International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

A review on Ethical and Legal Challenges of Deepfake Technology

First A. k.pavan kalyan , *Member*, IEEE, Second k.manu Jr., and Third sai reddy, and fourth venket sai

Abstract: Deepfake videos are becoming a significant social concern. These videos are created using artificial intelligence (AI) techniques, particularly deep learning, and they pose a growing challenge for society. Unscrupulous individuals exploit deepfake technology to disseminate false information, including fake images, videos, and audio clips. The rise of convincing fake content poses serious threats to politics, security, and personal privacy. Most methods for detecting deepfake videos rely heavily on data-driven approaches. This survey paper aims to provide a comprehensive analysis of both the generation and detection of deepfake videos. One of its key contributions is classifying the various challenges faced in detecting these deceptive videos. It delves into data-related issues, such as unbalanced datasets and insufficiently labeled training data. Training challenges are also highlighted, particularly the need for substantial computational resources. Additionally, the paper addresses reliability issues, including overconfidence in detection methods and the emergence of new manipulation techniques. The research underscores the prevalence of deep learning-based methods in deepfake detection, despite their computational demands and limitations in generalization. However, it also points out the drawbacks of these methods, such as their inefficiency and generalization challenges. Furthermore, the study critically assesses deepfake datasets, stressing the importance of high- quality datasets to enhance detection methods. It also identifies significant research gaps, paving the way for future investigations into deepfake detection, including the development of robust models for real-time detection.

Index Terms — Deepfake Generative Adversarial Networks (GANs) Artificial Intelligence (AI) Machine Learning (ML) Neural Networks Deep Learning Face Manipulation Deepfake Detection

I. INTRODUCTION

Deepfake images and videos might look real, but they're actually crafted using artificial intelligence algorithms. Spotting this kind of content can be tricky for the human eye since it's been technically manipulated. The term "deepfake" comes from a mix of "deep learning" and "fake" videos, where videos are digitally altered to create hyper-realistic portrayals of people saying and doing things that never actually happened. The process involves aligning the faces of two different individuals, using an autoencoder to capture features from one face (let's call it "face A") and then blending those features with another face (we'll call this one "face B"). The end result is a face that resembles B but doesn't truly reflect their real appearance (Alanazi & Asif 2023). Unfortunately, these facial reconstruction techniquesare often misused, especially for creating adult or explicit content on the black market. Deepfakes depend on neural networks that sift through vast datasets to learn how to imitate human facial features, expressions, and voices, making it incredibly hard for people to tell what's real and what's fake. Plus, you don't need to be an expert to create convincing deepfakes; even non-specialists can whip them up using easily accessible tools like Face2Face and Face Swap. Sadly, deepfakes are often used for harmful purposes, including scams that impersonate the voices of business professionals or are deployed in damaging situations, like politics and other deceptive contexts. Given these challenges, it's crucial to look into detection techniques and effective methods to reduce the potential risks associated with deepfake technology. This review paper aims to dive deep into the production and identification of deepfakes to better understand this technology and clarify its complex aspects. and concerning technology, offering helpful insights to navigate it and guard against its potential downsides.

The first part of this paper looks into how people make deepfakes in the world of deepfake tech. Next, it checks out the different software and apps behind creating deepfakes. After that, it talks about spotting deepfakes, which has two parts: finding fake pictures and fake videos. The fourth part of this paper zeroes in on changing pictures and videos that show human expressions in deepfakes. Then, it

examines how deepfakes affect society and what laws deal with them. To wrap up, the paper gives a rundown of the main findings and takeaways.

II. GENERATING DEEPFAKE

Deepfakes are created with deep neural networks, specifcally through the use of autoencoders (Juefei-Xuet al. 2022). This process includes the training of a neural network to encode and decode a video or image, as illustrated. The encoder is tasked with accepting the original input of an image or video and reducing it into a latent code, keeping the important features while fltering out unwanted details. Next, this latent code is sent to the decoder, which restores the original content using this code (Nguyen et al. 2019). During the production of fabricated content, both genuine and manipulated videos or images are trained together with the autoencoder. The encoder is learned to map real as well as deepfake materials into similar latent representations for both

types. At the same time, the decoder employs these forged latent codes to reconstruct the original input, eventually enabling the creation of extremely realistic deepfake conten.The creation of such deepfake content depends on an array of technologies, such as algorithms such as 3D ResNeXt and 3D ResNet (Alanazi & Asif 2023). Generative adversarial networks (GANs) are a robust family of deep neural networks that are becoming popular for producing deepfake content, for example, fake images and videos (Malik et al., 2022). A standard GAN architecture consists of two primary elements: a generator and a discriminator. The generator creates new samples of data, while the discriminator tests them against actual data to determine authenticity. During training, the generator attempts to deceive the discriminator, which then adjusts to improve its ability to detect fake data. This dynamic, however, is challenged when dealing with small datasets, needing large volumes of data to operate efectively and reliably, as observed by Almars (2021).The availability of manipulated images and videos highlights the need for robust detection methods for distinguishing authentic and fake content. To this end, Yang and colleagues (2022) suggest a system called deepfake network architecture attribution that identifies the specific generator architectures used in the production of fake images. The technique is effective even when applied to sophisticated models that have been retrained on various datasets. Going deeper into deepfake technology, particularly assignment can be addressed at two different levels: the architectural and the model specifc. Two approaches are evaluated in this study: one based on learned features and The principle followed by the deepfake content generation is typically this whereby the deepfake videos and images are comparatively less clear than the actual images and videos used in producing the output. The created content is lesser in resolution, but the ordinary person in first impression may confuse it with the original content. Deepfake integrates characteristics from various sources to produce an output that mimics the original one but there are few significant changes which modify the overall meaning of the image or the video. That feature can be smile, cry, body part, expression or color of skin.,

III. TOOLS AND SOFTWARE

The rapid development of deepfake making apps pushed by their popularity in secret markets, shows why we gotta keep getting better at spotting fakes (Shahzad et al. 2022). There's a bunch of tools out there to create deepfake stuff, and here's a couple for ya. DeepSwap is a big name for stitching up fake content for fun. People dig it 'cause it's super easy to use and you can find it online without a fuss. Lots of folks go for the no-cost version, which works on phones and laptops. This app stands out 'cause it's got two boss tricks up its sleeve. First, it's wicked fast, so you can whip up some fake-yet-realistic stuff super quick (Wilpert 2022). It's all about getting things done fast. Plus, it spits out images that look pretty much like the real deal making it tough to tell the real from the fake at first glance (Rankred 2022). DeepSwap plays by the rules, with a clear no-no on making or spreading dirty deepfakes. It tells users they better not put up, pass around, or send any not-okay stuff (De Silva De Alwis & Careylaw 2023). But not everything's rosy. Some peeps gripe about how it's a pain to cancel the service 'cause quitting seems way too complex. That's why not a whole lot of folks tell their friends to use DeepSwap; some are beefing about the hassle to ditch the app. So, you got just a few peeps saying good things about it to their pals. DeepFace Lab gets a lot of love from students and brainy types for tweaking pictures and videos on computers. It ain't newbie-friendly, but the nerds love its flexibility in picking the tech for the brainy machine stuff (Wilpert

2022). Sure, the setup's pretty straightforward, but it's gold for coders. Not to mention, this software vibes with a range of computers so more people can get in on the action (Rankred 2022).

DeepFace Lab is good at making super realistic videos and photos and it's open for anyone to use. It's got awesome features like making faces look younger in pictures. People who do research and famous folks like models and actors find it super valuable, but it can be tricky for everyday folks who aren't tech wizards. At the start, DeepFace Lab could handle two specific identities when swapping faces (Xu et al. 2022), but they've made it better now. New updates made things easier and now it works with lots more faces (Xu et al. 2022). There are two types of methods: the source-oriented kind looks at the original video's details, and the target-oriented kind fits the new video's features. This super cool tech, along with the thoughtful design that Perov and friends talked about in 2020, makes creating fake videos that look real a breeze, no matter what computer you're using. It's good for making amazing videos and also for finding fake ones, so it's become a must-have for both making movies and for tech experts.

Deep Nostalgia is another deepfake thing that's great at making high-def pics and videos that look almost real. It's great for making old photos look new and lively. Kidd and Nieto McAvoy (2023) said that this tool doesn't just make old photos look better but also makes them move like real people. It's super easy to use and you can share the stuff you make with it with your friends on social media. But some folks are worried about whether it's okay to use it, like for animating pics of people who have passed away or for selling stuff without permission (Kidd & Nieto McAvoy 2023). This whole situation shows how tech can mess with how we remember personal and shared history. It makes us think more about what it means for family trees and how people connect with each other.

Deep Art Effects works on computers and phones, but some people who use it on their phones aren't super happy iPhone users. The version that costs money seems to work better, but not many people like the free version. There are some issues with giving money back and picking out pictures, so it's not the most loved tool for making deepfakes (Wilpert 2022). In Table 1, there's a bunch of info that shows how these tools are different, what they can do, and what might be a bit of a pain to deal with

IV. DEEPFAKE DETECTION

Deepfakes are scaring folks a ton with their threats to our privacy, security, and how we run our democracy stuff. People are throwing around some ideas to spot these deepfakes. First tries were all about picking out weird unnatural stuff in those fake videos. But now, the brainiacs are leaning on deep learning to pick up on special features that scream "deepfake" (Chesney and Citron 2019). It's pretty much a game of "real vs fake," aiming to tell apart legit videos from the phonies. Trouble is, you need tons of both kinds of videos to get your detection machines learning right (de Lima et al. 2020).

Even though fake videos are popping up like crazy, we're running dry on solid ways to check how good we are at catching them. Jumping into the fray, Korshunov and Marcel (2018) whipped up this cool dataset just perfect for putting deepfake-spotting tools to the test. They took 620 video clips pumped out using this thing called FaceSwap-GAN. They grabbed some movies that are up for grabs on the VidTIMIT database jazzed them up into deepfakes with some on-point face moves, and used them as guinea pigs for testing out detection tricks.

Now, get this: even the top-dog face-checking systems that use VGG and Facenet are messing up on nailing deepfakes. Other tactics, like watching if the lips are moving right or if the video just looks too crispy when they use support vector machines (SVMs), are also goofing up a bunch with this brand-spanking-new dataset. This is shouting at us – we gotta cook up some stronger deepfake-fighting moves (Wen, Han, and Jain 2015). Hang tight – Next up, we're gonna dive into the different flavors of ways folks have come up with to catch these deepfakes.

V. FAKE IMAGE DETECTION

Face-swapping tech is pretty handy for things like editing videos creating cool portraits, and keeping folks' identities on the down-low by swapping out faces in pics. Yet, some bad actors use it to sneak into accounts and steal people's IDs (Korshunova et al. 2017). With today's fancy tools, like those brainy convolutional neural networks (CNNs) and crafty generative adversarial networks (GANs), spotting a

fake mug has gotten super tricky, 'cause they nail all the tiny details like where your face sits, the look you're giving, and even the lighting.

To figure out which faces are legit and which ones got a digital makeover, Zhang et al. (2017) whipped out this smart tactic called the “bag-of-words” move. They took a bunch of neat features and fed them to brainiac machines, like support vector machines (SVMs) and multi-layer perceptrons (MLPs). Out of all the messed-with pics, the ones made by GANs are a real head-scratcher, thanks to their top-notch realness and the GAN’s knack for mimicking complex stuff and dishing out pics that are indistinguishable from the real deal.

When it comes to spotting these GAN-crafted fakers, Agarwal and Varshney (2019) treated it like a bit of a guessing game. They got all science-y framing it in a statistician’s playground of info theory and proving who's who. They came up with this thing called the “oracle error” sorta like the tiniest gap between real deal pics and the GAN’s handiwork. They figured that the worse the GAN gets the bigger this gap gets making it easier to spot the faults in those sneaky deepfakes. This gets extra important when you're dealing with super sharp high-res images where the GANs roll up their sleeves and get down to making fakes that are crazy tough to spot (Nguyen et al. 2019)

VI. 7 ALTERING IMAGES AND VIDEOS WITH HUMAN

Altering static images is typically easier than dealing with moving images. However, altering videos with human expressions is a significant challenge in the field of deepfake content manipulation. Each individual has their own way of expressing themselves, and when coupled with their facial features, it results in distinctive visual outcomes. Deepfake videos, as defined by Groh et al. in 2021, are usually developed from publicly available datasets where human faces tend look as if they lack any significant expressions, looking like lifeless puppets. To overcome this limitation, newer deepfake technologies have emerged, focusing on the modification of a broad array of motions, such as facial and body gestures and expressions. Machine learning is applied to mimic human actions like walking, talking, smiling, crying and frowning. These models are subsequently applied to substitute the original identity. It should be noted that videos with less expressions and shorter durations is easier than those that contain complex expressions, several variations and longer durations. Sophisticated algorithms use elements of psychology, probability, kinematics, inverse kinematics and physics to detect deepfakes by examining the temporal aspects of videos. For deepfake detection in the field, neural network algorithms that focus on facial localization, e.g., CNNs, have proved exceptional accuracy. Their emphasis is on facial positioning rather than continuous emotional speech and expressions, as explained by Groh et al. in 2021.The identification of deepfake manipulation is a process that entails a detailed inspection of specific facial areas instead of the entire image. Algorithms use fusion techniques to detect changes by comparing these areas with a large training dataset that encompasses facial features across different populations. Multiple attributes, including facial expression, hair and eyes, are used as random markers to measure changes. Even minor distortions in facial areas, which are imperceptible to humans,
can have a significant effect on the final image. Algorithms are committed to closely tracking these chosen areas for accurate detection, as emphasized in the research of Tolosana etc al. in 2022 and Guarnera et al. in 2022.Deepfake content detection is more than just concentrating on the person being portrayed, it also involves taking into account background and scene features. Algorithmsare programmed to detect changes in scenes, starting with simple backgrounds and increasingly tackling more intricate situations. Scene element rotations and domain experts' insights facilitate the identification of important attributes specific to specific situations. Identifying changes in such features enables algorithms to classify deepfake images according to the identified changes, as Choras et al. reported in 2020 and Siegel et al. in 2021.
Data scientists and artificial intelligence specialists are currently researching methods for the identification of forged images and videos by examining both obvious features such as accents and nuance such as lighting conditions. Training data is carefully crafted to emphasize features such as poses, postures, lighting conditions and backgrounds to check for authenticity. The fundamental concepts of lighting physics provide promising potential for identifying deepfakes despite the fact that artifcial intelligence tools continue to develop in this area. Current research focuses on

enhancing deepfake forensics by examining the physics of lighting (Somers 2020). Nirkin et al. 2022) explain that face swapping may result in manipulation of face area that results in aligning a face to a new setting. The same process may be applied in order to retain the scenario and background but swap the face alone. Either way, the individual whose face is employed will be made to appear as a participant in an event in which he did not participate. The detection of this manipulation is achievable by keeping a close look on some indicative manipulations. The context of the face's hair, ears, neck, etc. can be tracked to identify copy-paste or other manipulations. Liu et al. (2021) mention that the consistency of the image becomes different when it gets manipulated. therefore, face swap also provides some inconsistencies that can be identified by implementing the face swap technique. Liu et al. (2021) argue that a forensic specialists must know inconsistencies that result from face swapping because only then they will be in a position to look for the right clues that lead to deepfake detection. This involves grain abnormalities in areas boundaries where face-swapping is suspected. The development of Generative Adversarial Networks has grown concern over the privacy and trust of online users because these networks can create deepfake content that is highly realistic. The GANs make the forged images through the addition of adversarial losses and perceptual losses, whichmakes the forgeries extremely visually appealing. Frame to frame face recognition and face reenactment augmentation make the videos created through GANs more realistic.

Within the various deepfake techniques, face swapping and face morphing are the most recognized, with face morphing being the process of combining faces of two or more people in a single image. It is important to have techniques for the detection of morphed images so that the recognition systems works properly, so morphing attack detection (MAD) is one method that can be used. GANs are also crucial in the fabrication of data as well as in the alteration of images because they produce fake high resolution images that cannot be easily differentiated from the real ones. The use of Deep Convolution Generative Networks is also important for training these types of GANs so that they can produce even more realistic images.Phoneme-viseme mismatches are utilized to identify deepfake videos, where the spoken sound does not match the mouth's shape (Agarwal et al. 2020). These fine but important differences are useful in identifying manipulations, and language experts are often used to identify deepfakes in different languages. Forensic processes that use human expertise are utilized, aided by deep learning algorithms to assist the decision-making process. Attention- based explainable deepfake detection algorithms allow experts to focus their attention on specifc regions in images and videos. Human intuition and cultural context consideration are other factors that help in the detection of deepfakes. Forensic experts exercise a manual approach through manual selection of specifc regions in content, which can then be further processed using software tools in order to improve the accuracy of detection. Deep fake detection forensic technique is employed where human intervention is needed. Silva et al. (2022) explain that forensic algorithms rely on human effort who apply deep learning detection algorithm and assist in decision making on whether content is original or fake. Various forensic methods exist, and Silva et al. (2022) prefer an attention-based explainable deepfake detection algorithm which assists in implementing detection networks for detecting faces and other aspects of images and videos. Humans may decide on what area to neglect, magnify or pay greater attention while identifying deepfake material. There are a number of features of images and videos which can be evaluated in some pretexts. Individuals comprehend their social and cultural pretexts more than machines in most situations. Therefore, human intervention and forensic method are often employed to identify deepfake. Human intuition also plays a part in this method of identification. The areas that are chosen manually by the forensic experts can then be processed with the help of tools and software so that deepfake can be detected correctly final. Face morphing and face swap are two primary methods employed in deepfake to modify images or video in order to create fake content. The major diference between them is face swap process involved substituting the face of one individual in an image or video with another person face, whereas the facmorphing process

involves merging the facial features of more than two individuals in order to construct a new hybrid face. Face morphing is a challenge to the recognition systems; therefore, it is imperative to come up with techniques for identifcation of facial morphing. The

threat of face morphing method in deep fake technologies is on its harmful use. This can be achieved by morphing an actual picture of themselves and a friend andcombining the facial elements to create morphed image as their photo for an ePassport (Dameron 2021). This makes them present themselves as the accomplice and cross the checkpoint without triggering any red fags, even though they are wanted by the police (Dameron 2021). Hence, it is essential to identify the fake images made through this method. Damer et al. (2019) suggested a detection approach known as landmark-based solution by using the live probe image of the face of a possible attacker as another source of information. The authors' idea aims at the facial landmarks in the reference and live probe images. The solution proposed supposes that it is feasible to identify specifc patterns in the facial landmarks' position change in the two images when a morphed reference is employed. Damer et .(2019) describe the workfow of the landmarks-based solution It begins with scanning of the facial landmarks in both reference and probe image to form features vector based on the landmark's location shift. This vector would then be applied to classify reference image as being either a morphing attack or a bona fde image. Damer et al. (2019) provide examples for landmarks shifts for attack and bona fde image pairs, accompanied by a definition of the techniques sed for facial landmark detection. these examples by Damer et al. (2019) of the facial landmarks in bona fde and reference images of the same two subjects, and their corresponding probe images.

VII. DEEPFAKE SOCIAL IMPACT AND LEGISLATION

folks thought of deepfake vids as fun for makers and stars alike. Now, movie studios are big on using deepfake tech to tweak scenes – helps them skip the cost and hassle of do- overs (Uddin Mahmud & Sharmin 2020). But whoa, it didn't take long for that tech to be for nasty stuff, like adult content and blackmail making a serious mess in society. Hancock and Bailenson (2021) got it right when they pointed out that deepfakes shake up how much we can bank on the media. These phony clips mess with our heads, getting us to buy into lies and making it tough to tell what's real. They can twist our memories and plant fake ones. Think about that – believing stuff about folks that never even happened (Hancock and Bailenson 2021).

As tech speeds ahead new ways to pull off crimes pop up. Our laws can't keep up making it crystal clear we need smarter rules to clamp down on cyber-baddies and make them pay for their tricks. The scare factor of deepfakes hit home with the 2018 Rohingya disaster in Myanmar where it looks likemade-up videos played a part (GOV.UK 2019). In Kenya's 2018 vote battle, rumors flew about sickly candidate vids done with deepfakes to mess with folks' minds.

Over in Europe, they've got this AI Act thing to keep things clear-cut so people know when they're dealing with made-up media by AI systems. The Act's got different hoops to jump through depending on the AI risk, all to help people stay sharp and choose (Europarl 2023). The EU's law is all about keeping the lights on with AI and guarding our basic rights against digital deception (EC 2024; Loughran 2024).

Back in the USA, judges get the danger of deepfakes in crime. States like Texas are getting tough with laws focusing on this tech. They've slapped rules on making and spreading deepfakes in vote times. Break these and you could be behind bars for a year or coughing up $4000 (Kigwiru 2022). It's part of a wider move in different places to tackle AI shenanigans and sneakiness. The FCC in the US has put the kibosh on fake robocalls from AI posing as VIPs part of a bigger fight against digital scams (Kan 2024; Yousif 2024). And China ain't playing either – their Cyber folks ban making deepfakes without the green light and say you gotta label AI stuff to keep personal and national security tight (CAC 2022).

So yeah, we're in dire need of tough laws on deepfakes. We gotta hit back at those who make them and think about all the hurt they cause, like freaking people out trashing their good name, or turning elections upside down with lies. Plus, media and government should get cracking on teaching folks how to spot these fakes and keep their cool so we don't all get duped by deepfakes (Alanazi et al. 2024).

VIII. CONCLUSION

Deepfake tech is getting better fast, and that's making folks worry about folks using it for sneaky stuff. We gotta keep the internet safe so laws are being made. Figuring out what's a deepfake is tough, but experts found some clues, like when someone's eye blink looks weird. Fake videos didn't get blinking right at first, but they're catching up. Spotting them fake blinks means getting computers to learn all the different ways people blink in different scenarios. AI is really upping the game in catching these fakes even if the face looks almost right. Those smart algorithms can pick up the tiny things that don't match up, like goofy smiles or blinks. That shows why we need tech on the case, not just people looking at stuff.

Deepfakes can be good or bad, but we gotta make rules to stop the bad stuff. We're talking laws from local to national, and rules on websites that say a big nope to those making mean- spirited fakes. We should also tell everyone what's cool and not cool with deepfakes. Working together is the way to go—if governments, tech companies, and regular folks team up, we can get good at finding and stopping deepfakes. Cybercops

need to level up too, to keep everyone on the internet safe. Keeping the innovations coming and setting up some rules will help us handle the deepfake problems.

there's a whole flow to this deepfake biz—from making 'em spreading 'em on social media, catching 'em, and then figuring out how to keep tabs on them. This whole dance involves making policies getting the word out, and that teamwork I mentioned. We can't forget to make sure we're looping back with updates on catching them so we're always on top of stopping deepfakes from getting around.

IX. REFERENCES

Adadi A (2021) A survey on data-efficient algorithms in big data era. J Big Data 8(1):1–54 Afchar D, Nozick V, Yamagishi J et al (2018) MesoNet: a compact facial video forgery detection network.
In: 2018 IEEE international workshop on information forensics and security (WIFS). IEEE, pp 1–7
Agarwal S, Farid H, Gu Y et al (2019) Protecting world leaders against deep fakes. In: CVPR workshops.
pp 38–45
Agarwal S, Farid H, Fried O et al (2020) Detecting deep-fake videos from phoneme-viseme mismatches. In:
Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. pp
660–661
Aghasanli A, Kangin D, Angelov P (2023) Interpretable-through-prototypes deepf
ake detection for diffusion
models. In: Proceedings of the IEEE/CVF international conference on computer vision. pp 467–474
Akhtar Z (2023) Deepfakes generation and detection: a short survey. J Imaging 9(1):18 Amerini I, Galteri L, Caldelli R et al (2019) Deepfake video detection through optical flow based CNN. In:
Proceedings of the IEEE/CVF international conference on computer vision workshops
Anjum A, Abdullah T, Tariq MF et al (2016) Video stream analysis in clouds: an object detection and classification
framework for high performance video analytics. IEEE Trans Cloud Comput 7(4):1152–1167
Bansal N, Aljrees T, Yadav DP et al (2023) Real- time advanced computational intelligence for deep fake
video detection. Appl Sci 13(5):3095Berthouzoz 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(5):120–1 Brock A, Donahue J, Simonyan K (2018) Large scale GAN training for high fidelity natural image synthesis.
arXiv Preprint http:// arxiv. org/ abs/ 1809. 11096
Brown T, Mann B, Ryder N et al (2020) Language models are few-shot learners. Adv Neural Inf Process Syst 33:1877–1901
Carlini N, Farid H (2020) Evading deepfake-image detectors with white-and black-box attacks. In: Proceedings
of the IEEE/CVF conference on computer vision and pattern recognition workshops. pp 658–659

Chan CCK, Kumar V, Delaney S et al (2020) Combating deepfakes: multi-LSTM and blockchain as proof
of authenticity for digital media. In: 2020 IEEE/ITU

international conference on artificial intelligence
for good (AI4G). IEEE, pp 55–62
Cheng S, Dong Y, Pang T et al (2019) Improving black-box adversarial attacks with a transfer-based prior.
In: Advances in neural information processing systems, vol 32
Child R (2020) Very deep VAEs generalize autoregressive models and can outperform them on images.
arXiv Preprint http:// arxiv. org/ abs/ 2011. 10650 Ciftci UA, Demir I, Yin L (2020) FakeCatcher: detection of synthetic portrait videos using biological signals.
IEEE Trans Pattern Anal Mach Intell. https:// doi. org/ 10. 1109/ TPAMI. 2020. 30092 87
Coccomini DA, Caldelli R, Falchi F et al (2022) Cross-forgery analysis of vision transformers and CNNs for
deepfake image detection. In: Proceedings of the 1st international workshop on multimedia AI against disinformation. pp 52–58
Cozzolino D, Thies J, Rössler A et al (2018) ForensicTransfer: weakly-supervised domain adaptation for
forgery detection. arXiv Preprint http:// arxiv. org/ abs/ 1812. 02510
Dang H, Liu F, Stehouwer J et al (2020) On the detection of digital face manipulation. In: Proceedings of
the IEEE/CVF conference on computer vision and pattern recognition. pp 5781–5790
Dhariwal P, Nichol A (2021) Diffusion models beat GANs on image synthesis. Adv Neural Inf Process Syst
34:8780–8794
Dolhansky B, Bitton J, Pflaum B et al (2020) The deepfake detection challenge (DFDC) dataset. arXiv Preprint
http:// arxiv. org/ abs/ 2006. 07397
Dufour N, Gully A (2019) Contributing data to deepfake detection research. Google AI Blog 1(2):3
Frank J, Eisenhofer T, Schönherr L et al (2020) Leveraging frequency analysis for deep fake image recognition.
In: International conference on machine learning. PMLR, pp 3247–3258
Fung S, Lu X, Zhang C et al (2021) DeepfakeUCL: deepfake detection via unsupervised contrastive learning.
In: 2021 international joint conference on neural networks (IJCNN). pp 1–8. https:// doi. org/ 10.
1109/ IJCNN 52387. 2021. 95340 89
Gambín ÁF, Yazidi A, Vasilakos A et al (2024) Deepfakes: current and future trends. Artif Intell Rev 57(3):64
George AS, George AH (2023) Deepfakes: the evolution of hyper realistic media manipulation. Partn Univers Innov Res Publ 1(2):58–74
Gong LY, Li XJ (2024) A contemporary survey on deepfake detection: datasets, algorithms, and challenges. Electronics 13(3):585
Goodfellow I, Pouget-Abadie J, Mirza M et al (2020) Generative adversarial networks. Commun ACM 63(11):139–144
Güera D, Delp EJ (2018) Deepfake video detection using recurrent neural networks. In: 2018 15th IEEE international conference on advanced video and signal based surveillance (AVSS). IEEE, pp 1–6 Guo B, Ding Y, Yao L et al (2020) The future of
false information detection on social media: new perspectives
and trends. ACM Comput Surv (CSUR) 53(4):1–36 He K, Zhang X, Ren S et al (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 770–778
Ho J, Jain A, Abbeel P (2020) Denoising diffusion probabilistic models. Adv Neural Inf Process Syst 33:6840–6851
Hou M, Wang L, Liu J et al (2021) A3Graph: adversarial attributed autoencoder for graph representation
learning. In: Proceedings of the 36th annual ACM symposium on applied computing. pp 1697–1704
Hulzebosch N, Ibrahimi S, Worring M (2020) Detecting CNN-generated facial images in real-world scenarios.In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. pp 642–643
Hussain S, Neekhara P, Jere M et al (2021) Adversarial deepfakes: evaluating vulnerability of deepfake
detectors to adversarial examples. In: Proceedings of the IEEE/CVF winter conference on applications
of computer vision. pp 3348–3357
Ivanovska M, Struc V (2024) On the vulnerability of deepfake detectors to attacks generated by denoising diffusion models. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp 1051–1060
Ji S, Xu W, Yang M et al (2012) 3D convolutional neural networks for human action recognition. IEEE Trans Pattern Anal Mach Intell 35(1):221–231
Jiang L, Li R, Wu W et al (2020) DeeperForensics-

1.0: a large-scale dataset for real-world face forgery