Deepfake Overview: Generation, Detection, Risks and Opportunities

Anupama Kaushik*[a], *Prabhjot Kaur*[b], *Shivani Gupta* [c]

[a,b,c] *Maharaja Surajmal Institute of Technology, C-4, Janakpuri, New Delhi-110058, India*
anupama@msit.in [a] *, prabhjot.kaur@msit.in* [b]*, shivanig6263@gmail.com* [c*]

Abstract: With the advent of Deep Learning, its impact can be seen on almost every sector of life. One such prominent application of Deep Learning is Deepfake technology. The term 'Deepfake' itself first originated in Reddit in 2017 when users altered certain videos using machine learning algorithms. Hence, Deepfake refers to the creation of fake digital content to make it seem as if someone is doing something they are not. It can be classified into 3 main types- audio, video and text based deepfakes. Deepfake creation has evolved over the years. There are many software and tools that enable a user to create a deepfake. Two highly used tools are Generative Adversarial Networks (GANs) (Malik, Kuribayashi, Abdullahi & Khan, 2022) and autoencoders (Thanh Thi Nguyen et al., 2022). Autoencoders compress input data of original image into latent code containing only the essential features. They then use the decoder of the target face to reconstruct the image using this latent code and achieve face swapping. While this technique works for simple cases, such face-swapping based images are easily discernible to the naked eye. Here, the GANs come in, they can generate extremely realistic images. It too has two parts – the Generator and the Discriminator. Deepfake detection techniques can be categorized into 3 types – deep learning based detection, machine learning based detection and statistical tools based techniques. To ensure the quality of the detection model, select a suitable dataset to train it on must be selected. The dataset must cover a wide range of deepfake content's types, qualities and more. Deepfake is a dual edged sword (Kaur et al., 2024). Currently, deepfake is being used in entertainment and media for visual effects, enhanced user experience and to save the cost of production. It can also be used for simulation training, rehabilitation, therapy and more. And yet, despite the many benefits of deepfake, its immense capacity to cause harm as well cannot be ignored. There have already been several instances where its use has led to much damage in the fields of politics, entertainment, law and more. Therefore, to comprehend the many applications and repercussions of this technology, a thorough understanding of it is needed. This paper seeks to analyse deepfake technology, its applications, risks, detections methods and future opportunities. By reviewing previous research and combining it with new viewpoints and findings, this study serves as an introductory guide to Deepfake technology.

Keywords: Deep Learning, Deepfake detection, Generative Adversarial Networks (GANs), Deepfake Classification, StarGAN v2, DeepSwap, DeepFaceLab, DeepNostalgia, DeepArtEffects, FaceForensics++, Celeb DF v2.0, WildDeepfake, Forgery Net, Diffusion Models

1. Introduction

Deepfakes are digital content that appear realistic but are not. Deepfakes can be classified into three domains - text, audio, and video (Tolosana, Vera-Rodriguez, Fierrez, et al. 2020). Text based deepfake are mostly used to provide falsified reviews on social media platforms. Audio-based deepfake use AI to create believable human speech with the help of voice-changers and text-to-speech converters that are nowadays readily available. Lastly, visual deepfakes involve the creation of deepfake images and videos. Videos may be falsified by the addition of lip-sync, face synthesis and attribute manipulation (altering the hair color, skin color etc.) whereas deepfake images mainly involve face swapping.

The term 'Deepfake' itself is coined after a reddit user's id who first created a deepfake video using machine learning algorithms – they substituted a celebrity's face over a pornographic video. In addition to pornographic usage, other harmful usage of deepfake include fake news, financial fraud and more. However, there are many opportunities offered by it in the fields of art, education, media etc.

Majority of the current research on deepfake is related to deepfake images and videos. Even for deepfake videos, most papers analyze spatial coherence instead of temporal inconsistencies. Further, much of these research have not been tested in the real world and there is a lack of high-quality datasets for deepfake video detection training.

Hence, with the development in the field of artificial intelligence and computer vision, it has become exceedingly difficult to distinguish fake and real content. As such a need has arisen to thoroughly analyze the ethical and legal implications of this emerging technology.

The remainder of this article is organized as follows. In section 2 and 3, the study discusses deepfake creation and deepfake detection respectively in detail. This is followed by looking at the numerous risks affiliated with deepfake in section 4 and the many opportunities it offers in section 5. Finally, section 6 provides the conclusion.

2. Deepfake Creation

There are many technologies and even applications for deepfake creation. They can be classified as traditional generation methods and deep learning based generation methods (Kaur et al., 2024).

Traditional generation methods fall short in comparison to deep learning methods. They are based on computer vision and image processing algorithm and can further be classified into 4 types – entire face synthesis, attribute modification, identity swapping and face replacement/expression swapping. Entire face synthesis uses image warping and morphing to create deepfake (Zhao et al., 2016**;** Berthouzoz et al., 2011). Attribute Modification is used to change a part of an image or video (Berthouzoz et al., 2011). Identity swapping aims to replace the face of one person with another (Kaur et al., 2024). In face replacement/expression swapping, the expression in target image or video is imitated on original content (Xu et al., 2022; Akhtar , 2023).

Deep learning based generations techniques revolutionized deepfake and can create extremely realistic fake content. Some such models are autoencoders, variational autoencoders, generative adversarial networks and diffusion models.

Autoencoders: It has two parts – encoder and decoder. Encoder is used to extract features from digital content and condense it to form a latent code (a filtered code in which unnecessary attributes have been removed) which is later decoded by decoder to recreate the original content. By training its encoder and decoder on the training data set (which should be a mix of real and deepfake videos), autoencoder gains the ability to transfer the feature of one image to another. Hence, it's primarily used for face swapping applications. It however has the severe drawback that the created image may easily be discernible by human eye. (Nguyen et al., 2019; Khalid & Woo,2020)

Generative Adversarial Networks: They too are made of two parts – the Generator (G) and Discriminator (D). The generator is used to create new data samples (fake data), and the discriminator evaluates them against the existing data to draw comparison and conclude whether the content is fake or real. GAN based creation techniques are primarily focused on human faces and are divided into two categories. First is 'Entire face synthesis', in this a face is created from scratch and second is 'Attribute Manipulation' in which a person's hair color or add glasses is changes. They are used to create highly realist synthetic images and require more data and tuning. However, they are limited to a single image domain and do not support multimodal outputs. (Goodfellow et al., 20200; Brock et al., 2018)

Attention should be given to the fact that GAN's undergo adversarial training while autoencoders don't. The generator is constantly improving against the discriminator by contenting to create realistic data that the discriminator cannot identify. Autoencoders lack this feature of competitiveness and thus create blurry images unlike the hyper realistic images generated by GANs. Furthermore, GANs can handle fine grained complex features like skin texture, lighting effect etc. while autoencoders cannot, they perform better for static data and scenarios.

A higher evolved form of GAN is StarGAN v2. It overcomes the limitations of GAN and is designed specifically for multimodal image-to-image translations. Their encoder captures the style information needed to convert the input image into target domain. It too has many limitations – need of large dataset for training, high computational cost etc (Choi et al., 2020; Guarnera et al., 2022).

Variational Autoencoders (VAE): They are an extension of traditional autoencoders in that they model the latent representation as a probability distribution (usually Gaussian). They follow the standard format of encoding and decoding of autoencoders and use neural networks to learn complex, non-linear relationships and probabilistic models. VAE's ability to combine both of their best features makes it a powerful and flexible tool. They have lower fidelity (ability to generate realistic images) but can cover more diversity in contrast to GANs. (Kaur et al., 2024**Error! Reference source not found.**; Child 2020).

Diffusion Models: They have a noise refinement process through which they iterate over the original noisy image and continuously denoise it to match the targe image's data distribution. This helps them create realistic quality images with sharper details and more discriminating features (Ho et al., 2020). They outdo GAN's and VAE's in certain metrics but are not as straightforward to train (Aghasanli et al., 2023 Dhariwal et al., 2021).

Some of the well-known tools used for generating Deepfake content are DeepSwap (Wilpert, 2022), DeepFaceLab (Rankred, 20220), DeepNostalgia and DeepArtEffects. DeepSwap is user friendly and easily accessible, but users face difficulty in unsubscribing from the application. DeepFaceLab is preferred by researchers and students as it enables them to select the machine learning model to be used for generating deepfake content. Its complex UI is, however, incompatible with non-technical users. Further, it does not offer voice replacement. DeepNostalgia distinguishes itself by animating images based on actual human gestures and offering high quality images and photo enhancement functionality (Kidd. et al., 20230). DeepArtEffects is another application available to the public and can be accessed through both PCs and mobile devices. This application is best suited for artists as it uses AI algorithms to replace faces from images with desired quality output (Wilpert, 20220) and can be accessed via both computer and mobiles.

3. Deepfake Detection

In deepfake detection, the first step is the selection of an appropriate dataset to train the model on. This dataset can be a mix of original and generated content or a pre-exiting dataset available to the public online and must cover a wide range of deepfake content's types, qualities and more. Some such datasets used in research are:

- FaceForensics++(Rössler, Cozzolino, Verdoliva et al., 2019): It consists of 1000 original video sequences that have been manipulated with four automated face manipulation methods: Deepfakes, Face2Face, FaceSwap and NeuralTextures. It's been sourced from 977 YouTube videos, ensuring realistic and high-quality forgeries.
- ForgeryNet (He, Gan, Chen et al., 2021): It's a dataset with 2.9 million images and 221,247 images. This dataset in particular stands out due to its scale and size.
- Deepfake Detection Challenge (Dolhansky, Howes,, Pflaum et al., 2019): This dataset contains a mix of fake and original videos, along with content generated by multiple GAN based models. It comprises of over 100,000 clips featuring 3,426 paid actors.
- Celeb DF v2.0 (Li, Yang, Sun et al., 2020): It contains 590 original videos and 5,629 modified videos from YouTube covering multiple ethnicities, age groups and genders. This data has mainly been used for research and learning purposes.
- WildDeepfake (Zi, Chang, Chen, et al., 2020): It has 7,314 face sequences from a source of 707 deepfake videos. This dataset is highly useful to make powerful models.

The size of the model, nature of project and deployment are some factors that affect the choice of dataset selected. For instance, ForgeryNet and Deepfake detection challenge are best suited for training large deep learning models due to their large size.

As discussed previously, both images and videos can be falsified and although similar both require different detection techniques. They both essentially entail binary classification to differentiate between real and fake data.

While analyzing an image, we look at static properties which are features or attributes of an image that rely on a single frame. Some examples of static properties would be image texture, shape, lighting, contours etc. Image detection requires less computational power and mainly involves face swapping or object insertion.

For deepfake video detection, in addition to static properties, consistency in temporal coherence must also be checked. In temporal coherence, the video is converted into individual frames before feeding it to the detection system (comprising of deep learning model, its testing and training, result determination) and hence, requires more computational power.

Deepfake detection techniques can be categorized into 3 types – deep learning based detection, machine learning based detection and statistical tools based techniques.

Multiple approaches have been published using various deep learning techniques for detection. Most notable ones are:

The work by Tao, Gao, Liao et al., 2017 proposed a "sub-pixel motion compensation" (SPMC) layer in CNN framework to work with FlowNet-S CNN framework and MCT (motion transformer module) to achieve the same. Basically, they organize multiple frames to get results of higher quality. They conveyed that proper frame alignment alongside motion compensation is essential for good results.

Mo et. al., 2018 in their paper applied a spatial high-pass filter on input images before feeding them to CNN to highlight fine details and amplify noise. The CNN model used three groups of max-pooling and convolution layers. Max-pooling helped extract the most prominent feature from the feature maps and thus, Down-sampled them. They used CELEBA-HQ dataset and achieved an accuracy of 99.4% in distinguishing real and fake images. This is a relatively simple approach to deepfake detection.

Güera D & Delp EJ, 2018 used CNN to extract temporal features and used RNN to perform further classification. Through this approach they managed to acquire an accuracy of 97.1% on high quality images.

Kolagati et al. 2022: have combined facial landmark detection with multilayer perceptron – a well-established neural network architecture – and used it to distinguish between real and fake videos. Facial landmark detection is a preprocessing step in feature extraction. Here, feature extraction is done by identifying facial landmarks such as nose, hair, ears, eyes etc. Through this approach they achieved an accuracy of 84% and an AUC score of 0.87.

Coccomini et al, 2022: have created a hybrid model combining the features of Vision Transformers (ViTs) and EfficientNet B0 and then proposed a voting mechanism to get to the result. ViTs are advanced deep learning models that treat images or frames as patches and EfficientNet B0 is a CNN used as a feature extractor. Through this approach, we see an AUC of 0.951 and a F1 score of 88%.

Machine learning based detection techniques make use of conventional ML models like Decision Tree, SVM, Random Forest Classifier etc. These methods can be used for finding small defects in deepfakes but encounter difficulties with advanced generative models. For proper results, however, the ML model must be trained on a diverse dataset. Furthermore, ML based detection provides transparency as it enables the user to understand the specific features that contribute to the model's decision-making process. An example of ML based detection would be the study by McCloskey and Albright (McCloskey, S. & Albright, 20190). They proposed a method to distinguish GAN-generated images using saturation cues. They observed that real images contain underexposed pixels due to light and camera properties while GAN-generated images lack these as the normalization step in GAN removes them. They measured the frequency of saturated and underexposed pixels in an image and used it as a classification feature fed to the SVM. They trained their model on 2 datasets – GAN Crop and GAN Full and achieved an AUC of 0.7. This is a simple and efficient approach for detection.

For Statistical based detection, a quantitative analysis of inconsistencies in digital content is performed to distinguish between authentic and fake. Pixel distribution and color patterns are common features assessed by these approaches.

While deep learning based approaches are in general excellent at detecting complex and realistic deepfake, they are computationally expensive, have slower inference time and require large datasets to function well. In contrast, the machine learning based approaches have less detection accuracy but are faster, less computationally expensive and can be deployed with minimal data. Hence, the choice between them must be made based on available resources, project goals and performance expectations.

All deepfake detection algorithms must be appropriately assessed and evaluated. Some of the metrics that help do this are Accuracy, Precision, F1-score, Recall and ROC-AUC (Receiver Operating Characteristic – Area Under Curve). Accuracy is a measure of overall correctness; precision is the ratio of true positives to detected positives, recall refers to identifying all relevant instances, F1-score gives the harmonic-mean of precision and recall and ROC-AUC shows tradeoffs between true positives and false positives.

4. Affiliated Risks of Deepfake

Deepfake possesses immense potential to cause harm. There have already been several instances where its use has led to much damage in the fields of politics, entertainment, law, entrepreneurship and more. Hence, an understanding of the many ethical and legal issues affiliated with deepfake is important.

In the field of politics, deepfakes have been used to create falsified videos of politicians and influence the masses to affect the elections. Recently, in 2022, a video of Ukrainian President Zelenskyy appeared asking the soldiers to lay down their weapons and surrender, thus, affecting their morale. Another incident would be the spread of fake video of Nancy Pelosi, an American politician 0(Buo, 2020). In it, she appeared intoxicated while mispronouncing her words. Initially, Facebook did not remove the video as it did not match their definition of deepfake. Sinc then, Facebook (now Meta) has updated its policies to more clearly address deepfakes. It must also be noted that the spread of such misinformation is harmful to media integrity.

Several media personalities have complained that their likeness has been imposed on adult content creators. Actress Scarlett Johnson has been a frequent target of this. Further, fake leaks, scandalous videos and recreation of dead celebrities for profit are some of the few evils of deepfake plaguing the entertainment industry.

The judicial system too suffers from its nefarious uses. Creation of realistic fake evidence may potentially affect the court's decision and understanding of the trial. Judiciaries across the world have seen an increase in caseloads, cost money and time due to authentication of evidence. (Buo., 20200; Pfefferkorn., 20200)

In addition to politics and judiciary, deepfakes' negative impact has also been seen on businesses. There are numerous instances wherein deepfakes have been used to defraud firms (Buo, 2020). Symantec, a cybersecurity company, has revealed that three CFO's have been defrauded using deepfake and social engineering (Sjouwerman, 2019). These attacks can lead to severe monetary losses.

This technology has, however, made multiple contributions in the fields of healthcare, entertainment, media, education and more.

By providing information in a compelling way, it has revolutionized the realm of education. For instance, historical figures could deliver lectures making class much more interesting and engaging (Sharma et al., 2024;

Chesney, R., & Citron, 2019). “LumiereNet” (Kim BH, Ganapathi V, 2019) is a system that is trying to enhance the current academic approach. Military and medical training can also be enhanced by facing realistic simulations of what they will be facing later. Virtual avatars created by deepfake can also be used in rehabilitation and therapy of patients.

Deepfakes are also used in entertainment, media and gaming industry to create special effects and improve realism. This also saves, an example would be the de-ageing of Harrison Ford in the Indiana Jones movie (Mok, 2022).

To conclude, deepfake as a technology offers many harms and benefits. It must be noted that though the ethical conundrums surrounding deepfake remain ever present in the form of consent, authenticity and potential harm that can be caused by such false digital content, it also offers many positives to multiple industries. Hence, it’s a dual edged sword which needs to be regulated and monitored to not harm the public.

5. Opportunities of Deepfake

There are still many avenues left unexplored where deepfake can be applied and improvements that can be made in this rising technology.

Some future avenues [Kaur et al. 2024, Sharma et al., 2024] that may be explored are:

- High Computational Complexity: The computational complexity of deepfake generation and detection models is still a lot. Future research may aim to reduce this.
- Hybrid models: They have neither been much explored nor practically used. They have immense scope for providing remarkable classification accuracy for real-time video detection.
- Larger datasets: Access to a larger dataset to researchers may enable them to reproduce the results of an already developed model over different datasets.
- Targeted Marketing: In digital marketing, deepfake can be used for targeted marketing. For instance, creating a simulating video of product for advertising.
- AR-VR industry: Improvement can be made in the existing AR-VR industry. An example would be the development of realistic avatars for social media platforms.
- Strong Benchmark Development: Lack of strong benchmarks for fair comparison between different models. The genuine performance of present future detectors cannot be evaluated properly. As such there is a need to develop a proper baseline and challenging datasets for the same.
- Resistance for Adversarial attacks: There is a need for detection models that can resist attacks in the form of gaussian noise, blurring, compression or deliberate adversarial perturbations. Adversarial training and model ensembling are some of the methods that can be explored for the same.

Some targeted research directions to mitigate deepfake-related risks in critical sectors and benefit the society are:

- Real time surveillance and threat detection: These can be integrated with national defense systems to prevent misinformation spread and identity theft.
- Blockchain based tracing: Using blockchain to trace the deepfake content that appears on media. This will stop the manipulated content from being shared widely. It will be particularly useful in the entertainment and media industries.
- Fraud Detection: Enhancing deepfake detection models for applications, video verifications, fake emails and conference meetings has particular use in corporate world, banking and even legislature.

Development in the above areas will only increase the influence of deepfake in the current world.

6. Conclusion

In conclusion, Deepfake technology is still in its development phase. There are multiple approaches for their development and detection, and it is essential to select the most appropriate method for the task at hand. Selection of a database is also a crucial factor for getting the desired result. Deepfakes possess great potential for both good and harm. To protect the public from its malicious usage and to mitigate its dangers, legislative bodies, researchers and policy makers need to collaborate to form appropriate laws and regulations. It also has many research avenues that have yet to be explored and must be kept in mind while such regulations are being drafted.

This paper has thus presented a brief overview of the technology used currently in deepfake generation and creation, and their advantages and disadvantages. This study has also analyzed the many risks and the future development areas affiliated with deepfake.

Acknowledgements

We acknowledge the role of Maharaja Surajmal Institute of Technology for providing us the environment of learning and research.

References

Malik, A., Kuribayashi, M., Abdullahi, S.M., & Khan, A.N. (2022). DeepFake detection for human face images and videos: A survey. *IEEE Access*, 10, 18757–18775.

Thanh Thi Nguyen, Quoc Viet Hung Nguyen, Dung Tien Nguyen, Duc Thanh Nguyen, Thien Huynh-The, Saeid Nahavandi, Thanh Tam Nguyen, Quoc-Viet Pham, Cuong M. Nguyen (2022) Deep learning for deepfakes creation and detection: A survey. *Computer Vision and Image Understanding*, 223, 103525, doi.org/10.1016/j.cviu.2022.103525.

Kaur, A., Hoshyar, A. N., Saikrishna, V., Firmin, S., & Xia, F. (2024). Deepfake video detection: Challenges and opportunities. *Artificial Intelligence Review* https://link.springer.com/article/10.1007/s10462-024-10810-6

Tolosana, R, Vera-Rodriguez, R., Fierrez, J. et al (2020) Deepfakes and beyond: a survey of face manipulation and fake detection. *Inf Fusion*, 64:131–148. https://doi.org/10.1016/j.infus.2020.06.014

Rössler, A., Cozzolino, D., Verdoliva, L. et al (2019) FaceForensics++: learning to detect manipulated facial images. In: *International Conference on Computer Vision (ICCV)*, https://doi.org/10.1109/iccv.2019. 00009

He, Y., Gan, B., Chen, S. et al (2021) Forgerynet: a versatile benchmark for comprehensive forgery analysis. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp 4360–43 https://doi.org/10.1109/cvpr46437.2021.00434

Dolhansky, B., Howes, R., Pflaum ,B. et al (2019) The deepfake detection challenge (dfdc) preview dataset. *Computer Vision and Pattern Recognition* https://doi.org/10.48550/arXiv.1910.08854

Li, Y., Yang, X., Sun, P. et al (2020c) Celeb-df: a large-scale challenging dataset for deepfake forensics. *In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 3207– 3321. https://doi.org/10.1109/cvpr42600.2020.00327

Zi, B., Chang, M., Chen, J., et al (2020) Wilddeepfake: a challenging real-world dataset for deepfake detection. *In: Proceedings of the 28th ACM* *international conference on multimedia*, pp 2382–2390. https://doi.org/10.1145/3394171.3413769

Tao, X., Gao, H., Liao, R. et al (2017) Detail-revealing deep video super-resolution. *In: Proceedings of the IEEE international conference on computer vision*, pp 4472–4480 https://doi.org/10.1109/iccv.2017. 479

Mo, H., Chen, B., & Luo, W. (2018) Fake faces identification via convolutional neural network. *In: Proceedings of the 6th ACM workshop on information hiding and multimedia security*, pp 43–47 https://doi.org/10. 1145/3206004.3206009

Güera, D. & Delp, E.J. (2018) Deepfake video detection using recurrent neural networks. *In: 15th IEEE international conference on Advanced Video and Signal based Surveillance* (AVSS), IEEE, pp 1–6 https://doi.org/10.1109/avss.2018.8639163

Kolagati, S., Priyadharshini,T. & Rajam, V.M.A. (2022) Exposing deepfakes using a deep multilayer perceptron-convolutional neural network model. *Int J Inf Manage Data Insights*.2(1):100054.https://doi.org/10.1016/j.jjimei.2021.100054

Coccomini, D.A., Messina, N., & Gennaro, C. et al (2022) Combining efficientnet and vision transformers for video deepfake detection. *In: International conference on image analysis and processing*. pp 219–22 https://doi.org/10.1007/978-3-031-06433-3_19

McCloskey, S. & Albright M (2019) Detecting gan-generated imagery using saturation cues. *In: 2019 IEEE International Conference on Image Processing*. pp 4584–4588 https://doi.org/10.1109/icip. 2019.8803661

Wilpert, C. (2022) 7 best deepfake software apps of 2022 (50 Tools Reviewed), content mavericks. Available at: https://contentmavericks.com/best-deepfake-software/ (Accessed: 24 December 2022,30 November 2024).

Kidd. J. & Nieto, McAvoy, E. (2023) Deep nostalgia: remediated memory, algorithmic nostalgia and technological ambivalence.*Convergence.* 29(3):620–640. https://doi.org/10.1177/13548565221149839

Rankred (2022) 8 Best Deepfake Apps and Tools In 2022, Rankred. Available at: https://www.rankred.com/best-deepfake-apps-tools/ (Accessed: 11 December 2023, 30 November 2024)

Sharma, V.K., Garg, R. & Caudron, Q. (2024) A systematic literature review on deepfake detection techniques M*ultimedia Tools and Applications* https://link.springer.com/article/10.1007/s11042-024-19906-1

Xu, F.J., Wang, R., & Huang, Y. et al (2022) Countering malicious deepfakes: survey, battleground, and horizon. *Int J Comput Vis*. https://doi.org/10.1007/s11263-022-01606-8

Zhao, J., Mathieu, M., & LeCun, Y. (2016) Energy-based generative adversarial network. arXiv Preprint. http:// arxiv.org/abs/1609.03126

Berthouzoz, F., Li, W., & Dontcheva, M. et al (2011) A framework for content-adaptive photo manipulation macros: application to face, landscape, and global manipulations. *ACM Trans Graph* . 30,120.

Akhtar, Z. (2023) Deepfakes generation and detection: a short survey. *J Imaging* , 9(1), 18

Khalid, H., & Woo, S.S. (2020) OC-FakeDect: classifying deepfakes using one-class variational autoencoder. *In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. pp 656–657

Goodfellow, I., Pouget-Abadie, J., & Mirza, M. et al (2020) Generative adversarial networks. *Commun ACM* . 63(11), 139–144

Brock, A., Donahue, J., & Simonyan, K. (2018) Large scale GAN training for high fdelity natural image synthesis. *arXiv Preprint* http://arxiv.org/abs/1809.11096

Child, R. (2020) Very deep VAEs generalize autoregressive models and can outperform them on images. *arXiv Preprint* http://arxiv.org/abs/2011.10650

Choi, Y., Uh, Y., Yoo, J.,& Ha, J.W. (2020) StarGAN v2: Diverse Image Synthesis for Multiple Domains. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8188–8197.

Guarnera, Luca et al. (2022) The Face Deepfake Detection Challenge. *Journal of Imaging*, 8(10), 263 https://doi.org/10.3390/jimaging8100263

Ho, J., Jain, A., & Abbeel, P. (2020) Denoising difusion probabilistic models. *Adv Neural Inf Process Syst* 33:6840–6851

Aghasanli, A., Kangin, D., & Angelov P (2023) Interpretable-through-prototypes deepfake detection for difusion models. *In: Proceedings of the IEEE/CVF international conference on computer vision.* pp 467–474

Dhariwal, P., & Nichol, A. (2021) Difusion models beat GANs on image synthesis. *Adv Neural Inf Process Syst* 34:8780–8794

Buo, S.A. (2020) The emerging threats of deepfake attacks and countermeasures. https://doi.org/10.48550/ arXiv.2012.07989

Pfefferkorn, R. (2020) Deepfakes in the courtroom. *Boston University Law Journal*, 29(2), 245-276.

Sjouwerman, S. (2019) The evolution of deepfakes: Fighting the next big threat. Available: https://techbeacon.com/security/evolution-deepfakes-fighting-nextbig-threat. [Accessed 09 November 2020].

Chesney, R., & Citron, D. (2019) Deep fakes: a looming challenge for privacy, democracy, and national security. *Calif L Rev* 107:1753. https://doi.org/10.2139/ssrn.3213954

Mok, A. (2022) Take a look at the digitally de-aged harrison ford in the trailer for the new indiana jones movie. Accessed 24 Dec 2022 https://shorturl.at/bhIU5

Kim BH, Ganapathi V (2019) Lumièrenet: lecture video synthesis from audio. arXiv:1907.02253 https:// doi.org/10.48550/arXiv.1907.02253