

INTERNATIONAL JOURNAL FOR LEGAL RESEARCH AND ANALYSIS



Open Access, Refereed Journal Multi Disciplinary
Peer Reviewed

www.ijlra.com

DISCLAIMER

No part of this publication may be reproduced, stored, transmitted, or distributed in any form or by any means, whether electronic, mechanical, photocopying, recording, or otherwise, without prior written permission of the Managing Editor of the *International Journal for Legal Research & Analysis (IJLRA)*.

The views, opinions, interpretations, and conclusions expressed in the articles published in this journal are solely those of the respective authors. They do not necessarily reflect the views of the Editorial Board, Editors, Reviewers, Advisors, or the Publisher of IJLRA.

Although every reasonable effort has been made to ensure the accuracy, authenticity, and proper citation of the content published in this journal, neither the Editorial Board nor IJLRA shall be held liable or responsible, in any manner whatsoever, for any loss, damage, or consequence arising from the use, reliance upon, or interpretation of the information contained in this publication.

The content published herein is intended solely for academic and informational purposes and shall not be construed as legal advice or professional opinion.

**Copyright © International Journal for Legal Research & Analysis.
All rights reserved.**

ABOUT US

The *International Journal for Legal Research & Analysis (IJLRA)* (ISSN: 2582-6433) is a peer-reviewed, academic, online journal published on a monthly basis. The journal aims to provide a comprehensive and interactive platform for the publication of original and high-quality legal research.

IJLRA publishes Short Articles, Long Articles, Research Papers, Case Comments, Book Reviews, Essays, and interdisciplinary studies in the field of law and allied disciplines. The journal seeks to promote critical analysis and informed discourse on contemporary legal, social, and policy issues.

The primary objective of IJLRA is to enhance academic engagement and scholarly dialogue among law students, researchers, academicians, legal professionals, and members of the Bar and Bench. The journal endeavours to establish itself as a credible and widely cited academic publication through the publication of original, well-researched, and analytically sound contributions.

IJLRA welcomes submissions from all branches of law, provided the work is original, unpublished, and submitted in accordance with the prescribed submission guidelines. All manuscripts are subject to a rigorous peer-review process to ensure academic quality, originality, and relevance.

Through its publications, the *International Journal for Legal Research & Analysis* aspires to contribute meaningfully to legal scholarship and the development of law as an instrument of justice and social progress.

PUBLICATION ETHICS, COPYRIGHT & AUTHOR RESPONSIBILITY STATEMENT

The *International Journal for Legal Research and Analysis (IJLRA)* is committed to upholding the highest standards of publication ethics and academic integrity. All manuscripts submitted to the journal must be original, unpublished, and free from plagiarism, data fabrication, falsification, or any form of unethical research or publication practice. Authors are solely responsible for the accuracy, originality, legality, and ethical compliance of their work and must ensure that all sources are properly cited and that necessary permissions for any third-party copyrighted material have been duly obtained prior to submission. Copyright in all published articles vests with IJLRA, unless otherwise expressly stated, and authors grant the journal the irrevocable right to publish, reproduce, distribute, and archive their work in print and electronic formats. The views and opinions expressed in the articles are those of the authors alone and do not reflect the views of the Editors, Editorial Board, Reviewers, or Publisher. IJLRA shall not be liable for any loss, damage, claim, or legal consequence arising from the use, reliance upon, or interpretation of the content published. By submitting a manuscript, the author(s) agree to fully indemnify and hold harmless the journal, its Editor-in-Chief, Editors, Editorial Board, Reviewers, Advisors, Publisher, and Management against any claims, liabilities, or legal proceedings arising out of plagiarism, copyright infringement, defamation, breach of confidentiality, or violation of third-party rights. The journal reserves the absolute right to reject, withdraw, retract, or remove any manuscript or published article in case of ethical or legal violations, without incurring any liability.

THE AGE OF DIGITAL IMPERSONATION: DEEPPFAKE AND THE LAW'S DILEMMA

AUTHORED BY - TARUSHI GUPTA & NAGA KAIVALYA

ABSTRACT

Today's courts face chaos as fake videos and images crafted with tools like GANs and diffusion models turn up everywhere. It's so easy now: criminals whip up convincing fakes to steal identities, fool banks, demand ransom, spread false stories, or leak intimate images without consent. Deepfakes can slip through a scene, showing moments that never happened, while doctored photos freeze a lie and push the same ugly agenda. People end up anxious, reputations shattered, bank accounts wiped out. Companies get hit, public trust drains away, and even elections feel shaky. Different countries draw the line in their own way. South Korea sticks strict rules into its sex crime laws, aiming only at specific abuses. India goes broad, relying on sweeping digital regulations from 2000, then layering on fresh updates in 2023. No border really stops these crimes if you want to fight them, laws must pinpoint the threat. When fake identities race across the web, helping people recover and defend themselves matters just as much as nabbing the crooks. It's not just about slapping on watermarks. Digital forensics the careful, methodical hunt for evidence packs just as much punch. Altered images slip past most defences, unnoticed. As AI gets more cunning, online theft sharpens up.

Keywords: Deepfakes, Morphing, AI-driven crimes, GANs, Cybercrime, Identity theft, financial fraud, Misinformation, Legal framework, Victim protection.

INTRODUCTION

Imagine you wake up one morning and you stumble on a video and you see a popular leader, a judge, or even a person you know personally addressing straight, confident and persuasive. The face is real, the voice is authentic, the expressions are natural, and the message is unquestionable. Yet, none of it ever happened. The individual did not utter such words, did not make such a statement, or appear before a camera. What you are looking at is not real, but it is a well-engineered illusion known as deepfake a potent weapon to mankind¹.

¹ Fortinet, 'What Is Deepfake: AI Endangering Your Cybersecurity?' (Fortinet2023)

Current developments explain how deepfake technology has evolved over the years, turning into a rare curiosity and a deep-rooted threat to society. In 2024, a financial worker at Arup fell victim to a scam in which a fake video showing senior executives was used to convince the employee to transfer an amount of 25.6 million dollars, after an AI-generated video call of verisimilitude quality scheduled the fake video session, and a Ukrainian President Volodymyr Zelensky was impersonated in a fake video spread via a hacked television network, urging soldiers to surrender and causing confusion in the middle of ongoing hostilities. Similarly, the cybersecurity company KnowBe4 accidentally hired a North Korean hacker who, with the help of a fake deepfake identity, passed several interviews and background checks, which highlights the gaps in the practice of remote hiring. In effect, all these cases demonstrate the fact that deepfakes have gone beyond defrauding individuals out of money to tampering with trust, security, and the ethical line. It clarifies that deepfakes are not a harmless joke anymore, but a systemic threat in the areas of finance, politics, jobs, and online culture².

CONCEPTUAL FRAMEWORK

Now shaping how people make and watch online content, artificial intelligence moves fast. A debated result? Deepfakes - media made by tech to look real even though it is fake. Behind them sits machine learning, systems studying huge image sets to copy facial details, speech, tones. Once trained, those models generate near-perfect likenesses, mimicking how humans appear in videos or audio clips. Though subtle, the shift changes trust in what eyes see digitally. Yet blending faces from several people, morphing methods lack high-tech complexity but still create serious flaws in ID checks. Although different in approach, each can enable fraud, trickery, or online criminal acts.

TYOLOGY OF CRIMES ENABLED BY DEEPFAKES AND MORPHING.

Fake Faces, Real Risks: Deepfake Identity Fraud

Now spreading through online networks, what used to belong in futuristic tales emerges silently - faces, voices, audio altered by synthetic means until they mirror truth almost completely. Not only do these fabrications bypass confidence checks, but they also expose weak spots in how we confirm facts across banking, governance, and private conversations. Beginning not with

<<https://www.fortinet.com/resources/cyberglossary/deepfake>>.

² Preeti Kulkarni, '20+ Real Deepfake Examples That Shocked the World' (*HyperVerge* 26 February 2025) <<https://hyperverge.co/blog/examples-of-deepfakes>>.

tools but ideas, deepfakes grow from complex training systems such as GANs and encoding models capable of producing visuals and noises so close to real that telling them apart becomes difficult. Central to this effort stands the Generative Adversarial Network (GAN), where one part invents false content while another, independent checker hunts down imperfections. Because of their continuous challenge - one creating, the other doubting - both improve steadily, pulling imitation nearer to authenticity. Elsewhere, some methods lean on compressed representations using encoders paired with gradual refinement processes; data enters, gets reduced, then reappears detailed and convincing. By doing so, they allow exact face swaps or voice shifts without losing subtle characteristics³.

From hidden corners of digital space, tools like deepfakes and automated dialogue systems now appear within messages tied to groups such as the Islamic State and networks linked to al-Qaida. Instead of relying on raw footage, these actors shape fake broadcasts that mimic real news segments using computer-made faces and voices. One way they spread influence involves tailored interactions through bots designed to mirror human conversation closely. Because of this shift, emotional appeal grows stronger among younger users navigating online forums or private apps. A sense of realism emerges even when none exists, helping false narratives gain ground quietly. Over time, manufactured clips make extremist movements seem larger or more accepted than they are. Since each altered video or message multiplies quickly, oversight becomes harder for governments and tech companies trying to respond. What once required teams now happens in seconds without physical presence. Reality blurs further whenever synthetic media avoids clear markers of manipulation. As methods evolve silently, tracking origins turns into an endless chase across fragmented channels.⁴

Forged Prints

A single technique, called Proxy Prints, builds artificial fingerprints much like deepfake photos - using neural networks to shape minute features, such as ridge endings and bifurcations, into lifelike prints. Instead of relying on real fingers, it generates synthetic ones from digital traces left behind in biometric databases. With help from low-cost 3D-printed molds, those designs take form through silicone replicas, reviving hidden template files as tangible prosthetics. Experiments reveal success rates near eighty percent across widely used sensors, even devices

³ IMI, 'Deepfake Deception in Digital Identity - Identity Management Institute®' (*Identity Management Institute*®10 February 2025) <<https://identitymanagementinstitute.org/deepfake-deception-in-digital-identity/>>.

⁴ADF, 'From Deepfakes to Chatbots, Terrorists Embrace AI to Spread Propaganda' (*Africa Defense Forum*9 September 2025) <<https://adf-magazine.com/2025/09/from-deepfakes-to-chatbots-terrorists-embrace-ai-to-spread-propaganda/>>.

made by *ZKTeco* and *SecuGen* surviving under such spoofing attempts. So many deepfake tools are available and very cheap that it is possible to use a realistic faked person. The threat is not a theory anymore; it has very well manifested itself in reality. One of the most fascinating areas where AI is expected to be used for illegitimate monetization is “synthetic eyes.” Artificial intelligence systems can deceive biometric systems using synthetic eyes with natural movements in the event of biometric access that recognizes use iris or retina for recognition. This can be evade easily. Fakeness now looks too much like truth for simple defences to catch. These tricks bypass checks designed to prove someone real is actually there.

Deepfakes pornography, an AI driven exploitation

Some people leverage artificial intelligence to generate false intimate images, placing a person's likeness into adult videos without consent. Instead of obtaining approval, they deploy technologies such as face-swap programs - powered by complex algorithms - to carry out the task. Often, pictures shared publicly get repurposed as source data, forming realistic-looking but entirely manufactured situations. These outputs spread quickly due to automated processes, undermining reputations and emotional well-being, with women facing higher risks. Most deepfake videos spotted online feature sexual content, studies cited by specialists indicate. Nearly all targets of these altered clips turn out to be female, data suggests. Women make up almost the entire group impacted, appearing in around 99 percent of known instances. Close analysis reveals how heavily gendered this misuse really is.

Impacts and Consequences

Deepfakes and morphing attacks can damage people, institutions, and democracy at the same time, and those harms often reinforce each other. Below is a structured view from the level of an individual victim up to the global system.⁵

Individual victims

Trauma impacted by psychological harm and Financial Losses and Life Opportunities

- Frequently, people who experience psychological trauma due to the unauthorized use of their facial image (in an image or video) or voice (in a recording, video, or any form of digital media) report having significant anxiety, depression and other post-traumatic

⁵Shashikant Shinde, ‘Deepfakes and Their Impact on Society’ (2024) Volume 4 International Journal of Advanced Research in Science, Communication and Technology 283.

stress disorder-like symptoms, especially when they are shown in sexualised or disparaging ways. Many people report feeling a significant loss of autonomy and ownership of who they are (the elements of their identity), and consider this feeling an actual "violation of self". People also report having recurrent fears about being shown in a fabricated or false manner at any time in the future, which negatively affects their ability to sleep, concentrate and function on a day-to-day basis. Reputational Damage.

- Deepfake content and the impersonation of one's identity can be permanent, so one must be aware. A false claim or idea spreads like wildfire but a true correction cannot do so. Often there is reputational damage due to a flawed perception.

Privacy and Bodily Autonomy

- The use of technology to create fake images that depict someone being sexual or demeaned (i.e., deepfakes) is an invasion of that person's privacy and their right to control their own body/image. Women and girls are disproportionately impacted as they account for most victims of abusive deepfakes. A single leak of that image and/or video may be copied, altered, and shared numerous times with different people, in different place, and it will be almost impossible to remove from existence.

Fraud and Financial Risk

- Deep in the world of CEO fraud and vishing we see the use of deepfake audio and video. In these we find that attackers pose as C-level executives or vendors to push for urgent wire transfers or to change payment info. Also, we see that these attacks are getting past traditional email defences which is due staff's tendency to trust a familiar voice or face which in turn may result in large single instance financial losses.

Compliance and Regulatory Failures

- Synthetic identities which have been digitally altered ID photos can get past remote KYC and on-board systems. This puts financial institutions, fintech companies and telecom providers at risk for violations of anti-money laundering and counter terrorism financing regulations. If regulators decide that due diligence and verification measures were not strong enough, they may put in place large fines, sanctions, and also require extensive remedial actions.

Operational Disruption and Security

- Deepfakes is a social engineering tactic that is increasingly gaining popularity among hackers to extract information from employees, customers, and suppliers. Additionally, they are also employing this method of attack to infiltrate corporate systems without authorization. These advanced attacks can result in the exploitation of multi-factor authentication, a loss of data, and ransomware in the attack. React.

Brand and Trust Erosion

- If as a customer we are presented with very realistic doctored videos of a company's spokesperson which contain offensive content or false financial info that damage the brand's image and that trust which had been built up between the company and the public can break down very quickly. To get back some of that which was lost in terms of reputation requires open communication from the company and also, they put in place very visible corrective actions; but there are some elements of reputation which may never fully recover. Societal and Democratic Harms

Misinformation and Polarization

- Very realistic doctored speeches, protests, or violent events which can spread very quickly via social media. This in turn may put out of balance public perception before fact checking entities have a chance to react. Also, we see that which of this content is put out plays into and amplifies present social and ethnic issues which in turn causes these into to grow in which people tend to seek out info that confirms what they already believe in thus creating ideological silos.

Threats to Democratic Processes

- Deepfakes are realistic enough videos that can make it appear as if someone said or did something that did not. Making a fake bribe, statement, or confession in political contexts can be damaging. Its negative publicity can damage a person's image; it can influence voters' perceptions; and more.

Undermining Rule of Law and Human Rights

- When audio or video evidence is compromised, it can greatly impact criminal prosecution and the victimization of victims. Because of this violent crime is primarily

reliant on evidence which is becoming untrustworthy. We cannot ignore that deepfakes can pose a serious threat as well.

Global and Cross-Border Dimensions

- Borderless Dissemination, synthetic media can be generated within one jurisdiction and instantaneously shared globally, frequently via encrypted or foreign-hosted platforms that present significant regulatory challenges for local authorities. Once a deepfake or altered image achieves widespread viral circulation, numerous mirror copies and remixes often appear across various services, rendering removal processes slow, incomplete, and technically complex.

Enforcement and Jurisdiction Challenges

- Laws pertaining to data protection, defamation, cybercrime, and image-based abuse exhibit substantial variations across different countries. This creates "safe havens" where perpetrators can operate with a diminished risk of legal repercussions. The process of gathering cross-border digital evidence, identifying pseudonymous offenders, and coordinating prosecutions frequently extends beyond the typical life cycle of a viral deepfake.

Asymmetry Between Attackers and Defenders

- The tools required for generating high-quality synthetic media are both inexpensive and globally accessible. Conversely, effective forensic detection and the implementation of legal remedies demand specialized expertise and robust cross-border collaboration. This inherent asymmetry implies that individuals and smaller organizations, particularly those in lower-resource countries, bear a disproportionate burden and possess limited recourse when they become targets.

Geopolitical and Information Warfare Risks

- State and non-state actors are capable of deploying deepfakes in influence operations. For instance, they can fabricate statements by foreign leaders or stage simulated atrocities, thereby amplifying existing propaganda strategies. Such campaigns have the potential to destabilize elections, peace processes, and financial markets across national borders, with the processes of attribution and accountability often proving highly complex.

Detection and Technical Countermeasures

- There are now systems used to detect if a content is generated by AI. They review the smaller details and patterns to figure out if the content is authentic or not. Deepfakes made with the help of advanced deep learning technology are so realistic that you can't tell if it is fake just by watching the video.

Detection and Technical Countermeasures

At present, deepfake videos are made with a level of quality that creates very few (if any) visible errors or glitches, making the end product feel more like real images or footage as opposed to the fake product that they are. Because of this, just looking at an image or a video will not validate if it is real. Many systems run behind the scenes analyzing the pixel pattern, the data rhythm behind the images and attempting to determine if the movement was created by a living being or not. These systems are meant to differentiate between the content that is produced naturally and that which was created artificially. They also analyze the methods used to create the content, who was targeted and if there are any factors influencing it. This process continues to develop each day through a form of continual observation but still will not achieve complete accuracy.

- *Forensic analysis & AI-based detectors (artifact detection)*

Through the careful examination of images, evidence that digital manipulation has occurred can be identified by close inspection of an image. One common indicator is unusual shadows around a person's face, another is an inconsistent pattern of lighting and/or a difference in compression levels from image to image. By placing these subtle signs, produced by various digital processes, in front of a program designed to detect forgery, it allows the program to learn from the examination of thousands of both real and digitally processed images concurrently. Additionally, some programs utilize multiple neural networks in order to analyze both static errors and moving errors of a video sequence. For example, a person's hand may move without synchronizing with their blinking, and then when viewed over a period can become a very identifiable error. Another example would be an inconsistent pattern of noise produced by a digital camera's sensor; when analyzed against other images, it will reveal its distinct noise characteristics once an individual has had time to study them. These tools become more precise as they have more and more examples to learn from. They adjust themselves over time to identify the differences between authenticity and deception without being

told exactly how to identify each difference. Even small variations in pixel blending impact how an item is detected.⁶

- *Liveness checks*

the purpose of liveness detection is to determine whether someone is really there, not just to confirm a recorded image. Liveness detection analyzes how someone reacts to the request of being told to shift position or respond quickly rather than only utilizing a single image. Humans when authentically present display small movements in their facial expressions and eye muscles that could not be easily replicated if the person was not physically present. Liveness detection evaluates such small movements as well as if they blink naturally or mechanically by examining the way in which they blink compared to a natural rate of speed. The purpose of liveness detection is to determine whether someone is really there, not just to confirm a recorded image. Liveness detection analyzes how someone reacts to the request of being told to shift position or respond quickly rather than only utilizing a single image. Humans when authentically present display small movements in their facial expressions and eye muscles that could not be easily replicated if the person was not physically present. Liveness detection evaluates such small movements as well as if they blink naturally or mechanically by examining the way in which they blink compared to a natural rate of speed. Some advanced versions track subtle shifts in face tone caused by blood pulsing under skin, picked up through careful lighting analysis. Even breathing motions, barely noticeable, become clues since most fake videos cannot mimic them accurately yet.⁷

- *Watermarking and provenance checking online.*

A watermark embedded in a digital file will only be detectable under particular conditions. Otherwise, watermarks are simply detected as soon as they are added during content creation or modification. Later on, using certain tools, experts were able to extract the watermark to determine where it was created or if it was generated using some helpful AI. Records get constructed with the file itself in addition to detecting after alteration and damage such as who modified it when and what device they used and what AI if any used. These files are going to get stored in chain of events that people cannot change. The receiving platform can later verify any claim against the event that

⁶Irene Amerini and others, 'Deepfake Media Forensics: Status and Future Challenges' (2025) 11 Journal of Imaging 73 <<https://www.mdpi.com/2313-433X/11/3/73>>.

⁷ Zerofox team, 'How to Detect DeepFakes: A Guide for Security Teams' (ZeroFox2025) <<https://www.zerofox.com/blog/how-to-detect-deepfakes/>>.

is stored. When determining trust, instead of looking for watermarks, check the written timeline as events happen.⁸

Limitations and the arms race

Most ways to spot fake content face the same core problem: progress in finding fakes often leads to smarter tools that hide or copy the clues systems look for. Watermark-free models, smoother transitions between frames, plus more natural motion help bypass detectors focused on flaws. At the same time, slight tweaks made on purpose can trick AI classifiers into seeing things wrong. Because of this, many specialists say relying only on one type of tool won't work well enough. A mix of strategies - like tracking origin data, adding hidden marks, using smart algorithms, checking if something is live, along with site rules - must work together to reduce threats. Even then, staying fully accurate forever isn't possible.

Legal and regulatory Landscape

US Laws against Deepfakes

One moment there's no real system, just bits and pieces across states and Washington. Then comes the Take it down act - landed in May 2025 - the country's first broad rule aimed at damaging fakes, especially explicit ones made without permission. Platforms now must wipe those posts fast, within two days after someone flags them, or face consequences; so, do those who make or spread them. Before that, only scattered federal powers applied: the FTC (federal trade commission) could step in when lies messed up markets, while older crime rules handled digital fraud or hacking through laws like the CFAA. Other ideas floated around, including the Defiance proposal meant to help harmed people fight back. Meanwhile, down below on state turf, nearly every one - 47 total - had already passed their own bans by late 2025, mostly against fake sex videos, election meddling, and swindling tricks. All these layers pile up quickly, making it tough to keep track as things shift.⁹

On March 20, 2025, the Senate passed legislation called the Take it down act, which criminalized the publication of a type of nonconsensual computer-generated imagery or "deepfakes." These include (but are not limited to) so-called "deepfake" nude photographs and movies and other computer-generated act(s) that depict individuals in nonconsensual intimate

⁸ Reality Defender Team, 'Deepfake Detection: Provenance vs. Inference' (*Reality Defender — Enterprise-Grade Deepfake Detection* 31 July 2025) <<https://www.realitydefender.com/insights/provenance-and-inference>>.

⁹ Henry Pattishman, 'Global Legal Actions against AI Deepfakes: Five Laws of 2025' (*Regula* 19 August 2025) <<https://regulaforensics.com/blog/deepfake-regulations/>>.

images. In a bipartisan effort by Senator Amy Klobuchar (D-MN) and Senator Ted Cruz (R-Texas), the bill was signed into law by President Trump on May 19, 2025, and goes into effect on May 19, 2026. The bill establishes a national law against this type of computer-generated imagery and, at the same time, works in concert with existing state-oriented laws against nonconsensual intimate imagery.¹⁰

The bill's disputation process does not penalize platform owners and similar entities for rigorous enforcement of computer-generated imagery in good faith if a person purporting to be the person in question is shown on the deepfake act and assertively requests removal of the act, the person possesses the legal authority to delink the act and reserve another access point for it, and if it is credibly positioned that the act was transferred without the person's consent.

To prevent abuse or prevent users from using computer-generated imagery to gain currency, the bill acknowledges the distinction between lawful and illegal deepfake activities. If the actions are lawful but illegal, then the criminal penalties will be pursued. However, when the activity is carried out in a manner that is not supranational to leverage or compel another user, then the actions will not be liable for the penalties set forth in the bill and will require information from law enforcement. The penalties to be set forth in the bill will be up to two years for those age 18 and older and three years for minors.¹¹

The defiance act, as part of the "DEFIANCE" (Dissemination of Equity for Intimate and Offensive Non-consensual Content) Act, provides legal recourse for victims of non-consensual intimate deepfakes, too. The Defiance act complements the Take it down act, as both create a narrow but meaningful federal civil right of action for victims of intimate digital forgeries. The Act (passed January 13, 2026), amending 15 U.S.C. §6851, adopts familiar elements of state "revenge porn" laws, but provides a national floor, allowing victims to file lawsuits against those who "create, distribute, and/or possess" their non-consensual intimate digital forgeries. This act also establishes a new Federal Court for all deepfake-related lawsuits, enabling victims to discover the identity of anonymous defendants—but allowing pseudonymous filings, too.¹² The law provides for "liquidated damages" of \$150,000 for the presumptively harmful violation of the victim's civil rights, with \$250,000 for "aggravated" cases. Victims can also

¹⁰ Cruz, Ted, 'Text - S.146 - 119th Congress (2025-2026): TAKE IT down Act' (*Congress.gov*2025) <<https://www.congress.gov/bill/119th-congress/senate-bill/146/text>>.

¹¹ Ashley R Lynem, Jacob A Sand and Patrick S Smith, 'TAKE IT down Act Targets Deepfakes: Are Online Platforms Caught in the Crosshairs?' (*Morganlewis.com*9 June 2025) <<https://www.morganlewis.com/pubs/2025/06/take-it-down-act-targets-deepfakes-are-online-platforms-caught-in-the-crosshairs>>.

¹² 'Senate DEFIANCE Act Drives Legal Reform - AI CERTs News' (*AI CERTs News*25 February 2026) <<https://www.aicerts.ai/news/senate-defiance-act-drives-legal-reform/>> accessed 23 April 2026.

seek compensation for their actual losses (if ascertainable), as well as the disgorgement of any profit made from the non-consensual intimate digital forgery.

If a victim must file suit to prove their case, the DEFIANCE Act mandates a fee payment by the violator to pay the victim's attorney, as well as an injunction against any further violations.¹³

The take it down act only requires platforms to remove content from their own environments. The defiance act provides a new federal civil cause of action.

Despite the shortcut, that route to accountability is not easy. In federal court, the plaintiff must prove the elements of their case to a jury, and the defendant has all the fact-finding tools of litigation at their disposal. Success—especially in the case of intent-based claims—is hardly assured. The defiance act ensures that when the victim chooses to go to court, they will be sufficiently incentivized to invest in the costs of proving the defendant's intent.

European Union fight against Deepfakes

European union AI act¹⁴

The AI Act of the EU is the first comprehensive regulation of synthetic media, including deepfakes, at the regional level. The European Union has decided on a risk-based framework in favour of a blanket ban that involves transparency, oversight, and bans. This tactic is intended to restrict the most damaging uses of deepfakes while continuing to allow helpful usage in entertainment, art, and research.

Transparency is one of the essential measures regarding the use of deepfakes under the AI Act. Technologies, targeted at developers and deployers of AI systems which generate or work with images, audio or video, require clear AI-generated content disclosures. Synthetic media is characterized by the mandatory indication on the platforms or by on the creators of synthetic media that allow a user to distinguish it from a real one. This is often implemented via a water mark or otherwise through metadata or visible notation that creates easy identification for general users.

Certain deepfake applications are considered 'high-risk' when used in political campaigns, for identity-related fraud, or for biometric surveillance. Risk assessments, human oversight, and data-quality requirements must be implemented by high-risk systems. The use of deepfakes

¹³Ben Coleman, 'Understanding the DEFIANCE Act Criminalizing Nonconsensual Deepfakes' (*Reality Defender — Enterprise-Grade Deepfake Detection* 31 July 2025) <<https://www.realitydefender.com/insights/the-defiance-act-and-combatting-nonconsensual-deepfakes>>.

¹⁴Alexandra Atsüren, 'EU AI Act 2024 | Regulations and Handling of Deepfakes - BioID' (*BIOID be recognized* 3 June 2024) <<https://www.bioid.com/2024/06/03/eu-ai-act-deepfake-regulations/>>.

that impersonate others in a manner that threatens their safety, privacy or democracy falls under this category.

In addition to these two aspects, traceability and accountability also serve as bid objectives. Producers of deepfake mode systems must keep a record of how content is generated, including the training data and the models used in creation. At the same time, deepfakes used for social-scoring or unlawful surveillance are clearly malicious practices that are classified as “unacceptable-risk” applications and banned across the EU.

AI Act aims to strike the right chord between capacity building and safeguarding the interests through revealing, reasonable supervision, and oppositions. The rule intends to halt deepfakes causing distrust, spreading false info, or violating rights whereas at the same time, it is motivating responsible trials and creation. With other countries planning to regulate synthetic media, the EU's strategy is turning out to be a model for democracies to manage deep fakes without suppressing technological advancements.

GDPR Implications¹⁵

The General Data Protection Regulation (GDPR) is an EU (European Union) law that helps protect the data privacy of EU citizens or residents (EEA) in the EU. GDPR applies to any organizations that process personal data in the context of an organization's work or location in the EU/EEA (European Economic Area). GDPR grants rights to citizens of the EU regarding how organizations handle their personal data. These rights include access, correction, deletion or destruction of data, objection, and so on. What makes GDPR different is the detail it goes into regarding how companies must handle data. For example, organizations must allow the lawful, fair, and transparent handling of personal data. GDPR imposes strict conditions on the use of personal data and breaches can result in fines of up to €20 million euros or 4% of a company's global turnover (whichever is higher). In today's world where technology is not only evolving but also changing the way data is being generated and consumed, the GDPR has far reaching implications for the future of technology.

There is a consensus that the increasingly sophisticated deepfake and AI (artificial intelligence) technologies have started to change the way personal data is shaped and used. Consent for using personal data has always been the cornerstone of data privacy. AI systems produce data by learning from multiple inputs. For AI created content to be accurately attributed to someone, their data would need to be processed without their consent. Such data processing could violate

¹⁵ Ben Welford, 'What Is GDPR, the Eu's New Data Protection Law?' (*GDPR.EU*2025) <<https://gdpr.eu/what-is-gdpr/>>.

not only the EU GDPR but also other consensus around the globe. For example, using a person's likeness in deepfake without their consent may legally infringe their ownership rights or equivalent rights in culture.

There are no straightforward solutions to address this growing challenge. AI governance, accountability and auditing are a few important aspects to explore for addressing potential harms from AI technology. However, to feel safe and secure, people will need to trust in the system of AI developers and the AI tools that are being built.

Deepfakes using private details need to be processed in accordance with clear legal frameworks. The first set of rules come from the GDPR, which determines the lawful handling of personal data. Subsequently EU's AI Act which, among other things, requires transparency and the careful management of risks. Together, these measures ensure that deepfake content depicting real individuals is produced and distributed within legal and regulatory boundaries.

South Korea and India

South Korea and India approach the issue of deepfake pornography in notably different ways, even though both operate within established legal systems. South Korea has taken a direct and specific route by amending the Act on Special Cases Concerning the Punishment of Sexual Crimes in September 2024 to explicitly criminalize deepfake pornography. The law makes it illegal not only to create and distribute such content but also to possess or view non-consensual deepfakes, with limited exceptions such as satire. The penalties are strict up to seven years of imprisonment for creation or distribution, and up to three years or heavy fines for possession or viewing. Enforcement is proactive, with specialized cybercrime units under bodies like the Ministry of Gender Equality and Family and the National Police Agency, along with obligations on digital platforms to remove such content quickly¹⁶.

In contrast, India does not yet have a law specifically targeting deepfakes. Instead, it relies on broader provisions under the Information Technology Act, 2000, which addresses offences like identity theft, impersonation, privacy violations, and obscene content through sections such as 66C, 66D, 66E, and 67/67A. Additionally, the Bharatiya Nyaya Sanhita, 2023 deals with related issues like misinformation and cyber offences. While these laws provide some coverage, they are not tailored to deepfakes, leading to a more indirect and case-by-case approach compared to South Korea's targeted framework.¹⁷

¹⁶Reuters, 'South Korea to Criminalize Watching or Possessing Sexually Explicit Deepfakes' (CNN27 September 2024) <<https://edition.cnn.com/2024/09/26/asia/south-korea-deepfake-bill-passed-intl-hnk>>.

¹⁷ IJLLR Journal, 'Deepfake Pornography and Criminal Law of India: The Crises of Legal Classification' (IJLLR

Comparative analysis

If we compare the two regimes to each other directly, one can see straight away that one of the systems is far more reactive in the immediate sense. South Korea, however, has looked for and South Korea's position advocates a targeted approach. Under the Act on Special Cases Concerning the Punishment of Sexual Crimes, it is a criminal offense to create, disseminate, or even view deepfake pornographic content. According to the report, the legislation effectively outlaws deepfake pornography from inception to passive viewing. This enables easier enforcement and serves as a deterrent. Those who violate the law will incur high costs which are surely a deterrent. Factors supporting it are enhanced policing and more rapid take-downs. However, according to this legislation, mere viewing of deepfake pornographic content is also a crime, which could be excessive. In the case of India, we see a more holistic approach. No laws or regulations exist for dealing with deepfakes at present. Nevertheless, the Information Technology Act 2000 and the Bharatiya Nyaya Sanhita 2023 contain provisions for related offences like impersonation, obscenity, cheating, and publication of false information, among others. As per reports, there are no specific laws to deal with damages arising out of malicious deepfakes due to non-consensual pornography, political misinformation, fraud, and such activities. As a result of these broader laws, there is ample flexibility.

International Response

United Nations officials say they've noticed a rise in AI-assisted sexual extortion schemes used to target children worldwide. Apparently, these criminal actors now use "synthetic explicit images," otherwise popularly known as deepfakes, to extort minors and cause noticeable worry about technology-facilitated abuse. It's notable how perpetrators can run transnational operations as a criminal enterprise.

UNICEF and the International Telecommunications Union (ITU) are calling for an immediate and holistic reassessment of tech companies' safety systems. Improving data protection practices may go a long way toward preventing threats. Governments also need to implement stricter regulations ensuring platform safety for children, while appreciating that the greatest policies can stumble in actual execution. Protecting future generations will take all of digital to accomplish, not just mom-and-pop warning labels.

The initiative, coordinated by the World Standards Cooperation a collaborative entity

comprising three international organizations: the International Electrotechnical Commission (IEC), ISO, and the International Telecommunication Union (ITU) convenes standards developers alongside technology leaders, policymakers, researchers, and civil society representatives. Its mandate includes a prompt response to existing risks presented by deepfakes, misinformation, and the malicious use of synthetic content, concurrently with the objective of maximizing artificial intelligence's societal benefits.

The purpose of the initiative is to build a new model of digital rights based on an inclusive and future-oriented system of transparency, accountability and ethical innovation. The information sharing collaboration is establishing an international standards ecosystem where creators and consumers can trust what they share, see, and hear. In addition, work contributes to the targets set out in the Global Digital Compact a roadmap recently adopted by the UN General Assembly and multiple stakeholders with an aim to maximise public benefit from AI and other technologies for humanity.¹⁸

The United Nations has expressed its support for technical specifications like C2PA through other aspects of its agenda, such as providing assurance for reliable and AI-responsible digital ecosystems and combating the problems caused by online disinformation and large-scale synthetic-media abuse. The UN encourages international standards bodies and industry consortia to support verifiable mechanisms for the provenance of content that will help protect human rights, democratic processes, and public confidence in digital information.

C2PA, or the Coalition for Content Provenance and Authenticity is an open, industry-led standard that binds tamper-evident “content credentials” to digital media (including images, videos, audio files and documents) to allow platforms and users to verify who created the content, when it was created and whether it has been edited or generated by AI. The credentials travel within the body of the file itself and can be verified by any compliant tool, establishing a cryptographically-secured chain of provenance to differentiate authorized content from deepfake or manipulated content, or non-consensually disseminated media.¹⁹

Founding members like Adobe, Arm, BBC, Intel, Microsoft, Sony and Truepic have been joined by an increasing number of well-known tech and media companies including Google, Amazon (and AWS), Meta, OpenAI, Publicis Groupe as a brand marketing agency with very close ties to the social media space as well as journalism; Reuters and The New York Times.

¹⁸ Angela Smith, ‘Standards and Policy Considerations for Multimedia Authenticity - ITU’ (*ITU* 11 July 2025) <<https://www.itu.int/hub/2025/07/standards-and-policy-considerations-for-multimedia-authenticity>> accessed 3 May 2026.

¹⁹ Grant Hatchimonji, ‘What Is the Coalition for Content Provenance and Authenticity (C2PA)?’ (*WhatIs2025*) <<https://www.techtarget.com/whatis/definition/Coalition-for-Content-Provenance-and-Authenticity-C2PA>>.

Camera-maker members including Canon, Nikon, Leica and Fujifilm are implementing C2PA-based signatures in their equipment for users to activate, while social and cloud platforms (LinkedIn and TikTok among those exploring integration) are establishing workflows compatible with C2PA to identify and validate content that is AI-generated or edited.²⁰

On the other hand, G20 South Africa 2025 has graduated from singular national nerves to one large, coherent taper around our shared and distinct acknowledgement that synthetic media can now be a source of electors, stable beginnings and individual safety. In a 2025 Issue Brief to the G20 Digital Economy Working Group, we frame deepfakes as high fidelity, AI-generated audio-visual content especially faces and voices produced with malicious intent (that is, for deceptive purposes aimed at damaging reputations or crowding out trustworthy information). Framing it this way changes the regulatory framing away from “is it fake?” to that of “in what context is it being used, and to what end?” which makes it a good candidate for bringing attention to harms like election-related disinformation, financial fraud, and non-consensual sexual imagery.

The brief is not a binding treaty in itself but serves as a policy-preparation tool that outlines the architecture of an eventual governance regime. It calls on members to develop foresight-based systems for observing deepfake risks, such as mapping attack- vectors surrounding elections, financial scams and child-sexual-exploitation-linked content. In particular, during times of extreme sensitivity like, say, elections or national emergencies the goal is to shift from reactive content moderation (from takedowns) towards proactive early detection, early warning and coordinated counter-narratives. This logic sees deepfakes not as one-off incidents but as a continuous, changing threat layer on digital infrastructure.

The G20 could devise a triad of options to deal with deepfakes: 1) a system for notifying each other when deepfakes surface during elections or around national events, so they can respond to those with counter-messaging and co-ordinate platform interventions; 2) a mechanism for pooling evidence on deepfake incidents that enables countries’ domestic legislation to be aligned with global patterns of misuse; and 3) calling for countries to adopt national rules about licensing high-risk AI models used for developing deepfakes so that platforms have to label and be transparent about ads, and face penalties for harmful deepfakes..²¹

²⁰ ccraigross, ‘C2PA Founding Press Release – Coalition for Content Provenance and Authenticity (C2PA)’ (C2pa.org2021) <<https://c2pa.org/c2pa-founding-press-release/>>.

²¹ ‘Priority 4: Resource Produced for G20 DWEG on generative AI and its evolving ability to produce high quality deepfakes at low cost Priority 4: Equitable, Inclusive and Just AI ISSUE BRIEF: Plumbing the Depths of Deepfakes: What’s at Stake, What’s to Be Done’ (2025) <https://www.dcdt.gov.za/images/DEWG_Issue_Brief_on_Generative_AI_and_its_evolving_ability_to_produce_high_quality_deep_fakes.pdf>.

Challenges for Law and Enforcement

When we put the two systems head-to-head what you see is that which country has which response to the issue. In the case of South Korea, they have put in a very detailed and strict regulatory framework. Their laws which include the Act on Special Cases Concerning the Punishment of Sexual Crimes which very precisely defines what is and is not deepfake porn and also what the penalties are for it plays out over the whole of the content's life from creation to view. That clarity in the law in turn leads to more uniform enforcement, and the very serious penalties put forward a strong deterrent. Also, authorities there have taken a proactive role in fielding special teams and putting in fast processes for taking down content which in turn brings quick relief to victims. However, this stringent approach, particularly the criminalization of private viewing, can elicit concerns regarding the demarcation between regulatory oversight and individual liberty.

In that regard India has a large yet broad legal framework. We see that laws like the Information Technology Act of 2000 and the Bharatiya Nyaya Sanhita of 2023 include what used to be separate issues of impersonation, sexual content, fraud and the like. This which may in fact cover a wider range of issues beyond that of the sexual doctored deepfakes to include financial fraud and misrepresentation. At the same time, it is also true that which these pieces of legislation were not designed with deepfakes in mind which has created some gaps in terms of issues of consent, political interference, and attribution of AI generated material. The enforcement measures can be long and inconsistent owing to reliance on general police systems, the enforcement challenges due to foreign-hosted content and concerns related to the constitutionally-guaranteed right to freedom of speech under Article 19. To sum up, Korea's model has a significant deterrent effect, and a victim-centred approach. India's approach is more flexible but less concrete and it is not applied consistently. Deepfakes have no boundaries and both countries share this problem.

Because the deepfake maker can be anonymous, offshore servers can be used, and the technology of deepfakes is rapidly evolving, it becomes very difficult for one country to deal with it independently. As a result, increasing international cooperation and the inclusion of technical preventative measures, such as watermarking, are viewed as more important than ever.

Recommendations

A balanced approach would integrate more precise legal provisions with practical technological safeguards. From a policy stand point it is up to countries like India to put in place very specific provisions within present structures which include the Information Technology Act of 2000 and the Bharatiya Nyaya Sanhita of 2023 which in turn will address the issue of deepfakes which are used in a malice filled way to violate consent or to cause intentional harm. Also, at the same time we must see to it that which victims have access to easy redressal which includes speedier takedown of content, easy to use complaint systems, and also for which we put in place compensation which in turn will reduce the burden of the individual to navigate through complex systems. Also, on the tech front we may see large scale benefit from global standards. For instance, we may put in place water marks for AI generated content which will help in tracing its origin, also we may see to it that we have much stronger identity verification tools which may include live checks within KYC systems—could curtail misuse by authenticating genuine individuals rather than manipulated imagery or video. Cumulatively, these measures would not only penalize transgressions but also contribute to their prevention, thereby establishing a system that is both responsive and forward-looking.

Awareness among people

Most people learn about fake videos through public programs that teach how to recognize tricks hiding inside digital content. A class from MIT, available at no cost, reached more than ten thousand participants starting in 2022. Instead of just explaining ideas, it shows real signs - such as blinking much less often or mismatched speech movements. When tested, those who trained were between a quarter and third less likely to pass fakes along. Some tools turn learning into games, encouraging users to check multiple sites before believing a story. This habit helps especially during elections, when false clips jumped fivefold across world votes in 2024.

Most people find it hard to judge if online info sounds believable. In the UK, seven out of ten grown-ups have trouble spotting false claims.²² Quick five-minute tests help some get better at noticing tricks used in misleading posts. These mini lessons teach how to check who made a message and why. Across Europe, half a million folks joined training meant to slow fake news. By 2025, those efforts cut sharing of lies by nearly one fifth. Classes came in many languages,

²² Government Office, 'Deepfakes and Media Literacy' (GOV.UK27 May 2025) <<https://www.gov.uk/government/publications/deepfakes-and-media-literacy/deepfakes-and-media-literacy>>.

plus tools let users flag suspicious clips. On sites such as YouTube, new marks appear on videos to signal origin. When more understand media clues, there are forty percent less hoaxes spreading fast.²³

Starting with people, not tools, helps cut through confusion when fake videos mislead millions. One path shows how teaching villagers to spot tricks can reduce false stories by almost half. In India, a project reached fifty thousand locals using games and sound samples that mimic real voices. Progress jumps when teams practice responses before problems hit hard. Training workers like actors running drills leads to sharper reactions under pressure. Schools add value too, lifting detection skills just enough to matter. Clear rules plus open methods make knowledge stick better than tech alone. When researchers, businesses, and officials share steps forward, updates flow faster. Machines change fast; trust stays only when humans stay in charge. Staying alert means adjusting without waiting for perfect answers.²⁴

Conclusion

In conclusion, the problems posed by deepfakes and morphing technologies are in many ways a reflection of the power and danger of AI in the digital age. While the tool itself has many legitimate uses from entertainment to education to innovation, its misuse in the form of AI-based crime has dangerously stretched the limits of the law. Deepfake-enabled crime is no longer limited to petty deception, and has been found to affect personal identity, financial security, public trust, and various democratic institutions and processes. The victim can suffer reputational damage, emotional and psychological trauma, privacy violation, loss of money and property. Clients and service providers, companies and governments now face the challenge of preserving credibility beyond national borders. Harmful content flows from country to country in seconds, and the victims find little recourse. In some countries such as South Korea, specific laws governing the creation and usage of deepfakes and AI-assisted videos now exist, and other countries stare at using their laws against cybercrime to fight the harms of synthetic media.

Though, the current legal system is unable to identify true/synthetic content quickly enough because true/synthetic content is changing too fast and too much to keep pace with the continued emergence of true/synthetic content on different websites. The failures then can be

²³‘EDMO Guidelines for Effective Media Literacy Initiatives EDMO Guidelines for Effective Media Literacy Initiatives’ (2024) <<https://edmo.eu/wp-content/uploads/2024/10/EDMO-Guidelines-for-Effective-Media-Literacy-Initiatives.pdf>>.

²⁴‘Media Information Literacy Initiatives’ (Digital Empowerment Foundation, DEF2023) <<https://www.defindia.org/media-information-literacy-initiatives/>>.

seen through the failure to define laws, the lack of protection for victims of crime, the lack of clarity with regards to international cooperation on these types of issues, etc. Watermarking, tracking, provenance checks, determine whether or not the content is "live" at the time of a particular event, and using AI-based forensic tools to analyze physical evidence will help assist in fighting back against true/synthetic content; however, in order for that to happen these types of technologies need to be updated faster as laws change. The results indicate from the findings to date that the need to fight back against AI-based criminal activity requires all four elements: legal, technical, public awareness, and international collaboration.

