Partners Universal Innovative Research Publication (PUIRP)
Volume: 01 Issue: 02 | November-December 2023 | www.puirp.com

Deepfakes: The Evolution of Hyper realistic Media Manipulation

Dr.A.Shaji George[1], A.S.Hovan George[2]

[1,2]*Independent Researcher, Chennai, Tamil Nadu, India.*

Abstract - Deepfakes, synthetic media created using artificial intelligence and machine learning techniques, allow for the creation of highly realistic fake videos and audio recordings. As deepfake technology has rapidly advanced in recent years, the potential for its misuse in disinformation campaigns, fraud, and other forms of deception has grown exponentially. This paper explores the current state and trajectory of deepfake technology, emerging safeguards designed to detect deepfakes, and the critical role of education and skepticism in inoculating society against their harms. The paper begins by providing background on what deepfakes are and how they are created using generative adversarial networks. It highlights the rising prevalence of deepfakes and the risks they pose for manipulating opinions, swaying elections, committing financial fraud, and damaging reputations. As deepfake creation tools become more accessible, the number of deepfakes in existence is likely to grow at an accelerated pace. Current deepfake techniques still leave subtle clues that allow detection by the human eye, like jerky movements, lighting shifts, and poor lip syncing. However, rapid improvements in realism and ease of generation mean technological safeguards are essential. The paper discusses digital authentication techniques like blockchain, AI-powered deepfake detection algorithms that identify image artifacts, and intentional video watermarking to disrupt deepfake creation. While technology can help label deepfakes, human discernment remains critical. Through education on how to spot signs of manipulated media, applying skepticism instead of blindly trusting videos and audio, and following cybersecurity best practices, individuals can minimize risks. However, the arms race between deepfake creation and detection means no solution will be foolproof. To conclude, the paper argues that relying solely on technological safeguards is insufficient. Rather, a multipronged societal response combining technological defenses, widespread public awareness, and conscientious skepticism is required to meet the epochal challenge posed by the evolution of deepfakes and other forms of synthetic media. As deepfakes grow more sophisticated, we must dedicate resources toward understanding and exposing manipulated media in order to inoculate ourselves against potentially significant social harms.

Keywords: Deepfakes, Generative Adversarial Networks (GANs), Authentication, Detection, Awareness, Skepticism, Cybersecurity, Disinformation, Arms race, Resilience.

1.INTRODUCTION

1.1 Definition and Background on Deepfakes

Deepfakes refer to hyperrealistic media that is synthetically generated by artificial intelligence. The term deepfake originated in 2017 from a Reddit user named "deepfakes" who pioneered the use of deep learning to swap celebrity faces onto pornographic videos. However, the techniques used in deepfakes build upon a long evolution of media and technology. From the early days of Photoshop to cutting-edge generative adversarial networks (GANs), the ability to manipulate images and video has grown exponentially. As

deepfake technology proliferates, the threats posed by disinformation and media manipulation have become a defining challenge of the digital age.

At a basic level, deepfakes leverage powerful AI systems to analyze and recreate attributes of human faces and voices with a high degree of verisimilitude. GANs consist of two neural networks - a generator and discriminator - that compete and refine the generated media until it is indistinguishable from reality. The generator creates fabricated images or audio while the discriminator tries to identify them as fake. This adversarial training process leads to unprecedented realism. Deepfake algorithms examine source media to learnpeech patterns, facial expressions, skin textures, mouth movements, and more. They can then transpose the face or voice of an existing person onto target media, combining realistic personal attributes and movements. The technical foundations for deepfakes began decades ago but have exploded in sophistication in recent years. As far back as the 1990s, basic machine learning techniques were used to swap faces in static images. However, generating convincing video required algorithms that could model human facial geometry and motion dynamics. In 2018, computer graphics researcher Hao Li pioneered techniques to map faces onto target videos, merging computer vision and CGI. The introduction of GANs supercharged deepfake creation by improving photorealism. Now, open-source tools like FakeApp and DeepFaceLab enable anyone to generate deepfakes with minimal technical knowledge.

Deepfakes have quickly proliferated across the internet. While they gained notoriety for celebrity pornographic videos, their potential harms are much vaster. Experts suggest deepfakes could be used for political disinformation, financial fraud, identity theft, defamation and more. In 2019, a deepfake video of Facebook CEO Mark Zuckerberg circulated as a warning. Political deepfakes have also emerged, including a fake video of former president Obama. As deepfake generation becomes accessible to anyone, the number of deepfakes and their destructive impacts could scale rapidly. A major challenge posed by deepfakes is eroding public trust in media authenticity. As synthesized videos and audio become indistinguishable from reality, people may dismiss all media as fake. This epistemic uncertainty could enable bad actors to dismiss authentic content as deepfakes. Additionally, biases in training data can propagate harmful stereotypes through generated media. Currently, deepfake detection relies on human discernment and forensic analysis by researchers. However, improved synthesis algorithms are quickly surpassing human perceptual capacities.

To confront the growing proliferation of deepfakes, researchers are exploring new detection techniques. Methods like digital watermarking, blockchain verification, and subtle artifact detection may help identify manipulated media. Social media platforms are also grappling with deepfake moderation policies. However, these responses have struggled to keep pace with rapid advances in generative AI. Tackling the malicious uses of deepfakes ultimately requires a comprehensive response engaging technology firms, lawmakers, journalists, and the public. Deepfakes utilize AI synthesis to achieve highly realistic media manipulation. Though nascent, deepfake technology has progressed rapidly from research experiments to widespread use for deception, humor, and entertainment. However, the capabilities of generative algorithms far exceed today's safeguards. As deepfake creation tools become more accessible, the threats posed by media manipulation will require vigilance, education, and technological progress to protect truth and society.

1.2 Rising Prevalence and Risks of Deepfakes

In recent years, the number of deepfake videos circulating online has exploded. While deepfakes were initially limited to fake celebrity pornographic videos, their usage has expanded across various domains.

Researchers estimate there are now over 15,000 deepfake videos online, with numbers rapidly rising as synthesis technology improves and becomes more accessible. The proliferation of deepfakes poses significant risks for manipulation, fraud, and disinformation across many facets of society. Several factors account for the increasing prevalence of deepfakes. First, open-source deepfake generation tools have lowered the barriers for creating manipulated media. Apps like DeepFaceLab, Zao, and FakeApp enable anyone to make deepfakes with minimal technical expertise. Second, wide availability of source training data in online image and video datasets improves deepfake quality and realism. Third, consumer-grade graphics cards allow deep learning models to be trained faster and more efficiently. Finally, online communities like Reddit provide platforms for sharing deepfake media and techniques.

These technical advances coupled with limited oversight have fueled deepfake proliferation. A 2019 study found that 96% of deepfakes online are pornographic, with 99% of those featuring female celebrities without consent. However, non-consensual deepfake pornography represents only a fraction of the potential harms. As deepfakes enter the mainstream, risks of political disinformation, financial fraud, and reputation damage rise substantially. The democratic process faces significant jeopardy from deepfake disinformation campaigns. Political deepfakes spread misinformation faster than rebuttals, making elections vulnerable to manipulation. In 2018, a deepfake video depicted President Obama insulting President Trump. If released without context, it could have created diplomatic turmoil. Adversarial states could use deepfakes to influence foreign elections through hyper-realistic fake videos of leaders making inflammatory remarks.

Likewise, deepfakes enable new forms of criminal and financial fraud. Deepfakes of senior executives could be used to provide false authority in scams targeting employees. Fraudsters might also use deepfaked audio to mimic a victim's voice for identity theft. Such social engineering attacks are far more convincing with synthesized media. The potential for monetary fraud is immense. On an individual level, deepfakes can inflict significant reputational damage. Deepfakes depicting public figures or ordinary citizens in compromising scenarios are difficult to remove from the internet. False accusations or mischaracterizations spread rapidly on social media regardless of debunking efforts. Victims of deepfake reputational attacks face stigma and psychological harm.

While today's deepfakes still have some detectable flaws, rapid improvements in AI synthesis create an inflection point for realistic media manipulation. Developments like Nvidia's Broadcast platform hint at the power of future real-time deepfake generation. To address rising deepfake risks, comprehensive technological, educational, and regulatory safeguards are needed. Situational awareness, media literacy, and technological countermeasures are required as deepfakes proliferate across the online media ecosystem. Failing to increase deepfake resilience could leave societies highly vulnerable to large-scale manipulation. In addition, the prevalence of deepfakes is increasing rapidly as creation tools become available to the public. While deepfakes are often associated with celebrity videos, their evolution enables dangerous new forms of hoax, fraud and slander. As deepfake capabilities approach photorealism, the window for developing societal defenses is closing rapidly. Through vigilance and a combination of technology, law, and education, society can mitigate the most pernicious deepfake risks before they cause significant social harms.

1.3 As Deepfake Technology Rapidly Advances, We Must Utilize a Combination of Technological Safeguards, Education, and Skepticism to Protect Ourselves.

As deepfake technology grows more sophisticated, the threats posed by synthetic media manipulation have reached an inflection point. Deepfakes, powered by generative adversarial networks, now allow for the creation of photorealistic fake videos, audio, and images. While deepfakes emerged only a few years ago, rapid advancements in AI synthesis mean they have outpaced many existing safeguards. This paper argues that addressing the malicious uses of deepfakes will require utilizing technological countermeasures, widespread public awareness, and a healthy skepticism of media authenticity. A multipronged societal response is essential as deepfakes become increasingly accessible tools of deception. Deepfakes leverage cutting-edge machine learning techniques to achieve unprecedented realism. By training neural networks on large datasets of faces and voices, deepfake algorithms can model human speech patterns, facial expressions, skin textures, and subtle mannerisms. This enables realistic media manipulation like transferring a person's face onto an actor's body or synthesizing audio in anyone's voice. While today's deepfakes still have some detectable flaws, rapid improvements in generative AI threaten to surpass human perceptual capacities in the near future.

The growing accessibility of deepfake generation poses threats across many domains. Deepfakes could enable political disinformation at new scales by depicting leaders making inflammatory remarks right before elections. Criminals can perpetrate fraud by impersonating victims with synthesized audio or video. Individual reputations and privacy face violation through non-consensual deepfake pornography and slander. As deepfake media proliferates unchecked, trust in online information could be fundamentally eroded. While the challenges seem daunting, deepfakes remain preventable given a proactive, multifaceted response. Firstly, developing robust technological safeguards provides the foundations for detecting deepfakes and tracing their origins. Emerging authentication methods like digital watermarking, blockchain ledgers, and subtle artifact detection may help label manipulated media. But technology alone is insufficient without public awareness. People must learn to identify telltale visual discrepancies that signal deepfakes. Through widespread media literacy education, individuals can apply critical thinking rather than blindly trusting unverified videos.

Finally, a dose of healthy skepticism is required in today's information ecosystem. Neither individuals nor algorithms can instantly determine a video's authenticity. By maintaining vigilant situational awareness and verifying sensational media, people can avoid knee-jerk reactions to potential deepfakes. A sober, level-headed populace, cognizant of deepfake capabilities, is less vulnerable to falsehoods. This paper argues that protecting society from hostile deepfake usage requires a symbiotic, multipronged response. Both sophisticated technological solutions and societal readiness are required to detect deepfakes and discourage their spread. If we enter an era where seeing is no longer believing, the foundations of trust in digital systems may be shaken beyond repair. Through prompt awareness, innovation, and discernment, we can meet the epochal challenge posed by the proliferation of deepfakes.

2. CURRENT STATE OF DEEPFAKE TECHNOLOGY

2.1 How Deepfakes are Created Through AI and GANs

Deepfakes leverage powerful artificial intelligence systems to synthesize fabricated audio, video, and images. The core techniques used to generate deepfakes originated in academic research on AI-based media synthesis. However, in recent years, these capabilities have proliferated far beyond labs as open-source tools enable the mass production of deepfakes. Understanding how deepfakes are crafted provides insights into their inner workings and vulnerabilities.

The vast majority of modern deepfakes are created using generative adversarial networks, or GANs. Developed in 2014 by Ian Goodfellow, GANs consist of two competing neural networks - a generator and discriminator. The generator creates synthetic media, while the discriminator tries to identify it as fake. By pitting the two networks against each other, GANs are trained through an adversarial process. To begin, the generator network is initialized with random noise. It then tries to create a fake image or video clip from this input data. Meanwhile, the adversary network acts as a detective to classify the media as either real or fake. Both networks receive feedback on their performance and are updated iteratively through backpropagation and gradient descent. Over thousands of training cycles, the generator gradually becomes skilled at creating synthetic media that fools the discriminator.

For facial deepfakes, the generator examines source media of a person to implicitly learn their facial geometry, skin textures, expressions, and speaking style. The discriminator serves as a discerning critic, identifying unnatural artifacts in the faked media. This adversarial dynamic drives the rapid improvement of the generator's outputs. With enough training data, the GAN generator can convincingly transpose an individual's face onto target footage in a photorealistic manner.

Several techniques are used to further refine deepfake media. Facial motion is modeled using computer vision landmarks placed on key face positions. The resulting motion vectors guide how facial expressions change. Source face textures are projected and blended into target videos using person-specific color corrections. Post-processing removes artifacts from warping and out-of-context faces. Voice cloning follows a similar principle using speech synthesis models. The rise of apps like FakeApp, DeepFaceLab, and Zao has made deepfake generation accessible to non-experts. These tools provide user-friendly interfaces and pre-trained models that automate GAN training. Some apps allow deepfake creation simply by providing source images or videos as input. The resulting media exhibits photorealistic facial swaps and vocal imitations primed for dissemination online or in disinformation campaigns.

In addition, generative adversarial networks powered by deep neural networks are the primary sources of modern deepfake technology. Through an adversarial training process, GAN generators implicitly learn how to synthesize realistic media that fools even robust discriminators. The proliferation of open-source tools has made such capabilities available to anyone with a computer. Understanding the techniques behind deepfakes is essential for illuminating their vulnerabilities and developing safeguards against their harms. But as algorithms, computing power, and data continually improve, deepfake technology threatens to progress beyond existing security measures absent a vigorous societal response.

2.2 Improvements in Realism and Ease of Creation

Since deepfakes first emerged in 2017, the realism of synthetic media has advanced rapidly while barriers to deepfake creation have lowered dramatically. Early deepfakes were characterized by visible glitches, distorted facial features, and unrealistic movements. However, generative AI capabilities have improved exponentially thanks to better algorithms, larger datasets, and increased computing power. At the same time, user-friendly deepfake apps have proliferated, putting sophisticated manipulation tools into the hands of the masses.

The first deepfake videos swapped celebrity faces onto pornographic footage using basic machine learning techniques. The results were artificial and blurry. However, by 2018, computer graphics expert Hao Li perfected techniques to map source faces onto target videos and model facial motion, taking

photorealism to the next level. The introduction of StyleGANs, Progressive Growing GANs, and other advances further improved image quality and training stability.

Another inflection point was the release of the DeepFake algorithm by the eponymous Reddit user in 2017. By combining AI face-swap techniques with autoencoders, DeepFake achieved unprecedented realism in mapping faces to target videos. Iterative refinement has led to hyperrealistic results that are difficult for humans to distinguish from original footage without scrutiny.

Equally important are improvements in training stability and efficiency. Techniques like attention mechanisms, normalization, and perceptual loss functions enable faster and more robust model convergence. Using multiple GPUs can dramatically shorten training times. The rise of generative repositories like VGGFace2 and PointRend provide diverse, high-quality datasets to augment deepfake creation.

Together, these technical leaps have translated to tangible boosts in deepfake media quality. State-of-the-art deepfakes exhibit smooth, natural facial movements and seamless integration into target footage. Photorealistic speech synchronization, blinks, and skin/lighting adjustments are common. Whereas early deepfakes could be spotted easily, the latest generation can deceive even expert human reviewers without forensic analysis.

At the same time, deepfake generation has become radically accessible to non-experts. User-friendly apps like FakeApp, DeepFaceLab, and Zao have pre-packaged complex ML pipelines into point-and-click interfaces. Some apps even generate deepfakes automatically from a single source image. These tools leverage open-source GAN implementations and cloud computing resources. Integrations with popular platforms like GitHub, Reddit, and Discord fuel community exchange.

The trajectory of deepfake advancement shows no signs of slowing down. Each day, researchers achieve new milestones in AI-synthesized media, while consumer apps integrate these capabilities. Ongoingtech leaps hint at a future where real-time deepfake generation on video calls becomes commonplace. The window to develop societal defenses before reaching this inflection point is rapidly shrinking as barriers to deep media manipulation disappear.

In addition, deepfakes have achieved immense success in quality and accessibility since their inception. Hyperscale datasets, cutting-edge generative networks, efficient model training, and user-friendly apps have fueled this evolution. While detection technologies and policies lag behind, the pace of change shows the urgent need for comprehensive countermeasures. As generating compelling deepfakes approaches the ease of writing fake news, the potential for harm if unchecked grows exponentially. Maintaining public trust and social cohesion in the face of such threats demands a multipronged response from technology, education, law, ethics and society.

2.3 Limitations That Still Allow Detection by the Human Eye

While deepfake technology has improved tremendously, current algorithms still leave behind subtle clues detectable by human observers. These visible artifacts act as cues to discern real from fake media. However, generative AI is evolving rapidly, and these tells may disappear in more advanced deepfakes. For now, limitations in modeling facial geometry, motion dynamics, and context provide an opportunity for human-based detection before machines surpass human perception.

One giveaway of deepfakes is jerky or unnatural movements. Most algorithms struggle to smoothly model intricate muscular dynamics like eyelids, mouths, and head turns. Slight jitters, distorted expressions, and asymmetric movements may signal manipulated footage. Quick head turns and prominent facial expressions often trip up deepfakes.

Likewise, oddities in eye blinking can indicate fakery. Deepfake algorithms often cannot replicate natural blink patterns and speeds. Faces may not blink at all or exhibit a frenzy of rapid, asynchronous blinking. They also may lack details like wetness, refraction, and shadows on CGI eyes. However, some algorithms now synthesize random, naturally variable blinks.

Facial geometry offers clues too. Mapping faces from different angles can create subtle distortions, especially around side profiles, chins, and noses. Temporary glitches may appear as faces move in and out of shadow. Deepfake models also smooth away natural skin imperfections. Pristine, pore-less skin likely signals fabrication.

Deepfakes also struggle with mouth and speech movements. Synthesized mouths may not synch perfectly with speech sounds or exhibit slight pixelation. Smiles and expressions may appear unnatural. Odd pronunciation of words can give away an AI's limited language mastery. However, high-quality deepfakes now involve manually fine-tuning mouth movements after AI generation.

Lighting and shadows act as another cue. If a face's shadows and tones do not match the target scene, it hints at compositing. Light reflecting in eyes may also be inconsistent with the environment. Flickering ambient illumination as a face moves can reveal unnatural artifacts. But deep learning techniques are making lighting convincingly photorealistic.

Finally, examining the contextual plausibility of a video can reveal deepfakes. Does the scene make logical sense? Do people's reactions fit? Would the participants realistically engage in such behavior? Spotting contextual irregularities requires vigilant critical thinking rather than passive watching.

While today's imperfections allow the trained eye to spot deepfakes, rapid improvements in AI synthesis create an inflection point for evaluating authenticity. Already, state-of-the-art deepfakes surpass untrained human discernment. As algorithms, datasets, and models continue to advance, even forensic experts may struggle to distinguish real from fabricated media. For society to protect truth and trust, technological safeguards and responsible policies must complement human skepticism.

3. EMERGING SAFEGUARDS AGAINST DEEPFAKES

3.1 Digital Authentication Techniques Like Blockchain and Watermarking

As deepfake technology proliferates, researchers are exploring various digital authentication techniques to help identify manipulated media. These include blockchain verification, timestamping, digital signatures, and video watermarking. Such safeguards act as tamper-evident seals to establish provenance and detect manipulation. However, no single solution is foolproof against increasingly sophisticated deepfakes.

Blockchain verification offers one emerging safeguard for certifying media authenticity. The video creator uploads a unique hash of the original footage onto a blockchain ledger. If even one pixel changes, the hash value also changes. To verify, one simply recomputes the hash of the video in question and checks if it matches the blockchain record. A mismatch indicates manipulation. Startups like Amber Video and TruePic are developing end-to-end blockchain architectures for image and video authentication.

A simpler technique is timestamping media in a trusted digital ledger when created. This establishes provenance by permanently recording origins. Several blockchain platforms like KodakOne and Binded allow timestamping to prove prior existence. By comparing media timestamps to distribution dates, altered videos can be identified. However, timestamps alone cannot detect deepfakes that use genuine older media as source material.

Cryptographic digital signatures also offer a mechanism for tamper detection. A unique signature is generated from the original media using the creator's private key. Alterations yield a mismatched signature. Microsoft recently proposed an algorithm called PhotoDNA that extracts robust hashes from media for signature generation. Still, signatures require trusted identity systems and widespread adoption to enable authentication at scale.

Finally, video watermarking intentionally plants subtle artifacts that are disrupted by manipulation. Watermarks exploit how deep learning models perceive images differently than humans. Imperceptible color modulations, noise patterns, and adversarial examples form a fingerprint detectable by algorithms but not visible to viewers. However, watermarks must be tuned carefully - too overt and they compromise visual quality, but too subtle and they may be destroyed by compression.

Despite promising advances, digital authentication remains an arms race against increasingly powerful AI synthesis models. Techniques like compressed deepfakes avoid common image artifacts that expose manipulation. Media provenance also requires comprehensive metadata standards and distributed Ledger infrastructure that do not yet exist. For deepfake detection to stay ahead of creation, continuous research and adaptation of authentication systems are essential.

In summary, emerging digital verification methods exploit tamper-evident signatures to expose deepfakes. Blockchain ledgers provide trusted records while techniques like video watermarking and timestamping can reveal origination history. However, technical limitations and adoption challenges remain. Authentication methods should therefore complement, not replace, education, skepticism, and legal deterrence in a multi-pronged strategy. As deepfakes grow more advanced, no singular solution can address all risks. Only an integrated societal response combining vigilance and safeguards across sectors can defend truth and trust.

3.2 Algorithms That Detect Deepfake Artifacts and Distortions

As deepfakes pose growing threats to truth and trust online, researchers are developing AI systems to detect manipulated media. These algorithms analyze videos for subtle technical artifacts, inconsistencies, and other signals indicative of synthesis. Promising approaches include using neural networks to model facial landmarks, light reflections, pulse signals, and pixel-level anomalies to identify deepfakes. However, generative models are also continually improving, leading to an escalating arms race between deepfake creation and detection. A popular approach trains convolutional neural networks (CNNs) on datasets of real and synthetic media to classify deepfakes. Facebook developed a model called Deepfake Detection Challenge that examines eyes, teeth, facial contours and textures for anomalies. Researchers at MIT trained a CNN to analyze photorealism, head poses, and blending. The winning model in a Google Deepfake Detection competition also relied on CNNs and external data to improve generalization.

Other methods look for inconsistencies and artifacts introduced in generating and rendering deepfakes. One technique models the geometry of facial landmarks to catch warping artifacts from mapping faces. Physical properties like light reflection, refraction, and subsurface scattering also provide clues. Pulse

signals from blood flow can reveal the lack of real cardiovascular dynamics. Horizontal artifact patterns generated during GAN training are another telltale sign of synthesis. Blockchain startup Deeptrace developed AI that targets eye blinking, teeth visibility, facial textures, and boundaries to spot deepfakes. Algorithms from Sensity scan videos for editing artifacts introduced in face-swapping. Startup Dessa leveraged photoplethysmography to extract blood flow signals and used the lack of pulse as evidence of deepfakes.

However, deepfake generation is also evolving rapidly. Distortions and anomalies are becoming less common in high-quality deepfakes as underlying techniques improve. Adversarial training regimes can also teach generative models to avoid leaving detectable forensic artifacts. Some startups are synthesizing realistic simulated heartbeat signals and blood flow dynamics to beat biometric detectors. This escalating arms race means deepfake detection algorithms require constant maintenance and updates. No algorithm today can identify all state-of-the-art deepfakes. Holistic approaches combining multiple detection signals, authentication methods, and human review may prove most resilient. But absent fundamental breakthroughs, debunking deepfakes seems destined to remain an endless game of catch-up. In conclusion, promising deepfake detection algorithms are emerging to identify manipulated media, but have limited shelf-lifes. As generation techniques evolve, new flaws and vulnerabilities open for detection. This cycle is unlikely to end given the dual-use potential of underlying AI systems. While detection algorithms are beneficial, societies cannot rely on them alone without a broader strategy encompassing education, regulation, verification, and resilience against misinformation. Multipronged proactive responses will remain essential to secure truth in the age of deepfakes.

3.3 Insertion of Artifacts to Intentionally Disrupt Deepfake Creation

An emerging defensive technique against deepfakes is to intentionally plant artifacts into authentic media to make it difficult to manipulate. These can include subtle adversarial patterns or noise that destroy information important for high-quality fake generation. Such disruption forces deepfake algorithms to generate obvious flaws, making the forgery detectable. However, practical adoption faces challenges as any perceptual artifacts compromise quality. One method, called DeepFakesON, inserts custom noise tailored to each face to mislead deepfake facial modeling algorithms. The noise exploits blind spots in deep networks to preserve quality while degrading deepfake performance. However, advanced generators can often see past such static noise through adversarial training. Periodically refreshing fingerprints also requires tracking media distribution.

Blockchain startup Truepic developed invisible noise patterns recognized by their detection algorithms. The patterns create a unique signature lost if faces are extracted for deepfakes. But clever editing around problematic face regions can defeat the fingerprinting. Dynamic watermarking using metadata attached to frames shows more promise to avoid extraction attacks. Researchers also experiment with adversarial examples - patterns decipherable by humans but trigger false outputs in deep learning models. One team synthesized videos with frame-to-frame adversarial noise to confuse deepfake generators. These relies on hijacking the idiosyncrasies of specific neural network architectures though. Updating models recovers deepfake quality by ignoring the adversarial signals. Other methods leverage 3D reconstructions of faces. By projecting faces in multiple dimensions, artifacts appear when mapping to a 2D target video. Facebook trained models on such projected faces to detect warping commonly seen in deepfakes today. However, advanced dimensional transformations and rendering can still achieve 2D/3D consistency.

Overall, disrupting deepfakes via artifacts remains an ongoing cat-and-mouse game. Much like malware and spam, the arms race favors attack innovation over static defense. And perceptual quality limits the strength of artifacts used. Both high-fidelity media authentication and resilient architectures are needed to enable trust online. Relying solely on disruption risks unintended consequences if legitimate speech is also inhibited by fingerprinting. Holistic governance and societal resilience may prove more lasting solutions. In summary, inserting artifacts to intentionally sabotage deepfake creation shows some promise but also limitations. Adversarial examples and noise tailored to media fingerprints can degrade deepfake quality and expose manipulation. But generative models continue evolving to see past simplistic tricks. And excessive artifacts impair media quality for legitimate uses. For anti-deepfake techniques to advance beyond a temporary patch, more fundamental innovations securing provenance and enabling authentication will likely prove essential.

4. THE ROLE OF EDUCATION AND SKEPTICISM

4.1 Educating People on How to Spot Signs of Deepfakes

Amid the proliferation of deepfake technology, educating the public on how to spot visual clues of manipulated media serves as a critical line of defense. As synthesis algorithms improve, technological detection of deepfakes grows increasingly difficult. However, the human eye remains remarkably perceptive at noticing subtle behavior clues that reveal fabrication. With proper training, people can rely on their intuitive discernment to identify questionable media as likely deepfakes.

Casual observers lack knowledge of typical deepfake limitations that create detectable artifacts. Through tutorials, courses, and awareness campaigns, both expert and lay audiences can learn which visual cues to look for when assessing media authenticity. This empowers individuals to leverage human perceptual abilities as a first line of defense rather than just blindly accepting unvalidated content.

Educational programs should focus on common deepfake flaws like odd blinking patterns, distorted mouth and speech movements, inconsistent lighting and shadows, skin smoothing, and other artifacts of facial geometries mapped from limited source images. The ability to critically inspect media details, instead of just passively watching, allows anomalous details to become conspicuous.

Mainstream technology companies like Facebook, Microsoft and Google have created online resources and trainings to improve deepfake spotting among everyday users. Non-profit organizations also offer tutorials to build deepfake detection skills for youth and seniors, the most vulnerable demographics. Short online games that allow players to distinguish real from fake videos also create engaging educational experiences.

Formal instruction in academic settings provides even deeper training. Some universities now offer courses focused on living in a post-truth world rife with synthetic media. These build critical thinking skills around discerning media manipulation techniques and weighing evidence sources. Such curricula delivered at scale can significantly improve societal resilience.

However, one-off education has limited impact. Repeated skill application across diverse contexts leads to lasting capabilities. Media literacy campaigns through games, resources and repeated training help individuals stay vigilant of new manipulation techniques. Maintaining public awareness ensures detection skills endure alongside rapidly evolving technologies of deception.

Of course, human discernment has limits. State-of-the-art deepfakes already surpass untrained eyes. But education makes society less vulnerable to manipulation by bad actors. An informed populace, widely inoculated with deepfake skepticism through continuous training, is far more resilient than one easily deceived by the latest hyperrealistic media fabrication.

In summary, public awareness and training in deepfake detection represent powerful lines of defense even as algorithms march ahead. With proper education, people can rely on intuitive skills in critiquing facial details, lighting, movements and context to identify likely manipulations. Widespread training and constant vigilance against evolving media fabrication techniques will remain essential given the pace of progress in AI synthesis.

4.2 Applying Healthy Skepticism Instead of Blind Trust

In an era where seeing is no longer believing, cultivating a judicious skepticism is essential for navigating deepfakes and disinformation. Instead of blindly trusting or sharing attention-grabbing content, individuals must apply critical thinking skills to evaluate media authenticity before reacting. Such discernment includes considering sources, motives, plausibility and corroborating evidence beyond just the content itself. With healthy skepticism as a default, people can avoid manipulation traps and make wise information consumption choices. Skepticism is an attitude of questioning and not accepting claims at face value. This mitigates confirmation bias where people believe information aligning with preconceived notions. Deep skepticism asks - who is sharing this information and why? Does the content match established facts or seem sensationalized? What's missing from the narrative? Cross-verifying suspicious info with credible independent sources provides perspective.

Before sharing unvalidated viral media, prudent skepticism prompts further scrutiny. What date was this footage recorded? Are identifiers like time and location corroborated? Do the people and scenes look natural or staged? Does it degrade opposing views instead of presenting objective facts? Pausing to question context helps avoid giving undue credence to manipulated content. Skepticism is not cynicism however. Wholesale dismissal of all media as fake risks disengagement. Instead, healthy skepticism is open-minded and evidence-based. It seeks credible authoritative sources to validate information. Facts and moral reasoning anchor conclusions, not just doubts and emotions.

Individually applying skepticism requires constant vigilance - an ongoing exercise rather than one-off solution. As deepfake technology evolves, new manipulation tactics emerge requiring updated diligence. Cultivating a lifelong habit of cautious discernment and deliberate analysis builds lasting immunity against deceit. Societally, a climate of constructive skepticism balanced with trust creates resilience. Shared heuristics, transparency and accountability around information sources foster trust in credible news and institutions. But mass awareness of deepfakes prevents blind faith. A discerning yet engaged public is difficult to manipulate but also open to truths. In summary, applying consistent healthy skepticism instead of blind trust has become imperative in the digital age. Questioning the veracity of all media - whether aligning with or opposing one's views - helps avoid reactionary responses to misinformation. Pausing to verify sources, assume positive intent, and find common ground despite divides builds societal resilience. With vigilance and discernment, truth sustained by facts, compassion and justice will prevail over disinformation.

4.3 Following Cybersecurity Best Practices to Minimize Manipulation Risks

While vigilance against media manipulation is crucial, following basic cybersecurity best practices also provides a valuable shield against deepfake risks. As deepfakes intersect with cybercrime, social engineering, and hacking, strong digital hygiene and data privacy habits limit vulnerabilities. Cybersecure infrastructure and responsible data sharing help minimize harm from stolen identities and inference attacks. With proper precautions, individuals and organizations can reduce their attack surface for malicious deepfake creation.

At an individual level, steps like enabling multi-factor authentication, using password managers, and performing software updates constrain data access. Monitoring financial statements and checking credentials prevents fraud enabled by stolen identities and synthesized voices. Backing up data securely offline also limits materials available to generate personal deepfakes.

Sharing personal images and videos online entails lasting risks of deepfake misuse. Adjusting social media privacy settings prudently, reporting exploitative accounts, and minimizing oversharing of sensitive media reduces exposure. The same vigilance used to spot misleading deepfakes should apply for sharing personal content that may become fodder.

For public figures like politicians and celebrities, proactive reputation management includes monitoring where their images spread online or purchasing rights. They run elevated risks of deepfakes influencing public perception. Staying vigilant against unauthorized use and false accounts issuing statements in their names limits reputational damage.

At an organizational level, vulnerabilities also stem from poor data governance. Insufficient access controls, security testing, and data compartmentalization enable breaches by malicious actors. Following established frameworks like the NIST Cybersecurity Framework hardens infrastructure against intrusions. Encrypting data also impedes deepfake generation using stolen datasets.

Most importantly, fostering an organizational culture of cyber-awareness through training minimizes human element risks. Educating employees on cyber hygiene, phishing attacks, suspicious communication policies, and reporting processes reduces entry points for social engineering. Staying vigilant against unusual activities, verifying requests, and enacting the principle of least privilege access across teams enhances resilience tremendously.

In summary, while deepfake risks feel novel, following cybersecurity fundamentals goes a long way. From strong passwords to controlled data sharing to constant vigilance, responsible digital habits build crucial immunity against disinformation. Alongside skepticism and technical countermeasures, earnest cybersecurity preparedness across sectors forms a durable immune system for society.

5. THE FUTURE OF DEEPFAKES

5.1 Projections on Improvements in Deepfake Technology

As deepfakes continue proliferating, improvements in underlying generative AI threaten to accelerate creation capabilities beyond current safeguards. Based on the technology trajectory and research frontiers, experts project deepfakes will achieve photorealism indistinguishable from truth, become trivially accessible through apps, and evolve adaptive adversarial capabilities against detection algorithms. This outlook underscores the urgent need for proactive solutions before reach an inflection point of unrestrained disinformation. Deepfake algorithms continue achieving exponential gains in image and video quality

through better training frameworks, neural architectures, and data. Startups are pairing GANs with 3D modeling, physics-based rendering, and animation techniques to improve realism. The creative possibilities are also expanding - deepfakes have generated synthetic news anchors, fictional celebrities, and non-existent people showcasing new levels of manipulation.

Accessibility is being enhanced through consumer apps, web tools, and automatic generation platforms. Chinese app Zao offers consumer-grade face-swapping built atop state-of-the-art research. Deepfake web tools make video creation as easy as writing text prompts. New services allow users to custom synthesize images simply by providing textual descriptions. Generating compelling personalized media is becoming as effortless as social media posting. Equally concerning is adversarial evolution to evade detection through techniques like attention masking, image blending, distortion removal, and anti-forensics. Startup Lyrebird simulated voice speech detection systems in the loop during training to optimize synthesized audio against detection. Similar evasion of biometrics, blockchain fingerprints, and other safeguards may follow. The capacity to evade and adapt will likely outmatch static analysis rules and models.

These improvements hint at a future where deepfake capabilities become democratized, ubiquitous and increasingly impervious to identification by humans or machines. Soon real-time video generation, continuous voice cloning, and lifelike VR avatars may redefine trust in digital interactions. While promising for creativity and access, preventing harms will require major initiatives in technology, law, and societal readiness before the point of no return. In summary, deepfakes look set for exponential advancements in quality, accessibility, and evasion abilities through constant AI innovation. As barriers disappear, we may reach an inflection point where salvaging authenticity and truth becomes impossible across media. To avoid this breakdown of reality, urgent action is needed on governance frameworks, trusted verification systems, public awareness, and comprehensive monitoring before the technological genie leaves the bottle.

5.2 Concerns Over Potential Large-scale Disinformation Campaigns

As deepfakes grow more accessible and realistic, concerns loom over weaponization through coordinated disinformation campaigns. The capacity to fabricate authentic-looking videos allows for unprecedented mass manipulation with viral false narratives. Without mitigation, the raw scale enabled by deepfakes could overwhelm the careful discernment required to identify falsified media. The risks span everything from political turmoil, market manipulation, hacked identities to doctored evidence in trials. Experts warn deepfakes may fuel significant social division and instability in coming years. Widespread synthetic video hoaxes personalized for different audiences could sow confusion and discord. Domestic and foreign adversaries could deploy deepfakes to weaken confidence in institutions through doctored yet credible-looking footage that is difficult to debunk at scale.

Democratic processes face high risks of interference. Adversaries could hijack political campaigns by depicting leaders making inflammatory remarks right before elections. Sharing thousands of personalized deepfakes showing a candidate disparaging target demographics may achieve recirculation before fact-checks contain damage. Voter impersonation, surveillance, and intimidation also grow easier with synthesized faces and voices. Financial fraud and market manipulation enabled by identity theft and falsified evidence poses another threat. Deepfaked video of executives making major announcements can move markets before credibility is questioned. False admission of guilt through synthesized media can also

tank stock prices during activism campaigns. For industries where reputation is paramount, mere deepfake allegations could inflict outsized damage regardless of veracity.

Law enforcement and courts face augmented threats of fabricated evidence undermining justice. Criminal deepfakes depicting individuals in incriminating scenarios may wrongfully influence verdicts in trials. Corrosive erosion of truth poses dangers even with rigorous scrutiny - doubts alone could paralyze the quest for accountability. The possibility space for injustice expands exponentially in system relying heavily on video evidence. While risks seem dire, deepfakes are hardly an existential threat to truth, but vigilance and safeguards remain vital. A resilient, discerning society, anchored in compassion over tribalism and facts over emotions, can overcome deception. Still, large investments into technological defenses, public awareness, law enforcement and cybersecurity readiness are prudent to prevent destabilizing manipulation. Time remains to erect societal antibodies against viral deepfakes - but the window is closing fast. In summary, deepfakes enable mass, personalized disinformation at unprecedented scales. The diffusion dynamics mean fact-checking often lags while damage is done. Without concerted efforts to enable rapid verification and inoculate public skepticism, deepfakes pose a powerful tactic for driving instability, distrust, and outrage without accountability. Averting this dangerous future requires prioritizing investments into multi-faceted solutions encompassing technology, law, markets, education and governance.

5.3 Ongoing "Arms Race" Between Deepfake Creation and Detection

As deepfake technology continues advancing, an escalating arms race is unfolding between the development of synthetic media and methods to detect it. Each new technique for creating realistic fakes spawns new forensic techniques for exposing them, and vice versa. This cycle of innovation appears likely to continue as capabilities improve on both sides. Absent game-changing breakthroughs, maintaining societal resilience against deepfakes will require embracing this constant cycle of enhancement rather than seeking a silver bullet. The dynamics of this arms race stem from the dual-use nature of generative AI. The same deep learning advances that refine deepfake creation also expose new signals for detection algorithms. For example, GAN fingerprints left during training enabled detection research to identify neural synthesis origins. But newer models are adapting to erase such fingerprints and evade detection through anti-forensics.

Likewise, algorithms leveraging photorealism cues and facial geometry analysis fueled breakthroughs in deepfake detection. Now, models are getting better at maintaining consistency across frames to outwit such analysis. In response, detection has incorporated temporal signals across multiple frames. This cycle of back-and-forth innovation continues across domains like audio, video, and image synthesis and authentication. With both creation and detection built atop rapidly evolving AI, the arms race is unlikely to end. Commercial incentives also exist on both sides, with startups pursuing adversarial deepfake and anti-deepfake technologies. Absent a plateau in research progress, keeping detection capabilities ahead of creation requires running just to stand still.

Rather than seeking a definitive solution, adapting to this cycle may prove more prudent. Developing agile, diversified detection combining biometrics, metadata analysis, blockchain ledgers, and human reviews enhances resilience. Societal antibodies also include public awareness, digital literacy, lawful deterrence, and infrastructure modernization to verify provenance. Holistic vigilance and mitigation across sectors offers the lasting way forward. In summary, an indefinite arms race characterizes the interplay between deepfake creation and detection. Each innovation spawns new vulnerabilities as capabilities improve on

both sides. Rather than chasing temporary patches, society must commit to perpetual enhancement of technological defenses, public skepticism, digital hygiene, and modern information architectures. By embracing reality of constant threats from increasingly sophisticated deepfakes, we can build durable and adaptive societal resilience.

6. CONCLUSION

6.1 Summary of Main Points

Deepfake technology poses an unprecedented challenge to truth and trust in the digital age. As AI synthesis capabilities improve exponentially, manipulated video, audio, and images are becoming indistinguishable from reality. This proliferation threatens to undermine belief in online information, enable new forms of disinformation, and erode social cohesion. Tackling the malicious uses of deepfakes demands a thoughtful, coordinated societal response across technology, education, law, and governance. We began by reviewing how deepfakes leverage generative adversarial networks to achieve photorealistic results. By implicitly learning facial expressions, skin features, and speech patterns, GAN algorithms can transpose an individual's likeness and voice onto target media. Early limitations are also disappearing - deepfakes exhibit fewer visual artifacts, synchronized speech, and smooth motion dynamics thanks to better techniques.

At the same time, deepfake creation is becoming widely accessible through consumer apps. Tools like DeepFaceLab, FakeApp and Zao enable anyone to fabricate realistic celebrity photos and videos through easy user interfaces. This low barrier for manipulating media presages a future where falsified content spreads virally before proper verification.

We then surveyed emerging safeguards against deepfakes. Digital authentication techniques like blockchain ledgers, cryptography, and video watermarking help establish provenance. AI detection algorithms reveal inconsistencies in faces, voices, reflections and other signals indicative of synthesis. While promising, technical countermeasures remain locked in an escalating arms race against improving generative models. Bolstering public awareness and skepticism provides another crucial line of defense. Educational campaigns teach individuals how to spot visual inconsistencies and questionable context in assessing media authenticity. Fostering a judicious skepticism instead of blind trust limits reactionary sharing of unverified content. Following cybersecurity best practices also reduces vulnerabilities.

Looking ahead, experts warn of exponential improvements in deepfake quality, accessibility, and evasion abilities. Deepfakes could enable widespread personalized disinformation undermining trust in institutions. Preventing harms requires prioritizing investments into multi-faceted solutions before reaching a point of uncontrolled proliferation. In summary, deepfakes present an extraordinary challenge, but not an insurmountable one. With vigilance, education, and a combination of technological, legal, and societal strategies, truth and trust can prevail. By making progress on early detection, media authentication, lawful deterrence, and inoculating public skepticism, society can maintain resilience against disinformation. Meeting the epochal test posed by deepfakes demands cooperation, wisdom and moral courage across all stakeholders.

6.2 Recommendations for Combating Deepfake Risks Through Technology, Education, and Skepticism

As deepfake technology proliferates, a comprehensive response encompassing technology, education, and skepticism provides the most prudent path for mitigating harms. Specifically, we must accelerate

research into authentication systems, foster greater public awareness, and cultivate healthy skepticism around online media. Combined with legal deterrence, ethical technology development, and modernized information architectures, such efforts offer a resilient bulwark against misuse. On the technology front, private sector and academic researchers should collaborate on forensic techniques for deepfake detection and provenance validation. Approaches like digital watermarking, metadata standards, perceptual AI models, and tamper-evident instrumentation show promise and warrant investment. Constant enhancement is required to keep pace with improvements in generative AI.

Equally important are information architectures and protocols for establishing media authenticity and pedigree. Developing blockchain verification networks, authenticated online repositories, and standards for preserving metadata like timestamps during distribution will enable better media tracing. Modernizing digital infrastructure to preserve and share provenance details is imperative. For rapid identification, platforms must enable robust verification pipelines. Integrating multiple signals like user reports, machine learning classifiers, authenticated source checks and manual review can contain deepfake spread. Prioritizing informational integrity over engagement and virality metrics will also discourage harmful misuse. On the education front, digital literacy programs, interactive tutorials, and repeated trainings should teach people how to exercise discernment in spotting and scrutinizing deepfakes. Curricula focusing on critical thinking and media analysis across all demographics will provide societal antibodies. Public awareness campaigns through games and media can further bolster immunity.

Platforms also bear responsibility in fostering wisdom and skepticism among users. Labels highlighting synthetic media sources, friction and pauses before sharing misinformation, and flagging unverified claims can nudge behaviors. Prominently rewarding contributors upholding dignity over division also helps. Finally, legal frameworks and international norms around deepfake use require development. Lawful restrictions on non-consensual deepfakes combined with anti-defamation laws can deter malicious actors. Peaceful advancement of technology necessitates ethics and wisdom. In summary, a multipronged response across technology, education, law, and behavioral design is recommended to counter deepfake risks. This combines continuous authentication research, modern verification pipelines, infrastructure upgrades, public awareness initiatives, healthy friction for virality, and lawful deterrence against misuse. With diligence and collective will, society can ascend over disinformation.

REFERENCES

[1] Takruri, L. (2023, July 19). What are deepfakes and how do fraudsters use them? | Onfido. Onfido. https://onfido.com/blog/what-are-deepfakes/
[2] L. (2023, May 3). What is Deepfake Technology? All You Need To Know. Forensics Insider. https://www.forensicsinsider.com/digital-forensics/what-is-deepfake-technology/
[3] Shaji George, D. A. (2023, October 25). Evolving with the Times: Renaming the IT Department to Attract Top Talent | Partners Universal International Innovation Journal. Evolving With the Times: Renaming the IT Department to Attract Top Talent | Partners Universal International Innovation Journal. https://doi.org/10.5281/zenodo.8436646
[4] Y. (2023, May 17). The Rise of Deepfakes: Navigating Legal Challenges in Synthetic Media. CBA's @theBar.
[5] Shaji George, D. A., Hovan George, A. S., Baskar, D. T., & Gabrio Martin, A. S. (2023, March 31). Human Insight AI: An Innovative Technology Bridging The Gap Between Humans And Machines For a Safe, Sustainable Future | Partners Universal International Research Journal. Human Insight AI: An Innovative Technology Bridging the Gap Between Humans and Machines for a Safe, Sustainable Future | Partners Universal International Research Journal. https://doi.org/10.5281/zenodo.7723117

[6] Aldredge, J. (2020, June 9). Is Deepfake Technology the Future of the Film Industry? The Beat: A Blog by Premium Beat. https://www.premiumbeat.com/blog/deepfake-technology-future-of-film-industry/
[7] The Dangers of Manipulated Media and Video: Deepfakes and More. (2021, February 8). ADL.
[8] Shaji George, D. A. (2023, September 25). Future Economic Implications of Artificial Intelligence | Partners Universal International Research Journal. Future Economic Implications of Artificial Intelligence | Partners Universal International Research Journal. https://doi.org/10.5281/zenodo.8347639
[9] Demystifying deepfake videos: The powerful fusion of technology and data science | Data Science Dojo. (n.d.). Data Science Dojo. https://datasciencedojo.com/blog/deepfake-videos-technology/
[10] Understanding the Technology Behind Deepfake Voices. (2023, April 28). Understanding the Technology Behind Deepfake Voices. https://murf.ai/resources/deepfake-voices/
[11] Shaji George, D. A., & Hovan George, A. S. (2023, October 11). The Rise of Robotic Children: Implications for Family, Caregiving, and Society | Partners Universal Innovative Research Publication. The Rise of Robotic Children: Implications for Family, Caregiving, and Society | Partners Universal Innovative Research Publication. https://doi.org/10.5281/zenodo.10045270
[12] Deepfake - Wikipedia. (2021, November 1). Deepfake - Wikipedia. https://en.m.wikipedia.org/wiki/Deepfake
[13] Shaji George, D. A. (2023, October 11). Securing the Future of Finance: How AI, Blockchain, and Machine Learning Safeguard Emerging Neobank Technology Against Evolving Cyber Threats | Partners Universal Innovative Research Publication. Securing the Future of Finance: How AI, Blockchain, and Machine Learning Safeguard Emerging Neobank Technology Against Evolving Cyber Threats | Partners Universal Innovative Research Publication. https://doi.org/10.5281/zenodo.10001735
[14] Top 7 Best Deepfake Video Makers 🚀 Speechify. (2022, September 27). Speechify.
[15] Frąckiewicz, M. (2023, July 5). Deepfakes: The Growing Threat of AI-Generated Synthetic Media. TS2 SPACE. https://ts2.space/en/deepfakes-the-growing-threat-of-ai-generated-synthetic-media/
[16] A. (2023, August 17). DeepFake Detection - Scaler Topics. Scaler Topics. https://www.scaler.com/topics/deepfake-detection/
[17] Rolling in the deepfakes: generative AI, privacy and regulation : Clyde & Co. (2023, October 2). Rolling in the Deepfakes: Generative AI, Privacy and Regulation : Clyde & Co. https://www.clydeco.com:443/insights/2023/10/rolling-in-the-deepfakes-generative-ai-privacy-and
[18] Deepfakes and Disinformation Pose a Growing Threat in Asia. (2023, March 11). Deepfakes and Disinformation Pose a Growing Threat in Asia – the Diplomat. https://thediplomat.com/2023/03/deepfakes-and-disinformation-pose-a-growing-threat-in-asia/
[19] D. (2022, June 25). Everything You Need to Know About Deepfake Technology. DeepSwap.
[20] Narang, R. (2022, September 4). #002 Deepfakes - The Creation and Detection of Deepfakes: A Survey - Master Data Science 04.09.2022. Master Data Science. https://datahacker.rs/009-the-creation-and-detection-of-deepfakes-a-survey/
[21] Generative AI Models Types and its Applications | Quick Guide. (2023, November 10). Generative AI Models Types and Its Applications | Quick Guide. https://www.xenonstack.com/blog/generative-ai-models
[22] Ezquer, E., D., & Writer, G. (2023, May 4). AI-Generated Media: A Guide to Understanding DeepFakes - Metaroids. Metaroids. https://metaroids.com/learn/ai-generated-media-a-guide-to-understanding-deepfakes/
[23] Teach, H. (2023, August 27). Plunging into the Reality of Deepfake AI: An Informative Guide » HEX Teach. HEX Teach. https://hexteach.com/blog/what-is-deepfake-ai
[24] Windmill Testing Framework. (n.d.). Windmill Testing Framework. https://getwindmill.com/best-cybersecurity-practices-for-business/