with the assistance of Enojas' mobile phone, which he left behind in his taxi.[121] One of the issues in the case was whether or not the text messages contained in the mobile phone of the accused was admissible in evidence, considering that the Rules on Electronic Evidence contains an express provision regarding its scope and it does not include its non-applicability to criminal actions and proceedings.[122] The Court ultimately held that the Regional Trial Court did not err when it admitted the text messages in evidence as this was in conformity with the with a previous Supreme Court Resolution[123] expanding the coverage of the Rules on Electronic Evidence to also cover criminal actions and proceedings.[124]

---

115. *Id.* rule 3, § 2; rule 5, § 2; rule 6, § 2; & rule 11, §§ 1-2.

116. *Id.* rule 11, § 1.

117. *Id.*

118. RULES ON ELECTRONIC EVIDENCE, rule 1, § 2.

119. People vs. Enojas, G.R. No. 204894, 718 SCRA 313 (2014).

120. *Id.* at 314.

121. *Id.* at 315.

122. *Id.* at 319.

123. EXPANSION OF THE COVERAGE OF THE RULES ON ELECTRONIC EVIDENCE, A.M. No. 01-7-01-SC (Sep. 24, 2002).

124. *Enojas*, 718 SCRA at 319.

With the 2019 amendment of the Revised Rules on Electronic Evidence, the definitions of the types of electronic evidence have been incorporated in the term documentary evidence.[125] The amendment included "recordings, photographs[,] ... words, sounds, numbers, ... or their equivalent" to the definition of documentary evidence.[126]

As a result, electronic evidence is now governed by both the 2019 Revised Rules on Evidence and the Rules on Electronic Evidence. Unfortunately, as of writing of this Note, there is no jurisprudence addressing which Rules would govern electronic evidence, especially since both Rules provide for ways to authenticate these pieces of evidence.[127]

C. Revised Rules on Evidence vs. Rules on Electronic Evidence

1. Implied Repeal of an Old Law

This problem may be resolved by using the rules on statutory construction.[128] It is settled doctrine in Philippine courts that implied repeals are frowned upon.[129] Consequently, the mere fact that a later law also covers the same subject matter as an older law, is not itself sufficient to warrant the repeal of the former law when both statues may be harmonized.[130] When two statutes cover the same subject matter, both statutes should first be harmonized, as long as both statutes are not absolutely irreconcilable.[131] Therefore, if both statues are able, by any reasonable construction, to stand together, then both statutes will be sustained.[132]

125. REVISED RULES ON EVIDENCE, rule 130, § 2.

126. *Id.*

127. REVISED RULES ON EVIDENCE, rule 132, § 20 & RULES ON ELECTRONIC EVIDENCE, rule 11, § 1.

128. A few rules on statutory construction are summarized by the Colorado General Assembly. Colorado General Assembly, Commonly Applied Rules of Statutory Construction, *available at* https://leg.colorado.gov/agencies/office-legislative-legal-services/commonly-applied-rules-statutory-construction (last accessed Oct. 31, 2023) [https://perma.cc/K2WE-T6W6].

129. Lichauco & Co. v. Apostol, 44 Phil. 138, 147 (1922).

130. Valera v. Tuason Jr., 80 Phil. 823, 827 (1948).

131. *Lichauco & Co.*, 44 Phil. at 147.

132. *Id.*

2. Implied Repeal of a Special Law

As established above, when both statutes are irreconcilable, the later law should repeal the former law. The rules on statutory construction, however, also provide that "a special law is not regarded as having been amended or repealed by a [later] general law, unless the intent to repeal or alter is manifest."[133] This holds true, regardless if the matters covered by the general statute are broad enough to cover matters specifically covered in the special statutes.[134]

3. Application of the Rules on Statutory Construction

It cannot be said that both the 2019 Revised Rules on Evidence and the Rules on Electronic Evidence are irreconcilable. The 2019 Revised Rules on Evidence did not include a provision which expressly repeals the Rules on Electronic Evidence.[135] Therefore, both Rules must be harmonized. There are multiple interpretations possible to provide both Rules effect. It can be said that compliance with both authentication rules is necessary for the admissibility of electronic evidence.[136] Furthermore, while the Revised Rules on Evidence define documentary evidence to include electronic evidence,[137] the authentication process of the Revised Rules on Evidence is more geared towards documentary evidence which are not electronic. Thus, documentary evidence which are electronic should be authenticated by using the authentication procedure as provided in the Rules on Electronic Evidence. In addition to these complications, there are different rules which modify the rules on authentication under particular circumstances such as the Judicial Affidavit Rule.[138] It is also important to note that while the Rules of Court generally only applies to judicial proceedings, the Rules on Electronic Evidence applies to judicial proceedings, quasi-judicial proceedings, and administrative proceedings.[139]

As this issue remains unsettled, the Author can only speculate as to the intentions of the Supreme Court when they incorporated the definitions of electronic evidence into the 2019 Revised Rules on Evidence.

---

133. *Valera*, 80 Phil. at 827.

134. *Id.*

135. REVISED RULES ON EVIDENCE.

136. *Id.*

137. *Id.*

138. JUDICIAL AFFIDAVIT RULE, A.M. No. 12-8-8-SC, § 9 (a) (Jan. 1, 2013).

139. RULES ON ELECTRONIC EVIDENCE, rule 1, § 2.

4. Opinions of Justice Singh and Atty. Jose Jesus Disini

In a webinar[140] hosted by Disini Law on the video conferencing platform, Zoom, and on Facebook, Atty. Jose Jesus ("JJ") Disini and Justice Maria Filomena Singh shared their thoughts on the interplay between the 2019 Revised Rules on Evidence and the old Rules on Electronic Evidence.[141] The Webinar began with a discussion on how the Rules on Electronic Evidence is treated in the 2019 Revised Rules on Evidence.[142] Its goal was to answer the question — "[a]re electronic documents governed by the [Revised] Rules on Evidence?"[143] Prior to the recent amendment of the Revised Rules on Evidence, there used to be a clear delineation between the Revised Rules on Evidence and the Rules on Electronic Evidence.[144] Before the promulgation of the Rules on Electronic Evidence, the sub-committee on electronic commerce, which Atty. JJ Disini was a member of, proposed an option to incorporate the then proposed Rules on Electronic Evidence in the Revised Rules on Evidence, instead of issuing a separate rule of procedure which would not form part of the traditional Rules of Court.[145] Unfortunately, the Court deemed it more advantageous to separate the Rules on Electronic Evidence from the Revised Rules on Evidence.[146] The Court did, however, indicate that it intends to incorporate the Rules on Electronic Evidence into the Revised Rules on Evidence at a future date.[147] Sad to say that the Court never actually incorporated the Rules on Electronic Evidence to the Revised Rules on Evidence until the 2019 amendment.[148]

The primary difference between the rules lies in the types of evidence they address and the authentication processes they require.[149] The 2019 Revised Rules on Evidence actually incorporated electronic evidence in the definition of documents when it included a definition of photographs; however, it failed to clearly state which rules would govern the authentication

140. Disini Law, *supra* note 95.

141. *Id.*

142. *Id.*

143. *Id.*

144. *Id.*

145. *Id.*

146. Disini Law, *supra* note 95.

147. *Id.*

148. *Id.*

149. *Id.*

process of such documents.[150] To address the question why the Court placed references to electronic evidence in the 2019 Revised Rules on Evidence despite its coverage under a separate rule, Atty. JJ Disini presents the following possible interpretations:

> (a) These references in the [2019 Revised Rules on Evidence] to electronic [documents] are drafting errors.
>
> (b) These references were necessary because of partial incorporation of electronic documents in the [2019 Revised Rules on Evidence].
>
> (c) These references [were] meant to be ignored and [electronic documents] are treated in the [Rules on Electronic Evidence while,] non-[electronic documents] are treated in [the 2019 Revised Rules on Evidence].[151]

Atty. JJ Disini prefers to use the third possible interpretation to keep the delineation between electronic evidence and non-electronic evidence.[152] He also recommends that the Court should incorporate all the Rules on Electronic Evidence in the Revised Rules on Evidence.[153] For Justice Singh, the Court incorporated electronic evidence in the 2019 Revised Rules on Evidence to fulfill its old promise to incorporate the Rules on Electronic Evidence in the Rules of Court.[154] Also, it is expressly provided in the Rules on Electronic Evidence that any reference in the Revised Rules on Evidence to documents should be understood to encompass electronic documents as well.[155] Justice Singh is of the opinion that the Court wanted to harmonize both rules.[156] This raises the question, however, of whether the Rules on Electronic Evidence should still be used since the 2019 Revised Rules on Evidence incorporated some of the forms of electronic evidence in its description.[157] As previously mentioned, implied repeals are frowned upon by doctrines on statutory construction as both rules should first be harmonized.[158] Justice Singh believes that the Rules on Electronic Evidence has not been superseded by the 2019 Revised Rules on Evidence. In fact, the Rules on Electronic Evidence has some provisions, which are important in assessing the

150. *Id.*

151. *Id.*

152. Disini Law, *supra* note 95.

153. *See* REVISED RULES ON EVIDENCE.

154. *Id.*

155. *Id.*

156. *Id.*

157. *Id.*

158. *Lichauco & Co.*, 44 Phil. at 147.

admissibility of electronic evidence that are nowhere to be found in the 2019 Revised Rules on Evidence.[159] Thus, she concludes that the provisions in the Rules on Electronic Evidence, which are not contrary to the 2019 Revised Rules on Evidence, are still good law.[160]

The Webinar was conducted to address the question of electronic documents; however, there is no reason that these opinions cannot be applied to other types of electronic pieces of evidence as well. The Author is of the opinion that there is greater reason to apply the Rules on Electronic Evidence on electronic evidence such as audio, video, and photo as the authentication process provided by these rules are more tailor-fit towards these pieces of evidence. Safe to say, the initial case involving electronic evidence that the Court will rule upon will be a much-awaited case to put to rest the question on which set of rules will apply.

IV. JURISPRUDENCE

This portion of the Note examines the different jurisprudence with regard to the authentication of electronic evidence. This Chapter aims to discuss the different methods of authentication depending on the type of evidence being offered in court. This Chapter also explores the different methods being used by foreign courts to draw inspiration on how they handle the threat of fake electronic evidence. Unfortunately, due to the novelty of deepfakes, the amount of jurisprudence which pertains to deepfakes are little to none. Hence, the Note examines analogous cases instead.

A. Philippine Jurisprudence on Electronic Evidence

Philippine jurisprudence regarding authentication and admission of electronic evidence is not as mature compared to other more contentious topics of the law, such as controversial areas of Political Law. Upon a perusal of the different cases involving electronic evidence, most of the cases involve the admissibility of ephemeral evidence, such as text messages and electronic mails.

In the case of *Ang v. Court of Appeals*,[161] the Court denied the claim of Rustan when he claimed that the photo constituted electronic evidence and needed to be authenticated according to the Rules on Electronic Evidence.[162] Although, technically the photo of Irish superimposed on naked woman could

159. Disini Law, *supra* note 95.

160. *Id.*

161. Ang v. Court of Appeals, G.R. No. 182835, 618 SCRA 592 (2010).

162. *Id.* at 604.

constitute as a form of electronic evidence according to the Rules, the Court denied to rule on the issue based on a procedural matter.[163] The problem was that Rustan objected too late, and was deemed to have waived the ground for objection.[164] The Court also mentioned that since the Rules on Electronic Evidence particularly states that the Rules only apply to "civil actions, quasi-judicial proceedings, and administrative proceedings[,]" then these should not apply to criminal cases.[165] As previously mentioned, the case of *Enojas* is the controlling doctrine on this matter. Despite the fact that the Rules themselves specifically state that they only apply to "civil actions and proceedings, as well as quasi-judicial and administrative cases[,]"[166] the Court held that the Rules on Electronic Evidence do, in fact, apply to criminal cases which consequently expounded the coverage of the Rules on Electronic Evidence.[167]

In the case of *Nuez v. Cruz-Apao*,[168] the Court admitted the text messages as ephemeral electronic evidence. This case was "an administrative case for [d]ishonesty and [g]rave [m]isconduct against Elvira Cruz-Apao [ ], [an] Executive Assistant II of the Acting Division Clerk of Court of the Fifteenth [ ] Division [of the] Court of Appeals [ ]."[169] Allegedly, respondent told the complainant that in exchange for one million pesos, his pending case would result in a speedy and favorable decision.[170] When the complainant attempted to negotiate for a lower amount, he was instead scolded and was told that they were not in a wet market.[171] The text messages, accompanied by the complainant's testimony, were introduced in evidence to prove that the respondent tried to extort one million pesos from the complainant to obtain a favorable judgement from his pending case in the Court of Appeals.[172] The Court quoted the Rules on Electronic Evidence in defining ephemeral electronic evidence, to wit —"'[e]phemeral electronic communication' refers to telephone conversations, text messages ... and other electronic forms of communication the evidence of which is not recorded or retained."[173]

163. *Id.* at 595-96.

164. *Id.* at 604.

165. *Id.*

166. RULES ON ELECTRONIC EVIDENCE, rule 1, § 2.

167. *Enojas*, 718 SCRA at 319.

168. Nuez v. Cruz-Apao, A.M. No. CA-05-18-P, 455 SCRA 288 (2005).

169. *Id.* at 290-91.

170. *Id.* at 293.

171. *Id.*

172. *Id.* at 299.

173. *Id.* (citing RULES ON ELECTRONIC EVIDENCE, rule 2, § 1 (k)).

According to the Rules on Electronic Evidence, "[e]phemeral electronic communications shall be proven by the testimony of a person who was a party to the same or who has personal knowledge thereof ... ."[174] Unfortunately for the respondent, the authentication requirements, as prescribed by the Rules on Electronic Evidence, were complied with.[175] It was actually the respondent herself who admitted that the cellphone number which sent the text messages to the complainant belonged to her.[176] To make matters worse, the respondent, together with her lawyer, confirmed that the exchange of text messages between her and the complainant were accurate.[177] Thus, the guilt of the respondent was proven through the text messages which were admitted as ephemeral electronic evidence.[178] This resulted in the respondent being dismissed from government service and she consequently lost her retirement benefits.[179] Therefore, with regard to ephemeral electronic communications such as text messages, they must "be proven by the testimony of a person who was a party to the communications or has personal knowledge thereof."[180]

The case of *INC Shipmanagement, Inc. v. Camporedondo*[181] briefly discussed the authentication of electronic mails. This case involved a review on *certiorari* of the decision promulgated by the Court of Appeals on a labor case.[182] The respondent was hired to be the chief cook of the M/V Fortunia.[183] The respondent had multiple inquiries to the captain of the ship with regard to the budget as well as the quality of the supplies he was given.[184] This angered the captain, hence, the respondent was admonished on a daily basis.[185] Sometime in September 2007, the respondent was given a ticket back to the Philippines

174. *Cruz-Apao*, 455 SCRA at 299 (citing RULES ON ELECTRONIC EVIDENCE, rule 11, § 2).

175. *Cruz-Apao*, 455 SCRA at 299-300.

176. *Id.* at 300.

177. *Id.*

178. *Id.*

179. *Id.* at 307.

180. Bartolome v. Maranan, G.R. No. P-11-2979, 740 SCRA 491, 501 (2014) (citing RULES ON ELECTRONIC EVIDENCE, rule 11, § 2).

181. INC Shipmanagement, Inc. v. Camporedondo, G.R. No. 199931, 769 SCRA 295 (2015).

182. *Id.* at 297-98.

183. *Id.* at 298.

184. *Id.*

185. *Id.*

to give him a vacation.[186] A day after the respondent received the ticket to return to the Philippines for a vacation, however, he was also served with a "report of dismissal, which he refused to accept."[187] This dismissal was allegedly because of the respondent's incompetence, as he failed to serve meals properly and keep his area in a clean condition.[188] As a result, the respondent filed a complaint for illegal dismissal against INC.[189] Electronic mails were presented to prove the petitioner's claim that the respondent was incompetent.[190] When the case reached the Court of Appeals, it "emphasized that electronic evidence, such as electronic mails [ ], must first be proved and authenticated before they are received in evidence."[191] The Court of Appeals did not give any probative value to the electronic mails because these were unauthenticated in accordance with the Rules on Electronic Evidence.[192] The Court agreed and added that these electronic mails even pertained to previous contracts of employment and is entirely unrelated to the case at bar.[193] Although it was never clearly tackled in this case as to which authentication process applies to electronic mails, the Author is of the opinion that it should be authenticated in accordance with rules pertaining to ephemeral electronic communications. This conclusion is based on previous cases decided by the Court, in which text messages and emails were often grouped together.[194]

Cambe v. Office of the Ombudsman[195] discusses the admissibility of audio evidence in a court of law.[196] This case involved a criminal charge of plunder, where funds totaling ₱517,000,000.00 were allegedly sourced from Senator Ramon 'Bong' Revilla Jr.'s Priority Development Assistance Fund (PDAF) from 2006 to 2010.[197] The crux of the Ombudsman's resolution was

186. *Id.*

187. *INC Shipmanagement, Inc.*, 769 SCRA at 298.

188. *Id.* at 299.

189. *Id.*

190. *Id.* at 307.

191. *Id.* at 302.

192. *Id.* at 307.

193. *INC Shipmanagement, Inc.*, 769 SCRA at 307.

194. Pacana, Jr. v. Pascual-Lopez, A.C. No. 8243, 594 SCRA 1, 13 (2009).

195. Cambe v. Office of the Ombudsman, G.R. Nos. 212014-15, 812 SCRA 537 (2016).

196. *Id.*

197. *Id.* at 565.

Cunanan's testimony.[198] "While he could have easily asked for a written confirmation of the authorization given by Revilla to Cambe, Cunanan himself admitted that he, instead, supposedly sought verification over the telephone."[199] Unfortunately, there was no audio recording of the telephone conversation that was presented or mentioned in court to support Cunanan's testimony.[200] The Court quoted the Rules on Electronic Evidence to illustrate the authentication process of audio evidence, such as a telephone conversation, to wit —

> Section 1. Audio, video[,] and similar evidence. – Audio, photographic[,] and video evidence of events, acts[,] or transactions shall be admissible provided [it] shall be shown, presented[,] or displayed to the court and shall be identified, explained[,] or authenticated by the person who made the recording or by some other person competent to testify on the accuracy thereof.[201]

The Court stressed that the identity of the person in the audio recording must be "reliably identified before the telephone conversation can be admitted in evidence and given probative value."[202] Thus, the identity of the person on the other end of the line must be reliably identified through "voice recognition or any other means[.]"[203] In *People v. Wagas*,[204] the Court also emphasized that for purposes of reliability and trustworthiness, a telephone conversation must be authenticated before it is admitted as evidence in court.[205] Since Cunanan's testimony was not supported by the actual audio evidence of the telephone conversation that transpired, his testimony is inadmissible in court.[206] It can be said that the authentication process of audio evidence presented in court should start with the identification of the speakers, followed by an explanation on how he or she recognizes their voices.[207]

198. *Id.* at 665.

199. *Id.*

200. *Id.*

201. *Cambe*, 812 SCRA at 666 (citing RULES ON ELECTRONIC EVIDENCE, rule 11, § 1).

202. *Cambe*, 812 SCRA at 666 (emphasis omitted).

203. *Id.* at 666-67 (emphasis omitted).

204. People v. Wagas, G.R. No. 157943, 705 SCRA 17 (2013).

205. *Cambe*, 812 SCRA at 667 (citing *Wagas*, 705 SCRA at 17).

206. *Cambe*, 812 SCRA at 668.

207. *Id.* & RULES ON ELECTRONIC EVIDENCE, rule 11, § 1.

These cases paint a small picture of the jurisprudential landscape of electronic evidence in the Philippines. It provides insight on how the different types of electronic evidence are treated in a court of law, particularly concerning their authentication.

V. INSUFFICIENCY OF THE RULES

This Chapter focuses on examining the authentication process of the current rules vis-à-vis the rise of deepfakes to determine their adequacy. The Author also references U.S. jurisprudence, where applicable, to address gaps in Philippine case law. Given the limited Philippine jurisprudence on electronic evidence, U.S jurisprudence may provide some guidance on how similar rules are to be implemented in a court of law.

A. Cheapfakes

Cheapfakes are not nearly as dangerous as deepfakes. As previously mentioned, cheapfakes are audio-visual manipulations which use conventional techniques like speeding, slowing, cutting, re-staging, or re-contextualizing footage to change the meaning and interpretation of media.[208] These alternations are significantly easier to detect as opposed to deepfakes. Nevertheless, it could still have detrimental effects. There have been cases where individuals have attempted to introduce fake or doctored evidence in court.

In the Philippines, the case of *Tecson v. Commission on Elections*[209] particularly stands out. Although the case does not actually deal with the presentation of electronic evidence, the petitioner attempted to submit fabricated evidence.[210] This came in the form of a marriage contract, which was edited through Photoshop.[211] In this case, the Court ultimately ruled that the evidence presented were mere fabrications.[212] These may not exactly be called as cheapfakes because what were presented were documentary evidence as opposed to electronic evidence; however, these documentary evidence were fabricated with the use of Adobe Photoshop, a program commonly used by people to create cheapfakes.[213] To prove this, the sworn statement of the

208. DeepTrust Alliance, *supra* note 42.

209. Tecson v. Commission on Elections, G.R. No. 161434, 424 SCRA 277 (2004).

210. *Id.* at 365-66.

211. *Id.* at 366.

212. *Id.*

213. *Id.*

person who was tasked to create the fabricated evidence was also presented.[214] Remmel G. Talabis narrated how he cleaned the signatures and appended these to the names contained in the marriage certificate.[215] Future cases which involve cheapfake evidence may not necessarily be decided the same way. What is unique in this case is that the creator of the fabricated evidence came forward to come clean about his actions.[216] Even if the creator of the fabricated evidence did not come forward, there were alternative ways to demonstrate that the evidence was a mere fabrication. While it is true that lawyers often employ creative strategies to protect their clients' interests, ethical considerations must also be taken into account when a lawyer attempts to present fabricated evidence in court. Nevertheless, there will still be some lawyers who will try to win at all costs despite crossing a few lines prohibited by the law.

In foreign jurisdictions, there have also been instances wherein people have tried to dupe the legal system by attempting to admit doctored pieces of evidence to get an advantage over the other party.[217] This will never change. There will always be attempts to fabricate evidence for as long as there are opportunities to do so. The evolving rules of authentication for pieces of evidence are driven by the reality that they are susceptible to fabrication and the likelihood that individuals will attempt to present such fabricated evidence in court.

Recently, in the U.K., there was an attempt to introduce a doctored audio in court to win a custody battle between the child of both parties.[218] Unfortunately, the wife's plan did not go as she intended as the lawyer of the husband was able to determine that the audio was a cheapfake by examining

214. *Id.*

215. *Tecson*, 424 SCRA at 366.

216. *Id.*

217. *See, e.g.*, Reynolds, *supra* note 23; David Mamone, Lafayette Attorney Brad Andrus Disbarred for Fabricating Invoices, Submitting False Evidence, *available at* https://web.archive.org/web/20221121133531/https://www.theadvocate.com/acadiana/news/courts/lafayette-attorney-brad-andrus-disbarred-for-fabricating-invoices-submitting-false-evidence/article_11df6b1e-7a5e-11ec-819e-0be4e6b5a158.html; & Debra Cassens Weiss, Lawyer Accused of Trying to File Fake News Article Doesn't Show Up for Sanctions Hearing, *available at* https://www.abajournal.com/news/article/lawyer-doesnt-show-up-for-hearing-weighing-sanctions-for-his-alleged-attempt-to-file-fake-news-article (last accessed Oct. 31, 2023) [https://perma.cc/5G4C-5NCJ].

218. Reynolds, *supra* note 23.

the metadata of the audio.[219] There are claims that the audio evidence presented was a deepfake instead of a mere cheapfake; however, the Author submits that the evidence could not have been a deepfake as the method used to detect the fabricated evidence was through the examination of the metadata of the audio evidence.[220] This is the typical method used when it comes to cheapfakes.[221] With regard to actually detecting deepfakes, there have been promising methods which usually involve artificial intelligence and machine learning.[222] It is quite poetic that the very technology that enables the creation of deepfakes is also the technology used to combat the same.

Although there are stark differences in the Revised Rules on Evidence and the Rules on Electronic Evidence as compared to the Federal Rules on Evidence, there are similarities when it comes to the authentication process. Therefore, it is still possible to examine U.S. jurisprudence and relate it to the Philippine context, especially given the scarcity of decided cases regarding various forms of electronic evidence in the Philippines.

B. Deepfakes

As previously mentioned, cases involving the presentation of deepfakes as evidence have yet to emerge, although there is an ongoing debate about whether the audio evidence in the previously mentioned U.K. custody case constituted a deepfake.[223] Nevertheless, the fact that these fabricated and spurious pieces of evidence are being presented in court even before the technology to create deepfakes has become available raises much concern because these deepfakes, although still in their infancy stage, have become quite mainstream. Thus, the Author opines that these deepfakes are likely to make their way into the courtroom in the near future.

---

219. *Id.*

220. Patrick Ryan, 'Deepfake' Audio Evidence Used in UK Court to Discredit Dubai Dad, *available at* https://www.thenationalnews.com/uae/courts/deepfake-audio-evidence-used-in-uk-court-to-discredit-dubai-dad-1.975764 (last accessed Oct. 31, 2023) [https://perma.cc/2B7U-9GC2].

221. Audio Forensic Expert, Authentication of Digital Audio Recordings, *available at* https://www.audioforensicexpert.com/authentication-of-digital-audio-recordings (last accessed Oct. 31, 2023) [https://perma.cc/8PCD-ZJ4J].

222. Kaveh Waddell, Defending Against Audio Deepfakes Before It's Too Late, *available at* https://www.axios.com/deepfake-audio-ai-impersonators-f736a8fc-162e-47f0-a582-e5eb8b8262ff.html (last accessed Oct. 31, 2023) [https://perma.cc/3X2P-JWVG].

223. Reynolds, *supra* note 23 & Ryan, *supra* note 220.

There are not many tools currently available to reliably detect deepfakes.[224] As mentioned earlier, deepfakes are usually detected through subtle imperfections in the fabricated evidence.[225] Through the use of artificial intelligence and machine learning, software may be invented to detect these subtle imperfections that might easily deceive people, including experts in the field.

It is a known fact that this deepfake phenomenon is a fairly recent occurrence.[226] The Rules on Electronic Evidence were promulgated back in 2001, almost 20 years ago.[227] It can be inferred reasonably that the Court had not taken into account the possibility of deepfake electronic evidence when it promulgated the Rules on Electronic Evidence. Although the 2019 Revised Rules on Evidence were promulgated and approved in 2019,[228] the Author is of the opinion that the possibility of deepfake electronic evidence being admitted in court was also not taken into account by the Court. This conclusion stems from the observation that the only significant changes to the authentication process of private documents under the 2019 Revised Rules on Evidence are the inclusion of a catch-all phrase, which the Court introduced to address potential gaps in the rules.[229] This is clearly insufficient, especially with regard to deepfakes. The lack of a proper procedure to address the gap in the rules could potentially wreak havoc in a court of law, as it would deteriorate one of the characteristics of the Philippine Judiciary — the characteristic of predictability.[230] It may be said that most of the laws in the Philippines are reactive laws as opposed to proactive laws. Unfortunately, this approach may prove to be ineffective, especially in light of the rise of modern technologies. The Law of Accelerating Returns states that "human progress

224. Hui, *supra* note 28.

225. *Id.*

226. Benjamin Goggin, From Porn to 'Game of Thrones': How Deepfakes and Realistic-Looking Fake Videos Hit It Big, *available at* https://www.businessinsider.com/deepfakes-explained-the-rise-of-fake-realistic-videos-online-2019-6 (last accessed Oct. 31, 2023) [https://perma.cc/7A8Y-TRVH].

227. *See* RULES ON ELECTRONIC EVIDENCE, rule 12, § 2.

228. *See* REVISED RULES ON EVIDENCE, whereas cl. para. 7.

229. *Id.* rule 132, § 20.

230. Land Bank of the Philippines v. Dalauta, G.R. No. 190004, 835 SCRA 1, 67-68 (2017) (citing Lazatin v. Desierto, G.R. No. 147097, 588 SCRA 285, 294-95 (2009)) (J. Jardeleza, concurring and dissenting opinion).

mov[es] quicker and quicker as time goes on[ ]"[231] — that technology grows exponentially.[232] With this premise, it is entirely possible that the reactive laws which Congress enacts would be regulating obsolete technology. By analogy, this could also apply to the rules that the Court promulgates to secure the processes of the Judiciary. Hence, a proactive approach is required to prevent these spurious pieces of evidence from entering the dockets of the courts.

1. Photo

In the conventional sense, photographs are pictures produced by a camera.[233] According to the 2019 Revised Rules on Evidence, photographs "include still pictures, drawings, stored images, x-ray films, motion pictures[,] or videos."[234] It should be noted that videos are included in photographs, but for the sake of this analysis, they will be discussed separately.

Although both the 2019 Revised Rules on Evidence and the Rules on Electronic Evidence contain their respective authentication procedures, it is clear that the authentication procedures of the Rules on Electronic Evidence is better fit in attempting to address the issue of deepfake electronic evidence.[235] Upon a quick perusal of the authentication process, as provided in the 2019 Revised Rules on Evidence, the second method of the authentication process states that "[b]efore any private document offered as authentic is received in evidence, its due execution and authenticity must be proved by any of the following means: ... (b) [b]y evidence of the genuineness of the signature or handwriting of the maker[.]"[236] A look at the wording used by the Court would immediately indicate that the rule pertains to a document in the traditional sense as opposed to a document as defined in the 2019 Revised Rules on Evidence. This is a similar conclusion reached by some experts although their discussion pertained to electronic document as opposed to electronic evidence in general.[237] The other authentication methods

231. Tim Urban, The AI Revolution: The Road to Superintelligence, *available at* https://waitbutwhy.com/2015/01/artificial-intelligence-revolution-1.html (last accessed Oct. 31, 2023) [https://perma.cc/FPD6-KX7P].

232. *Id.*

233. Merriam-Webster, Photograph, *available at* https://www.merriam-webster.com/dictionary/photograph (last accessed Oct. 31, 2023) [https://perma.cc/S59H-UQ3V].

234. REVISED RULES ON EVIDENCE, rule 130, § 2.

235. *See* REVISED RULES ON EVIDENCE, rule 132, § 20 & Disini Law, *supra* note 95.

236. REVISED RULES ON EVIDENCE, rule 132, § 20 (b).

237. *See* Disini Law, *supra* note 95.

prescribed by the Court also provide an indication that they pertain to traditional documents, save for the third method which is a sort of "catch-all" provision.[238] The Author posits that the 2019 Rules on Electronic Evidence's authentication procedures are insufficient to prevent deepfakes from being admitted as evidence in a court of law.

The Rules on Electronic Evidence have identical authentication rules for audio, video, and similar evidence.[239] The Rules' authentication process is extremely generic, such that it only requires that the audio, photographic, and video evidence "be shown, presented[,] or displayed to the court and shall be identified, explained[,] or authenticated by the person who made the recording or by some other person competent to testify on the accuracy thereof."[240] This authentication process promulgated by the Court could easily be circumvented with the use of deepfakes. Unfortunately, these forms of electronic evidence were usually taken at plain value because of the tendency of people to believe what they see.[241] The fact that these deepfakes are hyper realistic and extremely difficult to detect, that even experts would not be able to detect them reliably, would almost certainly spell doom for the country's litigation system.

In theory, one would simply need to present the photographic evidence to the Court and have it identified by a qualified individual who can attest to its accuracy for the evidence to be admitted. This would not, however, guarantee that photographic evidence would be accorded full probative weight as this would be left to the discretion of the judge. It is important to note that without complete information regarding the authenticity of the evidence, the judge's decision on its admissibility may be unreliable.

Unfortunately, Philippine jurisprudence regarding the admissibility of photographic evidence is notably deficient. The cases which dealt with the authentication requirements of electronic evidence, however, seem to be applying the black letter of the rules.[242] Therefore, it can be reasonably inferred that the courts will rely solely on Rules on Electronic Evidence to determine the authenticity, rather than establishing doctrines that outline requisites which are not expressly provided by the rules.

---

238. *See* REVISED RULES ON EVIDENCE, rule 132, § 20 (a) & (d); & Disini Law, *supra* note 95.

239. RULES ON ELECTRONIC EVIDENCE, rule 11, § 1.

240. *Id.*

241. *See* Marchant, *supra* note 64, at 151.

242. *See Cambe*, 812 SCRA at 666.

It should also be noted that unlike the 2019 Revised Rules on Evidence, the Rules on Electronic Evidence does not have a "catch-all" provision for photographic evidence.[243] Consequently, the authentication process provided is the sole method for authenticating photographic evidence so that it can be recognized and considered as evidence by the court.[244]

The authentication process contained in the Rules on Electronic Evidence does not ensure that the admitted pieces of evidence are genuinely authentic, as this process can be easily circumvented by a party attempting to present deepfake evidence. The Rules do not even require that the person who actually took the photo be presented to authenticate the photo, instead, they allow a competent person who can testify as to its accuracy and the circumstances surrounding the photo.

2. Audio

With regard to audio evidence, the same rule applies — that the same should be presented in court and identified or explained by a person who made the recording or some other person competent to testify as to such evidence.[245] Audio evidence provides for a better view as to why the current rules, both the 2019 Revised Rules on Evidence and the Rules on Electronic Evidence, are insufficient to prevent the admissibility of deepfake evidence.

The case of *Cambe* used the authentication process in the Rules on Electronic Evidence to determine the admissibility of the audio recording being proposed.[246] The discussion on audio evidence was framed within the context of a telephone conversation between the parties.[247] Unfortunately, a recording of the audio evidence was not presented, so the Court held that the "occurrence of the alleged telephone conversation is rendered highly suspect, if not improbable[.]"[248] This case illustrates the attitude of the Court by demonstrating how it approaches issues related to the admissibility and evaluation of evidence. Here, the court did not mention the Anti-

243. *See* RULES ON ELECTRONIC EVIDENCE.

244. *Id.*

245. RULES ON ELECTRONIC EVIDENCE, rule 11, § 1.

246. *Cambe*, 812 SCRA at 666.

247. *Id.*

248. *Id.*

Wiretapping Law[249] of the Philippines. This law provides for the inadmissibility of evidence acquired in violation to the Anti-Wiretapping Law, to wit —

> Any communication or spoken word, or the existence, contents, substance, purport, effect, or meaning of the same or any part thereof, or any information therein contained obtained or secured by any person in violation of the preceding sections of this Act shall not be admissible in evidence in any judicial, quasi-judicial, legislative[,] or administrative hearing or investigation.[250]

Therefore, it can be reasonably inferred that this is read into the Rules on Electronic Evidence.[251] Unfortunately, regardless if audio evidence acquired in violation of the Anti-Wiretapping Law is deemed inadmissible in evidence by the power of the legislature, this would still be insufficient to prevent audio deepfakes from being admitted into evidence.

The authentication process for audio evidence begins with one party presenting the audio evidence in court, followed by a witness's testimony. This witness must identify the speakers, clearly explain how they recognize the voices, and affirm that the recording was obtained in compliance with the Anti-Wiretapping Law.[252] With these, it can be seen that there are multiple loopholes when it comes to deepfake audio evidence. Since the Rules on Electronic Evidence provides that the testifying witness need not be the same person who actually conducted the recording of the audio, then a witness who is familiar with the voices of the people in the audio is sufficient.[253] This presents multiple opportunities to exploit the gap in the rules. Recent development in deepfake audio technology has made it a lot easier to make deepfake audio.[254] One would only need a five-second voice sample of a subject and artificial intelligence and deep learning technology would

249. An Act to Prohibit and Penalize Wire Tapping and Other Related Violations of the Privacy of Communication, and for Other Purposes, Republic Act No. 4200 (1965) (also known as the Anti-Wiretapping Law).

250. *Id.* § 4.

251. *Id.* § 5.

252. *See* RULES ON ELECTRONIC EVIDENCE & Republic Act No. 4200.

253. RULES ON ELECTRONIC EVIDENCE, rule 11, § 1.

254. Two Minute Papers, Video, *This AI Clones Your Voice After Listening for 5 Seconds*, Nov. 13, 2019, YOUTUBE, *available at* https://www.youtube.com/watch?v=0sR1rU3gLzQ (last accessed Oct. 31, 2023) [https://perma.cc/A4BH-WFN4].

synthesize the rest of a subject's voice.[255] This does not bode well for the argument on the sufficiency of the Rules of Court. Theoretically, deepfake audio evidence could be presented in court without the lawyer or the witness being aware that the audio in question is a deepfake. It is not that difficult to identify a speaker's voice that one is familiar with, especially if the speaker has a distinct voice. If these are the only safeguards provided by the rules with regard to the authentication of audio evidence, it is clear that these are insufficient to stop the rise of deepfakes.

3. Video

Video evidence may be likened to photographic evidence and audio evidence since video is essentially a combination of both types of electronic evidence. The tools used to detect these video deepfakes, however, rely on identifying subtle imperfections in the footage, such as unnatural blinking, inconsistent head movements, and irregular breathing patterns.[256] These slight imperfections are detected using the same technology that is behind deepfakes.[257] Currently, the applicable rule with regard to the authentication of video evidence is provided by the Rules on Electronic Evidence.[258] These procedures have not changed since its inception in 2001.[259] While it may be increasingly difficult to present deepfake videos as authentic evidence in court, there are numerous instances where isolated actions could potentially be manipulated into deepfake videos.

Due to the dearth of jurisprudence on the authentication of electronic evidence, the Court has not expounded much on the actual steps that one needs to take to prove the authenticity of the video evidence, other than the black letter of the Rules on Electronic Evidence. A brief examination of the processes employed by foreign jurisdictions could be beneficial. In the U.K., recommended best practices for properly authenticating electronic evidence include the use of "technical methods [such as] encryption, watermarking, or digital signature."[260] The authentication process in the U.K. courts, however, is not limited to these technical methods; the "layman's approach" remains an

255. *Id.*

256. Hui, *supra* note 28.

257. *Id.*

258. RULES ON ELECTRONIC EVIDENCE, rule 11, § 1.

259. *Id.* rule 12, § 2.

260. Francis Lim, *CCTV Footage as Evidence*, PHIL. DAILY INQ., Oct. 17, 2014, *available at* https://business.inquirer.net/180469/cctv-footage-as-evidence (last accessed Oct. 31, 2023) [https://perma.cc/6ARX-E7DJ].

acceptable option as well.[261] The aforesaid layman's approach is the conventional approach that is codified in the Philippine Rules on Electronic Evidence. It essentially provides that electronic evidence may be authenticated by any competent person who can testify "as to [its] exactness or accuracy ... ."[262] The use or non-use of the recommended technical methods provided by the U.K. rules of procedure does not directly impact the admissibility of electronic evidence; rather, it enhances its probative value.[263] This could also be used analogously in the Philippine context, as it is evident that the standards established by the Rules on Electronic Evidence are minimal and do not account for the potential challenges posed by deepfake technology. Since the admissibility of evidence is determined by its relevance and competency, it is adequate as long as these pieces of evidence meet the specified requirements.[264]

With these, it can be said that the current rules governing the admissibility of electronic evidence are insufficient to prevent deepfake videos from being admitted as evidence in court. The ease with which the rules can be circumvented, the absence of specific procedures established by the Court for authenticating electronic evidence, and the hyperrealism introduced by deepfakes all highlight the inadequacies of the current rules. Above all, deepfake technology is still in its infancy stage and has not even matured yet. The technology behind deepfakes is expected to advance and mature significantly in the coming years.

C. Expert Evidence

An argument against the probability of deepfakes being admitted into evidence is that expert testimony could be presented to rebut the authentication of deepfakes. While the Author admits that this could be a potential solution, it should, nevertheless, be noted that the technology behind deepfakes is still young and is constantly improving.

261. *Id.* (citing Sison v. People, G.R. Nos. 108280-83, 250 SCRA 58, 76 (1995) & Republic v. Court of Appeals, G.R. No. 103882, 299 SCRA 199, 324 (1998)).

262. *Id.*

263. *Id.*

264. REVISED RULES ON EVIDENCE, rule 128, § 3.

The number of experts in deepfake technology is relatively small compared to those in more established fields, as the deepfake phenomenon has emerged only in recent years.[265]

Fortunately, major tech companies have invested in the prevention of deepfakes.[266] Most experts in this field are either from abroad or affiliated with large foreign tech companies.[267] At present, there are no experts on deepfakes in the Philippines, and it is a "challenge [ ] to get experts with image manipulation to spot and combat [deepfakes]."[268] The scarcity of experts on the field of deepfakes would certainly be a roadblock in using experts to rebut the probable admissibility of deepfakes in court.

As previously mentioned, the continuous advancement of deepfake technology can even deceive experts, which is why leveraging artificial intelligence and machine learning offers a more effective solution.[269] Most of the tech companies' solution are based on the same technology which created deepfakes by detecting the subtle imperfections which are invisible to the naked eye.[270] This approach is preferable, as the human eye can often be unreliable.[271] Another approach is the use of deepfake algorithms.[272] Unfortunately, most of these tools are primarily designed to detect video deepfakes, as they focus on indicators like eye blinking and head movements. Audio deepfakes could potentially be detected using similar tools designed to identify imperfections in the synthesized syllables of a persom's voice, as compared to those enunciated in the original recording.

It should also be a concern that expert testimony could become expensive, potentially limiting access to justice for those who are less fortunate. The

265. Mika Westerlund, *The Emergence of Deepfake Technology: A Review*, TECH. INNOVATION MGMT. REV., Volume No. 9, Issue No. 11, at 48.

266. *See* Ryan Daws, Facebook Pledges Crackdown on Deepfakes Ahead of the US Presidential Election, *available at* https://artificialintelligence-news.com/2020/01/08/facebook-crackdown-deepfakes-us-presidential-election (last accessed Oct. 31, 2023) [https://perma.cc/5MZ8-B8EW].

267. *Id.*

268. Mikael Angelo Francisco, Make It 'Til You Fake It: How Deepfake Brought a Pinoy Showbiz Icon Back to Life, *available at* https://www.flipscience.ph/technology/deepfake-pinoy-icon (last accessed Oct. 31, 2023) [https://perma.cc/E6Y9-MX5R].

269. Marchant, *supra* note 64, at 151.

270. Waddell, *supra* note 222.

271. *Id.*

272. Vincent, *supra* note 9.

scarcity of experts based in the Philippines, or their complete absence, would make it increasingly difficult to challenge deepfake evidence in court, particularly for individuals lacking sufficient resources. This should be a grave concern, as it would turn the judiciary processes into a rich man's game. Another argument against the use of expert evidence is that this is not even required to be used in court, as the opinions of experts are subject to the discretion of the judge.[273] The court is not bound by the opinions given by the expert with regard to the matter of contention before the court.[274] At the end of day, judges must make their independent judgement to determine the authenticity of evidence being presented.[275]

D. Remedies

The technology behind deepfakes is constantly evolving and improving.[276] Deepfake detection tools are actively working to create software and develop algorithms aimed at identifying imperfections in deepfakes.[277] The situation remains a perpetual cat-and-mouse game because the more deepfake detectors combat the threat of deepfakes, the more deepfakers would create techniques or tools to circumvent the safeguards placed to detect deepfakes.[278] Thus, the possibility of mistakenly determining whether a piece of electronic evidence is a deepfake cannot be ruled out. There might come a time where perfect deepfakes can be created. It would also be entirely possible that a previous determination of a document's authenticity can be subsequently proven to be a deepfake with the use of future technology and techniques, which were previously unavailable.

1. Legal Implications of Erroneous Deepfake Certifications

In an ideal world, a determination that a piece of evidence is a deepfake should be binding on the parties to the specific case. Unfortunately, technology used to detect deepfakes is nowhere near perfect and inaccuracies are bound to happen. Thus, there should be contingencies put in place in cases where erroneous certifications are issued.

273. REVISED RULES ON EVIDENCE, rule 130, § 52.

274. *Id.*

275. *Id.*

276. Waddell, *supra* note 222.

277. Farid, *supra* note 39.

278. *Id.*

2. Liability of the Party Offering Deepfake Electronic Evidence

In cases where the pieces of electronic evidence being offered are proven to be deepfakes, the party presenting such spurious pieces of evidence shall be criminally liable in accordance with Articles 180 to 184 of the Revised Penal Code.[279] These provisions cover the different instances of false testimony.[280] It is elementary in the Revised Rules on Evidence that the presentation of evidence must be accompanied by a testimony regarding such piece of evidence.[281] With regard to the party presenting such testimony with the knowledge that the testimony being presented is false, he or she should be liable under Article 184 of the Revised Penal Code.[282] The provision reads —

> [Article] 184. Offering False Testimony in Evidence. — Any person who shall knowingly offer in evidence a false witness or testimony in any judicial or official proceeding, shall be punished as guilty of false testimony and shall suffer the respective penalties provided in this section.[283]

With regard to the person who actually gave the false testimony in court, he or she would be liable under Articles 180 to 183 of the Revised Penal Code, depending on whether the testimony is favorable or against the defendant in criminal cases, the testimony is offered in a civil case, or the testimony is offered in other cases.[284] Articles 180 to 183 provide —

> [Article] 180. False Testimony Against a Defendant. — Any person who shall give false testimony against the defendant in any criminal case shall suffer:
>
> (1) The penalty of *reclusión temporal*, if the defendant in said case shall have been sentenced to death;
>
> (2) The penalty of *prisión mayor*, if the defendant shall have been sentenced to *reclusión temporal* or *perpetua*;
>
> (3) The penalty of *prisión correccional*, if the defendant shall have been sentenced to any other afflictive penalty; and
>
> (4) The penalty of *arresto mayor*, if the defendant shall have been sentenced to a correctional penalty or a fine, or shall have been acquitted.

---

279. An Act Revising the Penal Code and Other Penal Laws [REV. PENAL CODE], Act No. 3815, arts. 180-184 (1930).

280. *Id.*

281. *See* REVISED RULES ON EVIDENCE, rule 132, § 20 & RULES ON ELECTRONIC EVIDENCE, rule 11.

282. REV. PENAL CODE, art. 184.

283. *Id.*

284. *Id.* arts. 180-183.

> (5) In cases provided in subdivisions 3 and 4 of this article the offender shall further suffer a fine not to exceed 1,000 pesos.
>
> [Article] 181. False Testimony Favorable to the Defendant. — Any person who shall give false testimony in favor of the defendant in a criminal case, shall suffer the penalties of *arresto mayor* in its maximum period to *prisión correccional* in its minimum period and a fine not to exceed 1,000 pesos, if the prosecution is for a felony punishable by an afflictive penalty, and the penalty of *arresto mayor* in any other case.
>
> [Article] 182. False Testimony in Civil Cases. — Any person found guilty of false testimony in a civil case shall suffer the penalty *of prisión correccional* in its minimum period and a fine not to exceed 6,000 pesos, if the amount in controversy shall exceed 5,000 pesos, and the penalty of *arresto mayor* in its maximum period to *prisión correccional* in its minimum period and a fine not to exceed 1,000 pesos, if the amount in controversy shall not exceed said amount or cannot be estimated.
>
> [Article] 183. False Testimony in Other Cases and Perjury in Solemn Affirmation. — The penalty of *arresto mayor* in its maximum period to *prisión correccional* in its minimum period shall be imposed upon any person who, knowingly making untruthful statements and not being included in the provisions of the next preceding articles, shall testify under oath, or make an affidavit, upon any material matter before a competent person authorized to administer an oath in cases in which the law so requires.
>
> Any person who, in case of a solemn affirmation made in lieu of an oath, shall commit any of the falsehoods mentioned in this and the three preceding articles of this section, shall suffer the respective penalties provided therein.[285]

The Author puts forward the argument that regardless if a party introduced spurious pieces of evidence, he or she should not be considered to have waived his or her right to present evidence because there are also instances when deepfake evidence is introduced in good faith. It may be argued that the party who presents deepfake electronic evidence in court may be held liable for indirect contempt because it may be considered as an "improper conduct tending, directly or indirectly, to impede, obstruct, or degrade the administration of justice[.]"[286]

285. *Id.*

286. 1997 RULES OF CIVIL PROCEDURE, rule 71, § 3 (d).

3. Liability of the Administrative Agency for Erroneous Certifications

Generally, if a party wants to appeal a decision of an administrative agency, the doctrine of exhaustion of administrative remedies must be followed.[287] "The premature invocation of a court's intervention is fatal to one's cause of action."[288] Thus, any party questioning the decision of the administrative agency tasked with issuing certifications regarding deepfakes should follow the doctrine of exhaustion of administrative remedies before seeking judicial intervention. Findings of fact by administrative agencies are generally given great weight and are not disturbed, unless "the decision was rendered as a result of fraud, imposition[,] or mistake, other than error of judgment, in estimating the value of the evidence."[289] If there is grave abuse of discretion with regard to the certification process by which the administrative agency handled, the remedy of the aggrieved party would be *certiorari*.[290]

4. Subsequent Procedural Remedies

As previously mentioned, there may be instances wherein new technology could prove that a piece of electronic evidence previously deemed authentic is now adjudged as a deepfake or conversely, a piece of electronic evidence previously deemed to be a deepfake is now adjudged to be authentic. In instances such as these, the Rules of Court does not leave the aggrieved party without a remedy.

In civil cases, the aggrieved party may file a petition for relief from judgment if the judgment is "taken against a party in any court through fraud, accident, mistake, or excusable negligence[.]"[291] So long as the petition is "filed within sixty (60) days after the petitioner learns of the judgment, final order, or other proceeding to be set aside, and not more than six (6) months after such judgment or final order was entered," and it is accompanied with affidavits stating the cause of action, this remedy may aid an aggrieved party.[292] It should be emphasized that the grounds to avail of this remedy are limited to those previously mentioned.[293]

287. HECTOR S. DE LEON & HECTOR M. DE LEON, JR., ADMINISTRATIVE LAW: TEXT AND CASES 360 (6th ed. 2010).

288. *Id.*

289. *Id.* at 426.

290. *Id.* at 326.

291. RULES OF CIVIL PROCEDURE, rule 38, § 1.

292. *Id.* rule 38, § 3.

293. *Id.* rule 38, § 1.

Another remedy would be a petition for annulment of judgement.[294] This remedy is only limited to two grounds, namely: (1) extrinsic fraud and (2) lack of jurisdiction. It must also be filed "within four (4) years from its discovery; and if based on lack of jurisdiction, before it is barred by laches or estoppel."[295]

In criminal cases, the 2000 Revised Rules of Criminal Procedure does not exactly have a provision similar to the ones previously mentioned in civil cases; however, the Revised Rules of Criminal Procedure does grant the remedy of a motion for a new trial or a motion for reconsideration in cases where —

> Section 2. Grounds for a new trial. — The court shall grant a new trial on any of the following grounds:
>
> (a) The errors of law or irregularities prejudicial to the substantial rights of the accused have been committed during the trial;
>
> (b) The new and material evidence has been discovered which the accused could not with reasonable diligence have discovered and produced at the trial and which if introduced and admitted would probably change the judgment.[296]

Unfortunately, this remedy is only available before a judgment of conviction becomes final.[297] Thus, the accused would not be able to avail of this remedy if the judgment has become final thereby limiting the chances that the accused can prove his or her innocence.

VI. CONCLUSION AND RECOMMENDATION

A. Conclusion

It can be said that both Rules which govern evidence in Philippine courts are not exactly in conflict with each other. The question as to which of these Rules govern the authentication and admissibility of electronic evidence may be solved with the aid of the doctrines of statutory construction. Although the Rules on Electronic Evidence are more specific, the newly amended Revised Rules on Evidence also tackles a similar subject such as the audio, video, and photographic evidence because of the inclusion of these types of evidence in the definition of documentary evidence.[298] The Author agrees with the opinions of Atty. JJ Disini and Justice Singh that the Rules on Electronic

294. *Id.* rule 47.

295. *Id.* rule 47, § 3.

296. 2000 REVISED RULES OF CRIMINAL PROCEDURE, rule 121, § 2.

297. *Id.* rule 121, § 1.

298. REVISED RULES ON EVIDENCE, rule 130, § 2.

Evidence, particularly the rules promulgated on the authentication of electronic documents, were not superseded by the 2019 Revised Rules on Evidence.[299] Although they focused on electronic documents, there is no reason why the same should not apply to the other forms of electronic evidence relevant to this Note. In addition, the use of the authentication procedures in the Rules on Electronic Evidence is more geared towards electronic evidence as opposed to the 2019 Revised Rules on Evidence which are more geared towards the conventional non-electronic types of evidence. It was also opined that the reason why the Court included electronic evidence in the definition of documentary evidence was to simply bridge the gap between the rules; however, it was not intended to supersede the Rules on Electronic Evidence.[300] Thus, it is also the position of the Author that the Rules on Electronic Evidence is the more appropriate rule to attempt to counteract the rise of deepfakes.

Regardless of which rule will govern the authenticity and admissibility of electronic evidence such as audio, video, and photographs, both Rules are insufficient to address the looming problem of deepfakes being introduced into evidence.

The 2019 Revised Rules on Evidence was not built to address electronic evidence. This can be gleaned from the wording used by the Court. They simply included a catch-all provision with regard to the proof of authenticity of private documents.[301] Other than the included catch-all provision, there was no more amendment to the original proof of private documents provision under the 2019 Revised Rules on Evidence.[302] The rule, before amendment, was never intended to address electronic evidence which was why the Court promulgated the Rules on Electronic Evidence. Thus, it is not reasonable to infer that the proof of private documents provision, regardless of the amendment, was not intended to address electronic evidence.

With regard to the Rules on Electronic Evidence, there is a complete lack of specific procedures to prevent deepfakes from being admitted into evidence. To rub salt into the wound, the dearth of jurisprudence regarding the process of admissibility leaves one in limbo as to the proper authentication procedures of electronic evidence. These Rules do not guarantee that the deepfake evidence would not get admitted in court. These Rules were promulgated at a time wherein deepfake technology has not yet been

299. Disini Law, *supra* note 95.

300. *Id.*

301. REVISED RULES ON EVIDENCE, rule 132, § 20.

302. *See id.*

conceived. Deepfakes, in the recent years, have greatly improved to the point that it can even trick experts. The authentication provided by both Rules are clearly insufficient to prevent the admission of deepfakes into evidence. The potential problems that could arise are unimaginable. At the very least, the mere existence of deepfakes would devalue other authentic electronic evidence especially since the Rules do not guarantee their detection. This would take away the predictability of the judiciary as the parties would be able to present spurious evidence to turn the tides of the decision.

B. Recommendation

To prevent the introduction of deepfake evidence into court processes, a provision that requires electronic evidence, such as audio, video, and photographic evidence, to undergo a mandatory authentication process to be determined by the National Bureau of Investigation (NBI) should be included in the Rules of Court as a special provision to address deepfake electronic evidence. According to the National Bureau of Investigation Reorganization and Modernization Act,[303] the NBI has the power to:

...

> (g) Establish and maintain a Forensic and Scientific Research Center which shall serve as the primary center for forensic and scientific research in furtherance of scientific knowledge in criminal investigation, detection, evidence collection[,] and preservation, and provide the necessary training therefor;
>
> (h) Establish and maintain a Cyber Investigation and Assessment Center which shall serve as the nerve center for computer information technologies, data on cybercrime cases, computer intrusion, threats, and other related crimes or activities;
>
> (i) Establish and maintain an integrated, comprehensive, and state-of-the art network of equipment and facilities to be used by the NBI in its criminal investigation, detection, and evidence gathering, and to provide the corresponding training in this regard;
>
> (j) Request the assistance of the Philippine National Police (PNP), Armed Forces of the Philippines, or any other agency of the government, including government-owned and/or -controlled corporations, in its anti-crime drive. Such assistance may include the use of the agency's

303. An Act Reorganizing and Modernizing the National Bureau of Investigation (NBI), and Providing Funds Therefor [National Bureau of Investigation Reorganization and Modernization Act], Republic Act No. 10867 (2015).

> personnel and facilities upon prior approval by the head of the agency concerned;
>
> ...
>
> (o) Perform such other functions as the President or the Secretary of Justice may assign.[304]

The powers given to the NBI by the aforesaid law places the NBI in the best position to sift through the pieces of electronic evidence which are suspected to be deepfakes.[305] Although the Author is aware that the Department of Justice, as well as the Philippine National Police, both have cybercrime divisions/offices, the powers possessed by these agencies are mainly focused on investigation and prosecution of cybercrime,[306] as opposed to the NBI, which is mandated to be the "nerve center for computer information technologies ... ."[307] Thus, the NBI should be the proper agency to lead the vanguard in the battle against deepfakes, not only to prevent the introduction of these spurious pieces of electronic evidence in court, but also to protect the general public against its harmful effects.

The current 2019 Revised Rules on Evidence's definition of documentary evidence has caused much confusion as to which of the Rules are applicable to electronic evidence.[308] The Court should clarify their reasons as to why they included the definition of electronic evidence in documentary evidence so that scholars, practitioners, and judges alike would be able to understand and apply the Rules as how the Supreme Court intended. The Author recommends that the Rules on Electronic Evidence be amended to contain a specific provision regarding the proper authentication procedure for pieces of evidence that are suspected of being deepfakes. The better approach is to completely overhaul the provision on the authentication of audio, video, and other similar evidence to include a directive that these pieces of electronic evidence must undergo a mandatory verification process conducted by the NBI to apprise the judiciary of the information on whether the evidence presented is a deepfake or not. An alternative approach might be to limit the

---

304. *Id.* § 4 (g)-(j) & (o).

305. *See id.*

306. *See* An Act Defining Cybercrime, Providing for the Prevention, Investigation, Suppression and the Imposition of Penalties Therefor and for Other Purposes [Cybercrime Prevention Act of 2012], Republic Act No. 10175, §§ 10-11 (2012) & Rules and Regulations Implementing the Cybercrime Prevention Act of 2012, Republic Act No. 10175, rule III, § 10 & rule 6, § 28 (2015).

307. National Bureau of Investigation Reorganization and Modernization Act, § 4 (h).

308. Disini Law, *supra* note 95.

mandatory verification process of the NBI to those pieces of electronic evidence which are specifically alleged to be deepfakes. This approach might be more efficient as it would not disrupt the operations of the courts to a great extent and would not unnecessarily clog the dockets of the courts.

This recommendation would still have some effect on the court's dockets as it proposes to add an extra step into the authentication process of electronic evidence. The Author believes, however, that the benefits would far outweigh the costs because accurate and predictable decisions are much better than inconsistent decisions. This would also provide the less fortunate the opportunity to contest deepfake evidence being presented against them. The lack of experts in the Philippines, as well as the costs and logistics behind having these experts testify in court will unduly favor those who have the resources to afford the analysis of experts. Thus, it would be a necessary sacrifice to protect every individual's rights.

The NBI should have the necessary powers and functions as given to it by the NBI Reorganization and Modernization Act. The powers and functions are general enough for it to be able to address deepfakes. To establish this additional function and inform the NBI, a law should be enacted as a precautionary measure. Due to the separation of powers, an amendment to the Rules would not suffice to order the NBI to verify electronic evidence. This would have to be addressed through a law, similar to when the NBI was reorganized in 2012. With the additional function, the NBI would need a larger budget to acquire the required expertise and equipment to effectively carry out the mandate of the law. The creation of the law would also facilitate the easier acquisition of a higher budget for the NBI from Congress because it was Congress itself, which expanded the powers of the NBI, and logic would dictate that a budget increase would be crucial for the NBI to perform its new role in the fight against deepfakes and other high technologies effectively and efficiently.

It was previously mentioned that the House of Representatives already has a pending bill which would attempt to address deepfakes. The pending bill will only need a short provision which would amend the NBI Reorganization and Modernization Act; adding a provision which would essentially compel it to issue verifications on evidence which are suspected to be deepfakes and mandate it to essentially be experts on deepfakes as well as potential harmful technologies. Realistically, it is close to impossible for the NBI to become experts on deepfakes overnight. Thus, the law must contain a provision on the possibility of outsourcing the verification process to a private entity or a government-owned and controlled corporation under the supervision and control of the NBI for at least a year; this is for the NBI to catch up with the advancements of deepfake technology. Generally, the cost of the equipment

to be used will be taken from the coffers of the National Treasury, however, each certification will require the party who seeks the certification of the NBI to cover the costs of regulation. The Author envisions the system to be one similar to how citizens acquire an NBI clearance. The person seeking the NBI clearance is also charged with the necessary costs of regulation.

Digital files can easily be tampered; these files are not made to be tamper-evident.[309] Fortunately, "digital forensic analysts may be able to identify some digital characteristics they can use to detect meddling, but these indicators [do not] always paint a reliable picture of whatever digital manipulations a photo has undergone."[310] These digital characteristics indicators, however, disappear when the digital files undergo "post-processing" such as uploading.[311] To potentially address the problem of deepfakes, researchers propose the adoption of software which is embedded in a camera to create a sort of a watermark, which is to be used to verify a video's authenticity with accuracy.[312] This would create a "fingerprint at the moment of a film's recording"[313] which will be used as a reference upon which any playback of the video can be compared with to determine likelihood of tampering.[314] Based on research, this would improve the detection rate by 100% bringing it to an impressive 90% manipulation detection rate.[315] Unfortunately, this technology is still at its infancy stage, and may initially be expensive. Although this method would give better assurance that the presented electronic evidence is authentic, mandating this software might also restrict the Rules as every instance of photo, video, and audio evidence presented in court would necessarily have to be taken using a camera with the appropriate software. This might arguably be a step backwards, but without the necessary tools to detect deepfakes, its impact could cause a lot of problems, especially in the modern digital age.

309. Lily Hay Newman, To Fight Deepfakes, Researchers Built a Smarter Camera, *available at* https://www.wired.com/story/detect-deepfakes-camera-watermark (last accessed Oct. 31, 2023) [https://perma.cc/2XV2-UNFS].

310. *Id.*

311. *Id.*

312. *Id.*

313. Simon Parkin, *The Rise of the Deepfake and the Threat to Democracy*, GUARDIAN, June 22, 2019, *available at* https://www.theguardian.com/technology/ng-interactive/2019/jun/22/the-rise-of-the-deepfake-and-the-threat-to-democracy (last accessed Oct. 31, 2023) [https://perma.cc/8JVZ-AJPL].

314. Newman, *supra* note 309.

315. *Id.*