Decoding Deception

Understanding Human Discrimination Ability in Differentiating Authentic Faces from Deepfake Deceits

Jilani, S.K.; Geradts, Z.; Abubakar, A.

DOI
10.1007/978-3-031-51023-6_39

Publication date
2024

Document Version
Final published version

Published in
Image Analysis and Processing - ICIAP 2023 Workshops

Decoding Deception: Understanding Human Discrimination Ability in Differentiating Authentic Faces from Deepfake Deceits

Shelina Khalid Jilani[1,2(✉)], Zeno Geradts[3,5], and Aliyu Abubakar[4]

[1] University of Bradford, Bradford, UK
[2] INTERPOL, Lyon, France
shelinajilani63@gmail.com
[3] Informatics Institute, University of Amsterdam, Sciencepark 900, 1098 XH Amsterdam, The Netherlands
[4] Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool, UK
[5] Netherlands Forensic Institute, Laan Van Ypenburg 6, 2497 GB Den Haag, The Netherlands

Abstract. Advances in innovative digital technologies present a maturing challenge in differentiating between authentic and manipulated media. The evolution of automated technology has specifically exacerbated this issue, with the emergence of DeepFake content. The degree of sophistication poses potential risks and raise concerns across multiple domains including forensic imagery analysis, especially for Facial Image Comparison (FIC) practitioners. It remains unclear as to whether DeepFake videos can be accurately distinguished from their authentic counterparts, when analysed by domain experts. In response, we present our study where two participant cohorts (FIC practitioners and novice subjects) were shown eleven videos (6 authentic videos and 5 DeepFake videos) and asked to make judgments about the authenticity of the faces. The research findings indicate that when distinguishing between DeepFake and authentic faces, FIC practitioners perform at a similar level to the untrained, novice cohort. Though, statistically, the novice cohort outperformed the practitioners with an overall performance surpassing 70%, relative to the FIC practitioners. This research is still in its infancy stage, yet it is already making significant contributions to the field by facilitating a deeper understanding of how DeepFake content could potentially influence the domain of Forensic Image Identification.

Keywords: DeepFake Detection · Face Identification · Artificial Intelligence · Forensic Practitioners and Deep Learning

1 Introduction

In recent years the proliferation of DeepFake media has emerged as a formidable challenge that poses a significant risk to multiple industries including human society, politics, democracy and forensic science [1, 2, 3]. The term came into existence when an individual known as "deepfakes," posting on Reddit, asserted in late 2017 that they had

G. L. Foresti et al. (Eds.): ICIAP 2023 Workshops, LNCS 14365, pp. 470–481, 2024.
https://doi.org/10.1007/978-3-031-51023-6_39

created a machine learning algorithm capable of superimposing celebrity faces onto adult content videos [4]. Ever since then, the domain of non-existent identities driven by artificial intelligence, has captivated the attention of researchers and the public alike. DeepFake is an umbrella term used for a broad range of synthetic, computer-generated media wherein the features of a target person in an original image or video are altered to resemble the facial characteristics of another individual. Such advanced technology typically produces media that is exceptionally lifelike in appearance, with many researchers reporting 'seeing is not believing' [5, 6, 7]. Concurrently, the use of biometric technology has propelled the use of physiological and/or behavioral attributes of an individual, to assist with person verification. In particular, the growth and ubiquitous nature of facial recognition-based technology has been multifarious, given that faces hold a pivotal position in human communication. A human face can share both verbal and non-verbal cues [8, 9], and the acquisition of face related material from a digital perspective, enables this external structure to hold prime position in the field of computer vision research.

The transformative aspect of DeepFake lies in its extensive reach, complexity, and magnitude of the underlying technology, which qualifies anyone with access to a computer, to create counterfeit videos that are indistinguishable from genuine media [10]. The issue is further fuelled with the availability of open-source software, which allows the public to test the latest technology; introducing them to the world of artificial intelligence, with a 'try before you buy' enticement. Further, the ease by which artificial faces are generated in images and videos is because of the: (*i*) availability of large-scale datasets [11, 12] and, (*ii*) the advancement of deep learning methods which reduce the need for manual editing, streamlining the process [13, 14].

In response to the increasingly sophisticated media, substantial endeavours are being carried out by researchers to understand the underlying processes and levels of accuracy associated with human discrimination ability. In the forensic sector, image falsification is not a novel challenge initiated exclusively by DeepFakes. The act of image manipulation through the means of editing software such as Photoshop, remains prevalent even today and the field of digital forensics has long been tackling this challenge [15]. An underdeveloped domain of study relates to the impact of DeepFake media on human facial perception. Furthermore, the question of whether forensic practitioners outperform inexperienced individuals in discerning manipulated media from genuine content remains an open inquiry.

It is already well documented that DeepFake media has the power to be misconstrued and accepted as authentic by human observers [16, 17, 18]. Human judgment is influenced by a range of factors inclusive of emotions. Recent studies in social psychology suggest that negative emotions have the potential to lower susceptibility to deception [19, 20], which may improve an individual's sensitivity. Anger is also reported to diminish cognitive processing depth by encouraging individuals to rely on stereotypes and pre-existing beliefs [21].

Nevertheless, existing scholarly work in the field of perceptual psychology and visual neuroscience indicates that the human visual system possesses specialised mechanisms designed for the perception of faces [22]. For example, within the Fusiform Gyrus of the human brain, there is a distinct region known as the Fusiform Face Area (FFA), which is dedicated to the processing of facial information. It has been reported that the FFA

exhibits selective activation to faces as opposed to other control stimuli [23]. Regardless of one's stance in the debate surrounding whether facial recognition is an inherent ability, or a skill acquired through experience the consensus is that the processing of face-related information tends to take place holistically for many [24, 25].

To investigate human capabilities in detecting DeepFakes, we created a survey named "Decoding Deception" using the Google Forms. The survey was accessible for anyone with an internet connection and featured the DeepFake videos with the original images sampled from the FakeAVCeleb dataset. Each participant had the opportunity to evaluate the level of difficulty or ease involved in distinguishing between the media. Each participant was asked to rate their level of confidence on a three-point scale (50% likened to someone being *unsure,* 75% suggested a *more than likely* response and 100% equated to *extremely confident),* when assigning their response.

Considering the research signifying the visual processing abilities of humans, it is reasonable to anticipate that the participants would exhibit proficient performance in identifying artificial face manipulations. Our hypothesis suggests that the group of forensic practitioners are expected to outperform, if not significantly outperform the group of novice participants.

1.1 Forms of Facial Manipulation

A photograph of a human face can be divided into two independent attributes as outlined in [26]. Firstly, there is the two-dimensional shape which encompasses the arrangement and contours of face features such as the eyes, nose, and mouth. Secondly, there is the representation of the facial surface which covers coloration, skin, hair, and luminosity and provide indicators of the three-dimensional face shape which can be influenced by lighting conditions. Facial manipulations can be classified into four distinct groups:

- **Entire Face Synthesis** [27]: This form of manipulation technique involves the creation of entirely fabricated facial images which are often achieved through Advanced Generative Adversarial Networks (GANs) architecture, such as StyleGAN [28] and StyleGAN2 [29]. This method of manipulation has reported remarkable outcomes, producing facial images of exceptional quality and realism.
- **Identity Swap** [30]: This manipulation involves the substitution of one person's face in a video with the face of another subject. In general, two approaches are considered, (i) conventional techniques such as FaceSwap [31] and (ii) newer deep learning practices commonly referred to a DeepFakes [32].
- **Face Editing/Retouching:** This involves the modification of facial characteristics such as hair, skin colour, age and the addition of accessories such as eyewear [33].
- **Face Reenactment:** Involves the seamless process of replacing a face in a video sequence whilst keeping the gestures and facial expressions of the target. A popular technique associated this form of face manipulation is NeutralTextures [34].

2 Methodology

DeepFakes have become a dominant form of deception in the realm of digital technology and fabricated media. Utilising advanced deep learning algorithms, particularly Generative Adversarial Networks (GANs), these manipulations generate astonishingly realistic content that can effectively mislead human observers.

2.1 Generative Adversarial Networks (GANs)

The field of Artificial Intelligence has been revolutionised by Generative Adversarial Networks (GANs), which have paved the way for generating highly realistic synthetic data. GANs operate through a competitive framework using a generator and a discriminator neural network, as depicted in Fig. 1. The generator's task is to produce synthetic data, while the discriminator's role is to evaluate and differentiate between authentic and generated samples. The ultimate objective of the generator is to create synthetic data that is indistinguishable from real data, challenging the discriminator's ability to discern between the two [35].

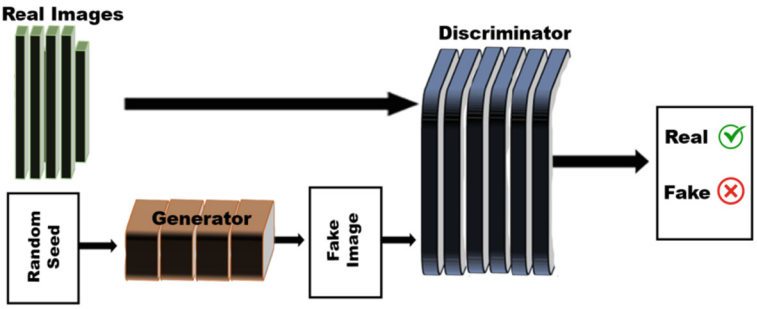

Fig. 1. Schematic representation of a Generative Adversarial Network (GAN) framework.

The generator in a GAN learns a distribution, $Dist_g$ over the data x. This is achieved by creating a mapping function from a prior noise distribution, $Dist_z(z)$ to the data space. The function is defined as $G(z; \theta_g)$. On the other hand, the discriminator, $D(x; \theta_d)$, provides a scalar output signifying the likelihood that x is derived from the training data instead of $Dist_g$. Both the generator and the discriminator are trained simultaneously. The parameters for G are adjusted to minimise $log(1 - D(G(z)))$, while parameters for D are adjusted to minimize $logD(X)$. This training procedure can be likened to a two-player min-max game, where the value function $V(G, D)$ is being optimized. The objective function is defined as:

$$min_G max_D V(D, G) = E_{xp_{data}(x)}\big[logD(x)\big] + E_{zp_z(z)}\big[log(1 - D(G(z)))\big] \quad (1)$$

2.2 Conditional GANs (CGANs)

In a conventional GAN framework (Fig. 1), both the generator and the discriminator operate without any constraints, allowing for unrestricted data generation. However, the lack of specific conditions can lead to inefficiency if the generated data is not required within a specific framework or context. In contrast, the architectural variant, CGANs, presents an option for conditionality in both the generator and the discriminator [36]. These conditions correspond to the class labels of images or other specified properties.

Thus, a traditional GAN model can be transitioned into a CGAN by introducing supplementary conditions to both the generator and the discriminator. For both the generator and the discriminator extra information y is added to the input x.

$$\begin{aligned} &min_G max_D V(D, G) \\ &= E_{xp_{data}(x)}\big[logD(x \vee y)\big] \\ &+E_{zp_z(z)}\big[log(1 - D(G(z \vee y)))\big] \end{aligned} \tag{2}$$

DeepFake videos employ various techniques such as lip-sync and faceswap to manipulate specific facial areas and create authentic, non-existent identities. Lip-sync entails synchronising mouth movements with an audio clip, whereas faceswap involves altering the entire face. Faceswap DeepFakes, employ a combination of two encoder-decoder pairs. The process involves extracting facial features such as the eyes, nose mouth, and ears using an encoder, and then reconstructing the face using a decoder. Typically, to accomplish faceswap, a pair of encoders and decoders are trained on both the source and target images or videos; the duration of the training process directly impacts the level of detail and specificity achieved in the final deepfake video. Once trained, the encoders and decoders are swapped, allowing the original encoder of the source and the decoder of the target to generate a manipulated video.

For lip-sync DeepFakes, a generator coupled with a lip-sync discriminator is employed. The generator learns to synchronise the mouth movements with the audio by using the target individual's data as a reference. By training on the target individual's data, the generator learns to produce realistic lip movements that align with audio. Figure 2 shows an illustrative example.

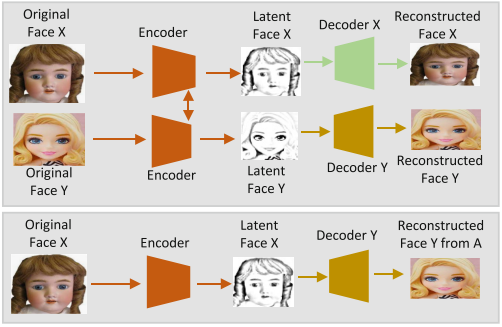

Fig. 2. Graphic to show the process of DeepFake development with encoder-decoder pairs.

The lip-sync DeepFake videos in this study were created using the Wav2lip application. The application employs a specialised model that focuses on synchronising lip movements in a video with an audio clip. The generator component of the model is trained on a range of audio samples and learns to align an individual's mouth movements with the corresponding audio content. To enhance accuracy, a lip-sync discriminator is utilised, assisting the generator in refining its output and rectifying any inconsistencies. By incorporating the lip-sync discriminator during training, the occurrence of artefacts in the final deepfake videos is minimised.

2.3 Faceswap

Faceswap encompasses the substitution of the source's face with another person's face (the target), while maintaining the target's facial expressions. Typically, this procedure entails a sequence of stages. During face detection and alignment, facial landmarks are identified for precise face alignment. Face encoding transforms features of the source and target faces into a numerical format, using deep neural networks. During the face swapping phase, the source face is overlaid onto the target face using techniques such as image warping and/or blending, the goal being to ensure the swapped face blends naturally. Texture blending involves the matching of colours and textures of the source face to the target face. Typically, using techniques such as colour correction and texture mapping. Lastly, during facial expression transfer phase, the swapped face exhibits the target's facial expressions.

2.4 Lip-Sync

Lip-sync techniques strive to achieve harmonisation between the lip movements of an individual in a video and the accompanying audio file. The objective is to ensure precise synchronisation between the spoken words and the corresponding lip movements. Again, similar to the process of faceswap, a series of sequential steps are required for lip-syncing. During audio analysis, the file containing sound information is processed to extract phonetic or timing information. This element of analysis helps to identify the specific sounds that need to be synchronised with the lip movements. Lip Motion Extraction analyses extracted lip shape and movements over time. This is achieved by tracking the movement of specific lip landmarks of the lips. For alignment, the extracted lip movements along with the phonetic/ timing information (from the audio file) are combined. Once the alignment is achieved, the lip movements are animated in a way that corresponds to the audio file. This typically involves warping or morphing the target person's lips to match the desired phonetic shapes or timing.

2.5 Dataset

To support the development of detection software, researchers have curated diverse datasets that serve as valuable resources for research. For this study, the FakeAVCeleb dataset [37] was utilised. FakeAVCeleb is one of the most recent dataset releases; a novel audio-visual DeepFake database which also includes synthesised lip-sync DeepFake audios.

A total of 11 authentic videos of varying image quality, duration and facial viewpoint were selected. Consideration was given to include videos that represented a range of racial backgrounds and maintain equal representation of genders. For each genuine video, a corresponding deepfake version was created utilising both the faceswap and lip-sync techniques, resulting in a total of 11 deepfake videos. The reason for creating a small data sample was to ensure manageability over the quality of the video files, over quantity. In addition, feasibility, to ensure the process wasn't time intensive, especially since our research serves purpose as an exploratory study which we endeavour broadens the discussion of identification abilities across a cohort of individuals.

2.6 Experimental Procedure

The experimental procedure involved presenting a series of videos (authentic and DeepFake), to two participant samples, (i) a group of facial image comparison (FIC) practitioners, from European forensic laboratories, and (ii) novice participants with no experience of working in the field of facial image comparison. The FIC practitioners were selected based on their expertise in the domain of facial image comparison. Each participant was tasked with carefully examining each video and deciding whether it was a DeepFake or not. Participants marked each video, accordingly, providing a clear indication of their judgment. To further evaluate the confidence of their assessments, participants were also asked to rate their level of confidence using a three-point scale.

The confidence scale included three categories: "likely" (50% confidence level), "very likely" (75% confidence level), and "extremely likely" (nearly 100% confidence level). By providing these confidence ratings, the experts were able to express the degree of certainty they had in their judgments regarding the authenticity of the videos. Such an experimental procedure ensured the objectivity and integrity of the assessment process.

3 Results and Discussion

In our experiment, a total of 51 participants were instructed to watch a series of vidoes and identify those of authentic nature and DeepFake. The survey results were analysed with the aim to determine the core elements of this study. Initially, it was hypothesised that given the level of expertise in unfamilair facial identification, Facial Image Comparison (FIC) practitioners would perform exceedingly better compared to the novice cohort. However considering the overall performance between both the participant cohorts, the novice participants marginally outperformed the FIC practitioners when identifying authentic faces in the survey amongst the DeepFakes (Fig. 3).

Upon closer inspection, for the *correct authentic* category, the median score is $>65\%$, and inclined towards the upper quartile of the data distribution which indicates that many of the FIC practitioners performed highly. The results are promising considering that only a small population sample were tested and that the practitioners only work with material consisting of true, authentic identities. Additionally, some if not all the FIC practitioners will not have had the opportunity (prior to this study), to test their discrimination ability using DeepFake material. In contrast, for the *incorrect authentic* category the median is $<35\%$ and closer to the lower quartile of the box. This suggests that several FIC practitioners struggled to make judgements about the authenticity of the videos.

Shifting our attention to the novice population cohort, the median score for the *correct authentic category* reached the boundary of the upper quartile ($>85\%$), indicating a higher discrimination ability relative to FIC practitioners. This may be reflective of the participants who work in the domain of digital forensics, but not facial image comparison.

In Fig. 4, the data suggests that both the FIC practitioners and the novices exhibit a comparable, average performance level in correctly identifying DeepFakes with a median score hovering around 60% for both groups. Likewise, a similar pattern is observed in the incorrect responses for DeepFake identities, with both cohorts exhibiting an average performance level of approximately 40%, except for a few responses. Amongst the FIC practitioners, the highest *correct DF* (DeepFake) distribution is 80%, with the

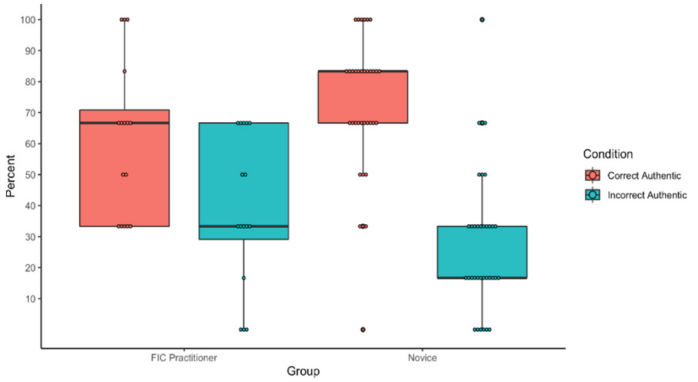

Fig. 3. A Boxplot to illustrate the discrimination abilities between Facial Image Comparison (FIC) Practitioners and novice subjects, when viewing authentic (non-computer-generated) videos with true, human identities.

maximum value depicted by the end of the 'whisker' in the box plot, reaching 100%. Conversely, the upper quartile for the *incorrect DF* distribution is at least 60%, with the 'whisker' extending to 80%. In summary, FIC practitioners generally perform highly in correctly identifying DeepFakes compared to incorrectly identifying them, with the majority performing above chance-level, at 60% for both cases.

In comparison, the novice cohort performance follows a similar pattern. The novice data reports that three-quarter of the incorrect responses is below 45%, indicating a low error rate.

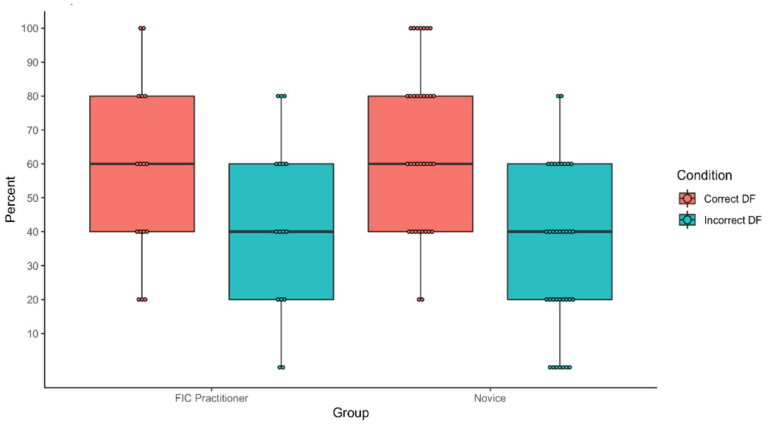

Fig. 4. A Boxplot to illustrate the discrimination abilities between Facial Image Comparison (FIC) Practitioners and novice subjects, when viewing computer-generated DeepFake videos.

Figure 5 (below) depicts the overall performance of both cohorts, and the findings reveal that FIC practitioners do not a perform as highly as their novice counterparts. Generally, the novice participants perform significantly better with a median score reaching

>70%. For the FIC participants, the incorrect distribution is positively skewed with an upper quartile of 55%. This suggests that at least 75% of their incorrect answers fall below the 55% mark. Instead, the incorrect distribution for novice participants shows an upper quartile value of approximately 45%. This suggests that at least three-quarters of the incorrect results from novice participants are lower than 45%. This indicates less accuracy in their results as compared to the FIC participants.

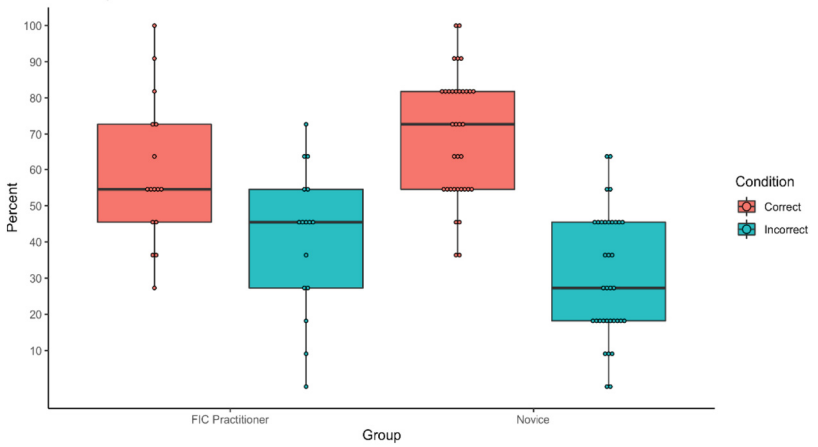

Fig. 5. A Boxplot to illustrate the overall performance between Facial Image Comparison (FIC) Practitioners and novice subjects, when viewing computer-generated DeepFake videos with authentic (non-computer-generated) videos with true, human identities.

Our experimental results appear to contradict current literature, which suggest that novice participants generally have limited ability to identify media manipulations. Our research has demonstrated that novice participants are better at identifying DeepFake faces, compared to FIC practitioners, especially when shown with a short viewing window. Such exceptional performance is to be considered with a caveat that there was a limitation to the number of times the videos could be shown and that too, without any voice information. In addition, there were only two participants in the novice cohort and one in the FIC practitioners' cohort, who performed exceptionally, achieving 100% accuracy across all eleven tested videos. Hence, it is evident that there are underlying variables which affect human performance, and our research highlights the requirement for further study.

Importantly, our research is not free from limitations. The greatest limitation is the participant size, especially for the practitioner cohort. The analysed data came from 51 participants, (16 FIC practitioners and 35 novices). This sample size is not particularly representative, although the results can sufficiently provide exploratory insights. Another potential limitation may be in the process of data collection itself, whilst the participants were asked to rate their level of confidence when making judgements on authenticity, they were not probed about what they were looking for within the videos, when determining whether the face was an authentic or a DeepFake. The judgement rating is not a core

focus of this paper, it was only included to ensure participants responded as honestly and confidently as possible. Such form of qualitative information may have provided a deeper insight into the similarities and/or differences between perception for the tested cohorts.

4 Conclusion

Detecting DeepFakes in this modern world is an increasingly challenging problem on various fronts. Such innovations have a significant impact on online safety, crime, forensic science, and society. In this paper, we have provided preliminary data to show the distinctions in human discrimination ability between an expert (facial image comparison practitioners) and a (non-expert) novice cohort. Our aspiration is that these discoveries will trigger more in-depth studies within the forensic science realm and explain the effects that DeepFakes have on facial image identification.

Acknowledgement. The authors would like to thank Bas Roosenstein, (Forensics Educational Institution, University of Applied Science, Amsterdam) and Dr. Reuben Morton (Open University, UK), for their valuable contributions to this article.

References

1. Borges, L., Martins, B., Calado, P.: Combining similarity features and deep representation learning for stance detection in the context of checking fake news. J. Data Inf. Q. (JDIQ) **11**(3), 1–26 (2019)
2. Dack, S.: Deep fakes, fake news, and what comes next. The Henry M. Jackson School of International Studies (2019)
3. Mansoor, N., Iliev, A.: Artificial intelligence in forensic science. In: Arai, K. (eds.) Advances in Information and Communication. FICC 2023. LNNS, vol. 652, pp. 155–163. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-28073-3_11
4. Bitesize, B.B.C.: deepfakes: what are they and why would I make one? (2019)
5. Maras, M.H., Alexandrou, A.: Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos. The International Journal of Evidence & Proof **23**(3), 255–262 (2019)
6. Cochran, J.D., Napshin, S.A.: Deepfakes: awareness, concerns, and platform accountability. Cyberpsychol. Behav. Soc. Netw. **24**(3), 164–172 (2021)
7. Hancock, J.T., Bailenson, J.N.: The social impact of deepfakes. Cyberpsychol. Behav. Soc. Netw. **24**(3), 149–152 (2021)
8. Jilani, S.K., Ugail, H., Logan, A.: Man vs machine: the ethnic verification of Pakistani and non-Pakistani mouth features. In: 41st ISTANBUL International Conference on "Advances in Science, Engineering & Technology" (IASET-22) (2022)
9. Adyapady, R.R., Annappa, B.: A comprehensive review of facial expression recognition techniques. Multimedia Syst. **29**(1), 73–103 (2023)
10. Fletcher, J.: Deepfakes, artificial intelligence, and some kind of dystopia: the new faces of online post-fact performance. Theatr. J. **70**(4), 455–471 (2018)
11. Narayan, K., Agarwal, H., Thakral, K., Mittal, S., Vatsa, M., Singh, R.: DF-Platter: multi-face heterogeneous Deepfake dataset. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9739–9748 (2023)

12. Korshunov, P., Marcel, S.: Vulnerability assessment and detection of deepfake videos. In: 2019 International Conference on Biometrics (ICB), pp. 1–6. IEEE, June 2019
13. Goodfellow, I., et al.: Generative adversarial networks. Commun. ACM **63**(11), 139–144 (2020)
14. Brophy, E., Wang, Z., She, Q., Ward, T.: Generative adversarial networks in time series: a systematic literature review. ACM Comput. Surv. **55**(10), 1–31 (2023)
15. Battiato, S., Giudice. O., Paratore, A.: Multimedia forensics: discovering the history of multimedia contents. In: Proceedings of the 17th International Conference on Computer Systems and Technologies 2016, pp. 5–16 (2016)
16. Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., Nießner, M.: Faceforensics: A large-scale video dataset for forgery detection in human faces (2018). arXiv preprint arXiv: 1803.09179
17. Vaccari, C., Chadwick, A.: Deepfakes and disinformation: exploring the impact of synthetic political video on deception, uncertainty, and trust in news. Soc. Media+ Soc. **6**(1), 2056305120903408 (2020)
18. Groh, M., Epstein, Z., Firestone, C., Picard, R.: Deepfake detection by human crowds, machines, and machine-informed crowds. Proc. Natl. Acad. Sci. **119**(1), e2110013119 (2022)
19. Forgas, J.P., East, R.: On being happy and gullible: mood effects on skepticism and the detection of deception. J. Exp. Soc. Psychol. **44**(5), 1362–1367 (2008)
20. Brashier, N.M., Marsh, E.J.: Judging truth. Annu. Rev. Psychol. **71**, 499–515 (2020)
21. Clore, G., et al.: Affective feelings as feedback: some cognitive consequences. In: Martin, L.L., Clore, G.L. (eds.) Theories of Mood and Cognition: A User's Handbook. pp. 27–62, L. Erlbaum, 2001
22. Sinha, P., Balas, B., Ostrovsky, Y., Russell, R.: Face recognition by humans: nineteen results all computer vision researchers should know about. Proc. IEEE **94**(11), 1948–1962 (2006)
23. Kanwisher, N., McDermott, J., Chun, M.M.: The fusiform face area: a module in human extrastriate cortex specialized for face perception. J. Neurosci. **17**(11), 4302–4311 (1997)
24. Richler, J.J., Gauthier, I.: A meta-analysis and review of holistic face processing. Psychol. Bull. **140**(5), 1281 (2014)
25. Young, A.W., Burton, A.M.: Are we face experts? Trends Cogn. Sci. **22**(2), 100–110 (2018)
26. Bruce, V., Young, A.W.: Face perception. Psychology Press, Milton Park (2012)
27. Sabel, J., Johansson, F.: On the robustness and generalizability of face synthesis detection methods. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 962–971 (2021)
28. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4401–4410 (2019)
29. Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., Aila, T.: Analyzing and improving the image quality of stylegan. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8110–8119 (2020)
30. Chen, R., Chen, X., Ni, B., Ge, Y.: SimSwap: an efficient framework for high fidelity face swapping. In: ACM Multimedia (2020)
31. Bitouk, D., Kumar, N., Dhillon, S., Belhumeur, P., Nayar, S.K.: Face swapping: automatically replacing faces in photographs. ACM Trans. Graph. **27**(3), 1–8 (2008)
32. Pu, J., et al.: Deepfake videos in the wild: analysis and detection. In: Proceedings of the Web Conference 2021, pp. 981–992, April 2021
33. Gonzalez-Sosa, E., Fierrez, J., Vera-Rodriguez, R., Alonso-Fernandez, F.: Facial soft biometrics for recognition in the wild: recent works, annotation, and COTS evaluation. IEEE Trans. Inf. Forensics Secur. **13**(8), 2001–2014 (2018)
34. Thies, J., Zollhöfer, M., Nießner, M.: Deferred neural rendering: image synthesis using neural textures. ACM Trans. Graph. (TOG) **38**(4), 1–12 (2019)

35. Soni, R., Arora, T.: A review of the techniques of images using GAN. In: Generative Adversarial Networks for Image-to-Image Translation, pp. 99–123 (2021)
36. Mirza, M., Osindero, S.: Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784 (2014)
37. Khalid, H., Tariq, S., Kim, M., Woo, S.S.: FakeAVCeleb: a novel audio-video multimodal deepfake dataset (2021). arXiv preprint arXiv:2108.05080