UNILAG LAW REVIEW

# FEDERAL RULES OF EVIDENCE IN THE UNITED STATES OF AMERICA AND THE CHALLENGES OF AUTHENTICATION IN THE AGE OF DEEPFAKE TECHNOLOGY

**Prince Amadi**

**VOLUME 7, NO. 1 (2024)**

# FEDERAL RULES OF EVIDENCE IN THE UNITED STATES OF AMERICA AND THE CHALLENGES OF AUTHENTICATION IN THE AGE OF DEEPFAKE TECHNOLOGY

**Prince Amadi***

## ABSTRACT

*Technology has deeply impacted society and how activities are conducted. This is not different from the development of deepfake technology. Deepfake is of serious concern to the public generally, but more particularly, it is of grave concern to the admission of evidence at trial. This paper espouses deepfake technology's inherent challenges to the Federal Rules of Evidence in the United States. As the use of this technology and its dangerous tendencies expands, it continues to bear on its multiplier effects in the judicial system concerning the authenticity of pieces of evidence in the courtroom. The paper argues that in the wake of this unprecedented development of deepfake technology, capable of manipulating evidence and misleading the courts, courts must deploy extra measures to ensure that evidence presented before it is the original. Thus, the paper proposes measures that courts may implement to address the challenges of deepfake.*

* Prince Amadi is the Lead Partner with the firm of Mathsman Attorneys & Solutions and a Research Fellow at the University of Ibadan Centre for Petroleum, Energy Economics and Law. Mr. Amadi's practice areas include human rights, criminal law (with a particular focus on financial and cybercrimes), cybersecurity, data privacy, and regulatory governance.

## 1.0 INTRODUCTION

The rise of deepfake technology has been rapid in recent years making it easy for individuals to generate artificial media files such as images, sound, or video.[1] With the aid of deepfake technology, almost perfect media are created to the deception of even the most careful among humans. To do this, a specific machine learning algorithm-deepfake is used to create images, audio, or video of individuals who never did nor said the things represented in the manipulated media.[2] In some quarters, it has been argued that deepfakes can be put to good use as seen in entertainment, where jokes and other comic reliefs were generated for the sole purpose of entertainment and modification in the film industry.[3] Still, the use of deepfakes has gone beyond entertainment to other realms of life, such as politics, and private and business lives; thereby resulting in the rise in misinformation, defamation, and alteration of political campaigns and causing personal harm to individuals.

The most targeted in this category are politicians, musicians, and movie stars.[4] A typical illustration of the use of this technology is the superimposition of the picture of former President Barack Obama over voice to create a video where it depicts him cussing and

[1] D. Harwell, "*Top AI* Researchers Race to Detect 'Deepfake' Videos: 'We are Outgunned," available at, In the race to detect deepfakes, AI researchers say they are "outgunned" - The Washington Post (accessed 7 May 2023).
[2] K Fagan "A Viral Video That Appeared To Show Obama Calling Trump a 'Dips---' Shows a Disturbing New Trend Called 'Deepfakes," available at: Deepfake: Fake Obama Video Calling Trump Dipshit Is a Disturbing Trend (businessinsider.com) (accessed 26 March 2023).
[3] L. Lamyanba, R. Maklachur, R, & K.J Soon, "Challenges and Applications of Face Deepfake" in J. Hieyong & S. Kazuhiko (eds.), *Communications in Computer and Information Science: Frontiers of Computer Vision*, 27th International Workshop, IW-FCV 2021 Daegu, South Korea, 22–23 February 2021, Revised Selected Papers, (Springer: 2021 Daegu, South Korea), p. 131.
[4] U. M. Bahar, S. Afsana, et al, *Deep Insights of Deepfake Technology: A Review* retrieved from [2105.00192] Deep Insights of Deepfake Technology : A Review (arxiv.org) (accessed 27 March 2023).

abusing former President Donald Trump.[5] In the viral video, Obama was represented as calling Trump "a dipshit."[6] The video was later discovered to be a deepfake. Despite BuzzFeed's pacification afterwards to the effect that the video was a product of deepfake technology, the ugly impact persists and reigns supreme within the political sphere. Obama is not alone in this. Other celebrities equally suffered similar and more horrifying experiences of deepfakes. The manipulation of former House Speaker, Nancy Pelosi's speech speaks volumes of the inherent danger of deepfakes.[7] In 2017, a face swap of Gal Gadot was made to create a video of her having sex with her stepbrother.[8] As expected, at first this was believed to be true, perhaps, still believed by many who have already seen the video. Sadly, the video was generated using a machine learning algorithm – deepfakes. Others whose images have been superimposed to create porn include Scarlett Johansson, Maisie Williams, Taylor Swift, and Aubrey Plaza; all of whom have suffered in their individual lives because of the deepfakes manipulation of their images.[9]

Outside the United States, a tapestry of deepfakes also exists. In 2018, the public was greeted by an atmosphere of speculation by the Gabonese social media regarding a purported New Year address to the country ascribed to the president. Many believe the address was a product of deepfakes, a manipulation to deceive the people of Gabon.[10]

---

[5] Ibid.
[6] Ibid.
[7] A. Henry, P. Giorgio, et al, "The State of Deepfakes: Landscape, Threats, and Impact" (September 2019) *Deeptrace*, p. 11.
[8] S. Cole, "AI-Assisted Fake Porn is Here and We Are All Fucked," available at: AI-Assisted Fake Porn Is Here and We're All Fucked (vice.com) (accessed 26 March 2023).
[9] Ibid.
[10] "The Bizarre and Terrifying Case of the "Deepfake" Video that Helped Bring an African Nation to the Brink" *MotherJones*, 15 March 2019, available at https://www.motherjones.com/politics/2019/03/deepfake-gabon-ali-bongo/ (accessed 26 March 2023).

This is a follow-up to the rumours of the health complications surrounding the then country's President, Ali Bongo, whose consistent absence from public functions for several months triggered concern amongst the people of Gabon. Although Deeptrace Labs thinks the video of President Ali Bongo is real, the people of Gabon, especially the political opponent of the president, think otherwise. On a similar note, in June 2019, a scandal emerged surrounding a sex tape allegedly featuring the Malaysian Minister of Economic Affairs, Mr. Azmin Ali, and a rival minister's male aid.[11] The aid claimed the video was real, however, Mr Ali disclaimed the truthfulness of the video and went on to add that the video "was a realistic deepfake" that was solely intended to destroy his political career by his opponents.[12] Although Mr Ali vehemently rejected the claim that the video was real, experts concluded that there was nothing in the video to suggest that it was deepfake or manipulated. This again reveals the political upheavals which deepfakes portends.

Just like the social and political sphere, the judicial system is no exception to the dangers of deepfakes. While the impact of deepfakes is felt across the board in the legal system, of particular interest is its consequences on evidence presented at trial under the Federal Rules of Evidence. Admissibility of evidence is primarily governed by the Federal Rules of Evidence in the federal courts within the United States.[13] The essence of the rules of evidence on admissibility rule is to ensure that the evidence presented at trial is credible, reliable, and authentic as it pertains to its originality and

[11] "A gay sex tape is threatening to end the political careers of two men in Malaysia" *SBSNews,* 17 June 2019, available at https://www.sbs.com.au/news/article/a-gay-sex-tape-is-threatening-to-end-the-political-careers-of-two-men-in-malaysia/ilgqdaqo5 *(*accessed 26 March 2023).
[12] Ibid.
[13] Section 104 of the Federal Rules of Evidence.

relevance to the material facts for determination before the court.[14] To this end, any evidence presented in court that seeks to establish a fact must be authenticated, which is to say, a particular piece of evidence must be what it is claimed to be in other to count on the reliability and trustworthiness of such piece of evidence.[15]

Having the above in mind, this article seeks to examine primarily the challenges posed by deepfake technology to the Federal Rules of Evidence as it pertains to the authentication of evidence in the United States of America. Although this article examines deepfakes and the way they impact proceedings in court, it is imperative to point out that this examination is narrowly tailored to the authentication of evidence in criminal trials following a challenge of audio-visual or voice-recording evidence. By focusing on the authentication of evidence for the avoidance of admission of deepfakes at criminal trials, it does not suggest that the question of deepfakes does not arise in civil proceedings. There are, of course, problems with deepfakes in civil trials as there has been discussion on the impacts of deepfakes on civil trials elsewhere.[16] Consequently, the cardinal focus of this article is the determination of authentication of a piece of evidence- that is, evidence sought to be tendered but objected to as a deepfake during criminal proceedings. Put differently, the article aims to determine whether a particular piece of evidence is indeed original or deepfake. It equally discusses the determination of the question of means or procedure to arrive at the authentication of the particular piece of evidence and when such a piece of

[14] Sections 102 and 901 of the Federal Rules of Evidence; R. A. Delfino, "Deepfakes on Trial: A Call To Expand the Trial Judge's Gatekeeping Role To Protect Legal Proceedings from Technological Fakery," *Hastings Law Journal,* Volume 74, Issue 2 at 321.
[15] Section 901(a) and (b) of the Federal Rules of Evidence.
[16] R. A. Delfino, *Supra*, note 14, p. 296.

evidence should be authenticated. In addition, it will also consider whose duty it is to authenticate and whether an objection to evidence as deepfake will be limited to certain circumstances or whether it will be a free-for-all practice. The foregoing forms the rubrics of this study. Indeed, the above queries form the gamut of this paper and the legal architectural foundation upon which the structure and the super-structure of this paper are built.

By way of caution, this article argues that courts must be extremely careful in admitting or rejecting evidence at trial due to the threat deepfakes hold today. Importantly, the article raises concern about the perversion of justice in the likely event that deepfake is admitted in evidence. In conclusion, the article suggests ways the courts may deploy to ameliorate the admissions of deepfakes at trial or rejection of original evidence in the belief that it is a deepfake. To that end, this paper is divided into four sections. The first section presents an introductory prelude to the paper. The second part explores the origin and rise of deepfake technology. Here the history and the meaning of deepfake is considered. The third section examines the traditional means of authentication of audio-visual, digital, and scientific images. The fourth section considers the authenticity of evidence under the Federal Rules of Evidence and the emergence of deepfakes. Here the challenges of deepfakes at trial are analysed. Part Five is the concluding part where suggestions are offered on how the court may navigate criminal trials and authentication of evidence to avoid the admission of deepfakes.

## 2.0 THE RISE OF DEEPFAKES

### 2.1. The Concept of Deepfakes

Deepfake Technology is an algorithm for the making of a fake version of an image, voice, or video with the sole aim of depicting an action by someone who never performed such action. Although its central purpose remains the same, writers have created categories or types of deepfakes depending on the objective of the creator. This is rightly captured in the words of Gamage et al., where the authors noted that "deepfake phenomenon must be positioned…and…be examined under the lens of the contexts of its uses (or potential uses), along with its effects."[17]

Following this sequence, deepfakes have come to be classified or typified into information deepfakes and 'DeepNude.' The patterns of use and intent of the makers reveal that information manipulation, on one hand, and fake nudes are top in the making of deepfakes. As Gamage et al noted in their study, "cyber security reports in 2019 predict 96% of all deepfakes to be pornographic."[18] This awful development includes child pornography. Of course, it will be splitting hair to differentiate child pornography from 'pornography.' However, the distinction lies in the danger of deepfake and its non-discriminatory use against children.

Apart from deepnude which is used to manipulate images, deepfakes are also used to create fake information, distorted information, or disinformation. More than deepnude, fake information causes more harm within the political arena. This is, of course, seen

[17] D. Gamage, et al., "Are Deepfakes Concerning? Analysing Conversations of Deepfakes on Reddit and Exploring Societal Implications," (2022), p. 3.
[18] Ibid.

in the political reactions of the public to the videos of Obama, Pelosi, and the President of Gabon. While such videos have been debunked as fake, a shadow of it somehow still clogs most of the people who have seen the videos. Although the videos of Barack Obama, Nancy Pelosi, and Ali Bongo, have been debunked, the public still harbours feelings of distrust. This is of course understandable as we have seen from earlier discussions that it is difficult to differentiate original images from deepfakes. In this sense, deepfakes create tension in the body polity of society.

### 2.2 The History of Deepfakes

The exact origin of deepfake technology is obscure as its use has been long within the film industry. Still, the year 1997 has been noted as the year of a breakthrough in the notion of AI-powered algorithms with the ability to replicate images and videos.[19] Attribution of the concept of deepfake was given to the work of Bregler et al.[20] In their paper, Bregler et al gave an illustration of the use of what they describe as dubbing "using a footage (sic) to create automatically new video of a person mouthing words that she did not speak in the original footage." However, the wave of deepfakes began to gain prominence in the field of computer science in 2014. At its earliest stage of development, deepfake was used in the manipulation of mainly images and voice. Not much was achieved at this time until 2016. In 2016, more researchers began paying attention to the emerging technology to understudy its workings. In that line,

[19] C. Bregler, M. Covell, et al, "Video Rewrite: Driving Visual Speech with Audio," available at, https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/human/bregler-sig97.pdf (accessed 27 March 2023).

[20] Ibid.; Borges, Luis, et. al, "Combining Similarity Features and Deep Representation Learning for Stance Detection in the Context of Checking Fake News" (2019) *ACM Journal of Data and Information Quality*, Vol. 9, No. 4, Article 39.

researchers with the aid of a neural network developed somewhat realistic facial expressions in videos with what was termed Face2Face.[21]

With years of improvement the technology became more sophisticated and by 2017 the term 'Deepfake' was first used following a posting of a video made using the technology algorithm set on Redditor.[22] The video was a face swap of celebrities with the body "of a porn actor" to create a somewhat fake action of sexual activity as though it was performed by celebrities and politicians.[23] Since then, deepfakes have been used against celebrities, politicians, and even private individuals to either coerce them or intimidate them into going against their will. The makers of deepfakes can achieve their aim because it is not easy to detect deepfake videos and images. This paper explores the dangers posed by deepfakes, particularly, the challenges of distinguishing between deepfakes and real images or videos in criminal trials.

## 3.0 TRADITIONAL MEANS OF AUTHENTICATION OF EVIDENCE AT TRIAL

As the heading of this section suggests, this part will explore how evidence was traditionally considered and authenticated during criminal trials before the *tsunami* of deepfakes. The essence of this is to demonstrate the idea that the traditional means of authentication of evidence is no longer tenable to fuel the aim of justice and fact-finding and truth-seeking through the ascertainment of facts relevant to the

[21] R. MD Shohel, N. N. Mohammad, et al., "Deepfake Detection: A Systematic Literature Review," available at IEEE Xplore Full-Text PDF: (accessed 27 March 2023).
[22] L. Lamyanba et al., *Supra*, note 3, p. 132.
[23] L. Lamyanba et al., *Supra,* note 3, p. 132; James Vincent, "Watch Jordan Peele use AI to make Barack Obama deliver a PSA about fake news" *The Verge,* 17 April 2018 https://www.theverge.com/tldr/2018/4/17/17247334/ai-fake-news-video-barack-obama-jordan-peele-buzzfeed (accessed 27 March 2023).

administration of criminal justice. To wit, this section is divided into theories of authentication of evidence and the treatment of audio-visual, digital images, and computer and scientific evidence under the rules of evidence.

### 3.1 Theories of Authentication of Evidence

Theories of evidence suggest principles by which a piece of evidence is subjected to in determining whether such evidence is authentic and admissible in evidence.[24] Given the rules of practice and history of evidence, the authenticity of every piece of evidence comes under the gripe arm of the theories herein below to decipher whether it meets the requirements of the law for admissibility.

#### *3.1.1. The Pictorial Testimony Theory*

The pictorial testimony or witness theory is to the effect that a witness testifies before the court regarding the content of a picture or video, that it is what it claims to be.[25] It involves the fair and accurate description or depiction of the event depicted in the picture or video,[26] both in fairness and accuracy.[27] The rule is best described as the witness with the best knowledge theory since the parameter for a witness to qualify under this theory hinges on nothing other than how well the witness knows of the event. The witness need not be the photographer or video recorder to qualify under this rule of authentication of evidence.[28]

[24] It is, however, important to observe that it is not in all circumstances that original or authentic evidence is admissible. Such evidence may be excluded where the method of its acquisition does not conform with the laid procedure for obtaining such evidence.
[25] In the Matter of the Welfare of L.J.L. (Unpublished), available in Westlaw and cited 2006 WL 3719652.
[26] Ibid.
[27] Ibid.
[28] United States of America v Kenneth Stephens, 202 F.Supp.2d 1361 (April 25, 2002); New York, S. & W. R. Co. v Moore, 105 F. 728, 728 (2d Cir. 1901).

As noted by Delfino, the premise upon which the "pictorial communication theory" is rested is on the overall idea that "any photographic or video evidence" is only but a "graphic portrayal of oral testimony," which, like every category of evidence requires verification by the witness.[29]

### *3.1.2 Silent Evidence Theory*

In the category of theories of authentication of evidence is the silent evidence theory. The philosophy of this theory holds the view that once a photographic image or video certifies the basic requirements as to its source and process of production, that is sufficient to establish a fact, it is thus admitted in evidence without the necessity of calling for the oral testimony of a witness in confirmation of what is contained in the photograph or video.[30] The rationale for this argument is built on the reliance on the process by which the photographic image of video was created.[31] Thus, once a proper foundation is laid establishing trust and reliance on the process through which the picture or video is generated, the need for oral testimony is dispensed with.

The theory of "silent evidence" was more elaborated in the case of *Wise* v *State of Indiana,*[32] where the court, while quoting *Mays* v *State of Indiana*[33] in approval, opined that "[T]here must be a strong showing of authenticity and competency" for the silent evidence theory to apply. The court went on to hold that:

> When automatic cameras are involved, there should be evidence as to how and when the camera was loaded, how frequently the camera was activated, when

[29] R.A. Delfino, *Supra*, note 14, p. 327-328.
[30] *United States* v *Gray*, 531 F.2d 933 (8th Cir. 1976); *Berner* v *State of Indiana* 397 N.E.2nd 1012 (Dec. 12, 1979).
[31] *United States* v *Gray, supra*
[32] 26 N.E.3d 137 (Feb. 13, 2015).
[33] 907 N.E.2d, 131-32.

> the photographs were taken, and the processing and the changing of the film after its removal from the camera.[34]

In judicial cases, silent evidence is properly applied if the factors to silence the need for oral testimony are met. From the cases, these factors are competency, that is, the instrument's efficiency in capturing the evidence. Second, the frequency of use is enough to establish the good condition of the instrument. Third, security of the content- image or video at the time of removal from the instrument. These factors are jointly considered and once met, a photographic image or video may be admitted in evidence without calling for the oral testimony of someone with the knowledge of the process of making the photograph or video.

During the trial, it is either pictorial or silent witness theories that are used to authenticate a piece of photographic or video-recording evidence for purposes of admissibility. How these theories are used in specific types of evidence during the trial will form the basis of the discussion of the next segment of this section. The aim is to x-ray how evidence has been examined in the past years for authentication. In the end, the argument in this segment will be that these methods are no longer viable to sift altered or manipulated creation of deepfake technologies.

### 3.2 Treatment of Audio-visual, Photographs, Digital Images, and Scientific Evidence under the Rules of Evidence

In the discourse of the treatment of the above-listed categories of evidence, it is germane to note that the Federal Rules of Evidence under its authentication provisions does not specifically mention the theories of authentication of evidence in the way they

[34] *Mays* v *State of Indiana*, Ibid.

have been discussed in this paper. The theories have roots in the English common law which was inherited by America and developed as part of its legal system.[35] Still, by way of legal integration, the vestiges of the theories are silently incorporated into the Federal Rules of Evidence and applied by the courts.[36]

### *3.2.1 Audio-visual Evidence*

Audio and video recording was not a problem until it became an issue at trial. This is of course the consequence of advancement in technology, just as with deepfake technology. With the introduction of audio and video recording came the need for adaptation in judicial proceedings - that is to say, the authentication requirement for admissibility. With the evolution of audio and video recording, the courts applied "strict and elaborate" requirements for authentication before any audio or video recording could be accepted as a true and accurate depiction of what is claimed.[37] As noted by Clifford S. Fishman, the "authentication regime" for audio recordings, and by extension video recordings was first formulated by the Georgia Court.[38] In *Steve M. Solomon, Jr.,* v *Edger,*[39] the Georgia court opined that:

> A proper foundation for [the use of a mechanical transcription device] must be laid as follows: (1) It must be shown that the mechanical transcription device was capable of taking testimony. (2) It must be shown that the operator of the device was competent to operate the device. (3) The authenticity and correctness of the recording must be established. (4) It must be shown that changes, additions, or deletions have not been made. (5) The manner of preservation of the record must be shown. (6) Speakers must be identified. (7)

[35] W. Jonathan, *Imagining the Law- Common Law and the Foundations of the American Legal System*, (Norman F. Cantor HarperCollins Publishers: New York, 1998), at p. 35.
[36] Rule 901 of the Federal Rules of Evidence.
[37] S. F. Clifford, "Recordings, Transcripts and Translations as Evidence," (2006), Vol. 81 *Washington Law Review*, 473 at 478.
[38] Ibid.
[39] 88 S.E.2nd 167 (Ga. Ct. App. 1955).

> It must be shown that the testimony elicited was freely and voluntarily made, without any kind of duress.

These optimum requirements will later be adopted with approval in the case of *United States* v *Mckeever.*[40] These requirements, upon the findings of the court in *Mckeever,* became the canon of interpretation in so far as authentication of audio and video recordings were concerned. These canons of interpretation for the authentication of evidence remained the standard of practice until they were codified under the aegis of Federal Rules of Evidence.

The Federal Rules of Evidence changed the requirement from the long list of requirements as laid down in *Steve M. Solomon, Jr.,* v *Edger* and adopted in *United States* v *McKeever* as a "badge of honour" for authentication of audio-visual recordings.[41] The Federal Rule of Evidence, Rule 901(a), as against the long list of requirements in *McKeever,* only requires a proponent of any audiovisual recording to satisfy that available evidence is "sufficient to support a finding that the item is what the proponent claims it is."[42] Although the overall standard for fulfilling authentication requirements is "sufficient to support a finding", Rule 901(b) provides instances for meeting this condition and there are a total of nine instances.

While the instances in Rule 901(b) are nine in number, of particular importance to this discourse are those enumerated in Rule 901(b) (1, 5, 6, and 9). Accordingly, under Rule 901(b)(1), the testimony of a participant to the recording suffixes in establishing the

[40] 169 F.Supp. 426, 430 ((S.D.N.Y. 1958).
[41] United States v Liberto, 4459219, WL, 1, 4 (USDC, D, Maryland, 2021); United States v Vidacak, 553 F.3d 334, 349 (4th Cir. 2009).
[42] Rule 901(a) Federal Rules of Evidence.

authenticity of an audio-visual recording.[43] The interpretations of the courts have been that it need not be strictly someone who took part in the conversation, for example, a telephone conversation. But one who witnessed such a conversation is enough to give evidence as to its authenticity to meet the requirement of Rule 901(a). Still, it is noteworthy that the other elements as detailed in 901(b) (5, 6, and 9) are taken together, not in the strictest of terms, but showing that they have been met.[44] This approach has been considered a "more liberal approach" to sustaining the standard of authentication under Rule 901(a).[45]

### *3.2.2 Photographs*

This subsection discusses digital images. However, digital images here include photography and X-rays. Writing historically about the sequence of development of photographs and digital images, Delfino observed that "although photographic evidence became a means of persuading the jury in legal proceedings by the end of the 19th century, the courts were initially hesitant to admit photographs into evidence."[46] Delfino further observed that the reason for the courts' hesitation centred around the logic of a witness testifying "on behalf of a photograph."[47] Of course, such hesitations were timely and precautionary as photographic evidence was a recent phenomenon and no rules of evidence were in place at that time to test the accuracy of photographs and ascertain their authenticity.

[43] United States v Brown, 136 F.3d 1176, 1181 (7th Cir. 1998).
[44] United States v White, 116 F.3d 903, 920-21 (D.C. Cir. 1997); Alonzi v People, 597 P.2d 560, 562 (Colo. 1979); United States v Fuller, 441 F.2d 755, 762 (4th Cir. 1971).
[45] S. F. Clifford on Translation and Transcription, *Supra*, note 37, p. 480.
[46] R. A. Delfino on Deepfakes on Trial, *Supra*, note 14, p. 414.
[47] Ibid.

The general scepticism about a photograph and its admissibility in evidence was only allayed after the court in *United States* v *Ortiz,*[48] allowed its admission in evidence. In *United States* v *Ortiz,* the court allowed the admission of an enlarged photographic signature. The court allowed the photograph after the testimony of the photographer "by whom the photographs were made" and "the accuracy of the method pursued" for authenticity "and the results obtained by him" which established "knowledge of the process and the accuracy of the photograph in a land suit.[49] The Supreme Court's decision in *Ortiz* became a prelude to the wave of relaxation of the admissibility of photographs and the standard requirement for their authentication. In that vein, and after several years apart, the court came up with the standard of "showing sufficient to permit a reasonable juror to find the evidence is what its proponent claims."[50] In *Rembert,*[51] the court cited the cases of *Jackson* v *United States*[52] *and United States* v *Smith,*[53] in approval of the "showing sufficient…" as the standard method for the authentication of a photograph for admission in evidence.

The method of photographic authentication is 'flexible' and 'liberal.' From the cases above, the threshold is minimal, and the burden is light on the proponent of a photographic image. Notably, the same principle applies to the authentication of X-ray images as the proof of accuracy and the chain of custody requirements also apply to the process of making X-ray images. Thus, the central element in the determination of

[48] 176 U.S. 422, 431 (Feb. 1900).
[49] United States v Ortiz, Ibid.
[50] United States v Rembert, 863 F.2d 1023, 1026 (D.C. Cir. 1988).
[51] Ibid.
[52] 395 F.2d 519 (D.C. Cir. 1968).
[53] 490 F.2d 789 (D.C. Cir. 1974).

the authenticity of X-ray photographs is the requirement that the instrument by which it was taken and the process through which it was produced were "trustworthy and that they were properly taken."[54]

### *3.2.3 Digital and Computer Images*

The terms 'digital images' and 'computer images' are used interchangeably. A reference to a digital machine, in most instances, would mean a reference to a computer with some degree of sophistication. The rhetoric regarding whether digital is the same as a computer or if there exists some iota of difference is unresolved, and the aim of this paper is not to inundate readers with the dilemma of the similarities or dissimilarities of these terms. Instead, here the analysis will be squared pegged on how images resulting from digital or computer machines are authenticated in court for purposes of admission.[55]

Just as seen in the discussion of photograph and audio-visual evidence, the court also viewed digital images with the same scepticism it viewed photographs and audio-visual recordings at the beginning.[56] The implication of such scepticism was a strict standard of authentication. The strict standard rule for the authentication of digital evidence by the court was later relaxed, and eventually replaced by the general standard of authentication under Rule 901(a) and (b).

Rule 901(a) laid the general condition for authentication. The principle is one of general application in so far as authentication of evidence is concerned. On digital images, the

[54] American Jurisprudence, Second Edition, Section 973.
[55] C. Gutherie & M. Brittan, "The Swinton Six: The Impact of *State* v *Swinton* on the Authentication of Digital Images" (2007) 36 *Stetson Law Review*, 661,
[56] Cunningham v Fair Haven & Westville R.R. Co., 43 A. 1047, 1048-1049 (Conn. 1899).

relevant provision to this discussion is Rule 901(b)((9). The Rule makes provision for the authentication of digital images. The clarity of and the relevance of the Rule is with the clear references to evidence of "a process or system showing that it produces an accurate result."

### *3.2.4 Scientific Evidence*

The practice amongst courts in the determination of authentication of scientific evidence is made through the application of Rule 702 of the Federal Rules of Evidence. This development is warranted by the courts' abandonment of the principle in *Frye* v *United States,*[57] where it was initially decided that the authentication of scientific evidence is sufficient where there is evidence "sufficiently established to have gained general acceptance in the particular field in which it belongs."[58] Thus, before the abandonment of the standard in *Frye,* the condition for establishing the authenticity of scientific evidence was through the oral testimony of an expert in the field of science called in question demonstrating "general acceptance" of the process used in producing the evidence within the community of practitioners.[59]

As noted above, the court made a U-turn to the standard in *Frye*, reasoning that it did not conform to the Rules of Evidence. In its place, the court came up with a new standard to the effect that scientific evidence can be authenticated on "the requirement that an expert's testimony" of "scientific knowledge" establishes a standard of evidentiary reliability."[60] This standard is further elaborated following the

[57] 293 F. 1013 (D.C. Cir. 1923).
[58] Frye v United States, ibid at 1014.
[59] Ibid.
[60] Ibid at 590.

expansion of rules of evidence. From the letters of Rule 702, to authenticate scientific evidence, the proponent must satisfy that the testimony of the expert witness is such that it will aid understanding of the "evidence or to determine the fact," based on "facts or data," that the testimony "is based on a reliable principle and method," and that the testimony is "reliably applied to the principles and methods to the facts of the case."[61]

## 4.0 AUTHENTICATION UNDER THE FEDERAL RULES OF EVIDENCE AND THE CHALLENGES OF DEEPFAKES

As noted in the introduction of this paper, the central focus of the discussion here is on the effect of evidence alleged to be deepfake in criminal proceedings and this section is at the core of this concern. To that end, the unanswered questions that have triggered this paper and the minds of other scholars and writers on the effect of deepfake at trial will be addressed.[62] In doing so, this section is further subdivided into three. The first subsection will consider the current standard of authentication under the Federal Rules of Evidence against the backdrop of deepfake technology. The finding will reveal that the standards and practices available under the authentication rule are not sufficient to tackle the menace of deepfake as they appear in court. The subsequent subsection will dovetail into areas of major concern in the authentication conundrum, which previous researchers have not turned attention to, a more fundamental in the

[61] Rule 702(a – d).
[62] R. A. Delfino, "Deepfake Defense – Exploring the Limits of the Law and Ethical Norms in Prosecuting Legal Proceeding from Lying Lawyers" (2023), *Loyola Law School, Los Angeles Legal Studies Research Paper* No. 2023-02, 84.5 Ohio St. L.J. 1068, available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4355140 (accessed 9 May 2023); M. Agnieszka, "The Threat of Deepfakes in Litigation: Raising the Authentication Bar to Combat Falsehood," 2021, 23 *Vanderbilt Journal of Entertainment and Technology Law,* available at https://scholarship.law.vanderbilt.edu/jetlaw/vol23/iss2/5 (accessed 9 May 2023).

discussion on the dangers of deepfake in the trial of fact. The final subsection will make suggestions on the way forward. Here, suggestions will be made regarding efforts, stakeholders in the administration of criminal justice must make to ensure the curtailment of the challenges of deepfakes.

### 4.1 Procedure under the Federal Rules of Evidence

Rule 901 of the Federal Rule of Evidence, as noted severally in this paper, governs the authentication of evidence. After a turn away from the common law rules through which the two theories on the authentication of evidence discussed above were formulated, the Federal Rules of Evidence came up with a single standard for the authentication of evidence.[63] As a general standard, the rule provides that to authenticate evidence of any nature, what is required to satisfy the requirement of authentication is for the proponent "to produce evidence sufficient to support a finding that the item is what the proponent claims it is."[64]

The rule applies generally so far as the authentication of evidence is concerned. Nevertheless, in Rule 901(b), examples of instances of authentication were provided as a guide and "not a complete list."[65] However, because the scope of this paper is narrow,

[63] Rule 901(a) of the Federal Rules of Evidence.
[64] Ibid.
[65] Rule 901(b): Here, the rule provides the following as list of instances of authentication of evidence (The following are examples only—not a complete list—of evidence that satisfies the requirement: (1) Testimony of a Witness with Knowledge: Testimony that an item is what it is claimed to be. (2) Non-expert Opinion About Handwriting: A non-expert's opinion that handwriting is genuine, based on a familiarity with it that was not acquired for the current litigation. (3) Comparison by an Expert Witness or the Trier of Fact: A comparison with an authenticated specimen by an expert witness or the trier of fact. (4) Distinctive Characteristics and the Likes: The appearance, contents, substance, internal patterns, or other distinctive characteristics of the item, taken together with all the circumstances. (5) Opinion About a Voice: An opinion identifying a person's voice—whether heard firsthand or through mechanical or electronic transmission or recording—based on hearing the voice at any time under circumstances that connect it with the alleged speaker. (6) Evidence About a Telephone Conversation: For a telephone conversation, evidence that a call was made to the number assigned at the time to; a particular person, if circumstances, including self-identification, show that the person answering was the one called; or a

the discussion here will be limited to the instances in subrules (1, 5, 6, and 9) of Rule 901(b); as they are the provisions affected by deepfake, whether as audio, video, or voice recording.

### 4.2 Testimony of Witness with Knowledge

In Rule 901(b)(1), an item sought as evidence can be authenticated by the testimony of a "witness with the knowledge" that the item is exactly what it is called in question.[66] Under this provision, the central requirement for authentication is the knowledge of the testifying witness of the item which is subject to authentication.[67] Rule 901(b)(1) does not provide for the level of knowledge required to qualify as "evidence sufficient" to support the finding of fact. Simply put, what is required to fulfil the condition under 901(b)(1) is ordinary knowledge of the item and the testifier need not have personal knowledge of the item. The witness only requires personal knowledge of the item.

In *Chao* v *Westside Drywall,*[68] while considering whether a witness could testify as to the authenticity of a document, the court held that the witness could not testify as she had "no personal knowledge" of the item which she purports to authenticate through her testimony. The court noted that "the origin, contents, and significance of

particular business, if the call was made to a business and the call related to business reasonably transacted over the telephone. (7) Evidence About Public Records: Evidence that a document was recorded or filed in a public office as authorised by law or Rule 902 of the Federal Rules of Evidence; or a purported public record or statement is from the office where items of this kind are kept. (8) Evidence About Ancient Documents or Data Compilations: For a document or data compilation, evidence that it; (A) is in a condition that creates no suspicion about its authenticity; (B) was in a place where, if authentic, it would likely be; and (C) is at least 20 years old when offered. (9) Evidence About a Process or System: Evidence describing a process or system and showing that it produces an accurate result. (10) Methods Provided by a Statute or Rule: Any method of authentication or identification allowed by a federal statute, or a rule prescribed by the Supreme Court).

[66] Lorraine v Markel American Insurance Company, 241 F.R.D 534. 538 (D.C. MA, 2007).

[67] Ibid.

[68] 709 F.Supp.2d 1037, 1049.

documents" were not discussed and that "the documents are facially devoid of any identifying information supporting any conclusion about their author."[69] Given the shortcoming, the court concluded that there are no perceived "applicable alternative methods of authentication under" Rules 901 or 902, in the circumstance other than those requiring personal knowledge of the item by the witness.[70]

The item in this case was a document page in a book, sought to be identified and authenticated as genuine. While the rule as it applies might be sufficient in authenticating items of such nature for admissibility, it would be difficult, if not impossible to put a witness with only personal knowledge in the box to authenticate a deepfake. Several reasons will account for this difficulty. For example, evidence as to the production, trustworthiness, and credibility. A detailed discussion of these challenges will be deferred to the next subsection of this section.

### 4.3 Opinion on Voice Recording and Telephone Conversation

The next examples under Rule 901(b) are those concerning voice recording and telephone conversations.[71] To authenticate a voice, the rule requires and stipulates that "an opinion identifying a person's voice" which is heard "whether first-hand or through mechanical or electronic transmission or recording." This is on the basis of "hearing the voice at any time under circumstances that connect it with the alleged speaker." Authentication requirements under the example given in 901(b)(5) are like those of 901(b)(6). The only difference is that while the example under 901(b)(5)

[69] Chao v Westside Drywall, Ibid.
[70] Ibid.
[71] Rule 901(b) (5 – 6) of the Federal Rules of Evidence.

concerns voice identification in general, the later provision in 901(b)(6) speaks of the identification of a telephone conversation.[72]

Rule 901(b)(6) provides that to authenticate the existence of a telephone conversion, the proponent must show "that a call was made to the number assigned at the time to" whether a particular individual or a particular business entity. In the case of an individual, the conversation can further be authenticated "if circumstances, including self-identification, show that the person answering was the one called." In the case of business, it further provides that "if the call was made to a business and the call related to business reasonably transacted over the telephone."[73]

### 4.4 Evidence about a Process or System

The final example under Rule 901(b) is the evidence about a process or system. It provides for evidence about a process or system that can be authenticated by a sufficient showing of "evidence describing a process or system and showing that it produces an accurate result."[74] Writing in *Lorraine,*[75] the court opined that "methods of authentication listed in Rule 901(b) relate for the most part to documents" with relative attention that "[has been] given to… computer print-outs," which were particularly described in "Rule 901(b)(9), which was drafted with recent developments in computer technology in mind."[76]

[72] Ibid.
[73] Ibid.
[74] Rule 901(b)(9) of the Federal Rules of Evidence.
[75] Lorraine v Markel American Insurance Company, supra, note 67.
[76] Ibid.

In *Lorraine,* the court analogised the provisions of Rule 901(b)(7) & (9). The court reasoned that under Rule 901(b)(7), unlike Rule 901(b)(9), "there is no need to show that the computer system producing the public records was reliable or the records accurate."[77] The court proceeded to hold that when it comes to Rule 901(b)(9), it "recognises one method of authentication that is particularly useful in authenticating electronic evidence stored in or generated by computers."[78] Under these circumstances, two conditions were necessary to satisfy the requirements of Rule 901(b)(9). The first condition is "evidence describing a process or system used to produce a result." The other requirement is a "showing that the process or system produces accurate results."[79] Tracing the origin of the rule, the court espouses the idea that the rationale for its existence is for circumstances where the "accuracy of a result is dependent upon a process or system which produces it."

As noted by the court in *Lorraine* above, Rule 901(b)(9) was designed, having the developments in computer technology in mind. Still, it is doubtful if audio-visual images and recordings can conveniently be situated within the confines of "computer-generated" documents and thus allow the application of Rule 901(b)(9) as a method for its authentication. The succeeding subsection will answer this poser. The finding will reveal that authentication of evidence to differentiate original from deepfake cannot be properly carried out under the ambiance of the meaning of authentication of electronic or computer evidence as detailed in Rule 901(b)(9).

[77] Ibid.
[78] Ibid at p.548.
[79] Ibid at p. 549.

## 5.0 LIMITATIONS TO THE RULES OF EVIDENCE AND CHALLENGES OF AUTHENTICATION

Although Rule 901 is the primary rule for the authentication of evidence, other provisions of the Federal Rules of Evidence also regulate the authentication of evidence. Rule 901(b) and Rule 901(10) give credence to this assertion. Categorically, Rule 901(10) states that "any method of authentication or identification allowed by a federal statute, or a rule prescribed by the Supreme Court" may be used for the end of authentication. However, this paper argues that the present methods are inadequate to address the challenges of deepfake. Other scholars share this view.[80]

Given the complex nature of deepfakes, a different path should be taken while authenticating any evidence challenged as deepfakes. According to Delfino, the challenges with authenticating deepfakes include the problem of proof, deepfakes defence, and juror scepticism.[81] However, beyond the issues with deepfake identified by Delfino, more fundamental are the following questions; Firstly, when do we authenticate audio, video, or voice recordings alleged to be deepfakes? In other words, does every allegation that evidence is deepfake automatically warrant the necessity to authenticate?

Certainly, where this is done, the cost implication is that the theory of "deepfake defence" will become a standard of practice, slower than the wheel of justice. Acknowledging that justice is not a one-way traffic, it is then expedient to direct it in a manner that will not only serve the interest of one party, but that of the victim, the

[80] R.A. Delfino on Deepfakes on Trial, *Supra,* note 15.
[81] Ibid p. 340-348.

accused, and the society. Having this in mind, it is suggested that to answer this question, the allegation of deepfake should be limited to matters or facts essential and fundamental to establishing the guilt or innocence of the accused person standing trial.

The second question is, where an allegation of deepfake is raised in the trial of fact, to whom lies the responsibility of authentication? Should this burden be on the proponent of the evidence as provided in Rule 901 or the party alleging deepfakes? If it lies on the proponent of the evidence, will the proponent be trusted enough to impugn his evidence where the truth or otherwise of such evidence will harm his case? Human experience shows that man is a specie of sentiment who would naturally want to protect his interest against an adversary. Only a few manage to suppress this *innate bias*. Considering this human trait, will justice be served to trust the authentication of evidence, the result of which will determine the fate of a given matter? This is unlikely. With this in mind, we then turn to the next question, which is, how do we authenticate?

The authentication as discussed by judicial authorities has been theoretical. At best, it is demonstrative of the truth of a matter. While the procedure used over the years in the authentication of other forms of evidence might suffice, the procedure and theories as enunciated by the courts as the standard of establishing the genuineness of evidence will not work in the same manner with deepfakes. Deepfakes are products of complex algorithms. Although its creation has become common these days due to the availability of algorithms used in creating them, it is by no means suggestive that it is easy to explain. Amongst experts in the field of artificial intelligence, it is general knowledge that deepfakes are not easily detected. The argument then is that to guarantee the authenticity of an audio-visual or voice recording alleged to be deepfakes, neither the

theoretical rules of evidence nor the court's standard is sufficient in establishing the authenticity of a particular piece of evidence.

## 6.0 CONCLUSION

The challenges of deepfakes as observed in the section above go beyond the issue of proof, deepfakes defence, and juror scepticism. It lies more in the actual or real authentication of the alleged evidence. Before then, the question is, should every video, audio, or voice recording be challenged as deepfake and at what point should this challenge be raised? Allowing the occasion that permits every piece of evidence to be challenged as deepfakes is dangerous to the administration of criminal justice. Objection to evidence is not only made by the accused person but also by the prosecution. Therefore, where objection to deepfakes is allowed without limit, the impact on the cost of justice will be enormous. It is therefore suggested that only matters or facts that go to the root of the cause should be allowed to be tested on the allegation of deepfake.

The next point is who bears the responsibility? Given that deepfake is a novel technology that determines the justice of a case, the state should bear the responsibility of determining the authenticity or otherwise of the evidence in a criminal trial. Whereas the proponent should bear the responsibility for the authentication of such piece of evidence.

Finally, the means of authentication of evidence where the allegation of deepfakes is raised should be by way of laboratory examination by an expert in the field. The degree of knowledge required of the expert should be above average in the field of artificial

intelligence technology. Still, to ensure that deepfakes are not admitted, such results should be further compared with tests by a different expert(s). This way, the court is certain of the piece of evidence it admits in the trial of fact as an authentic and accurate representation of what it claims it is.