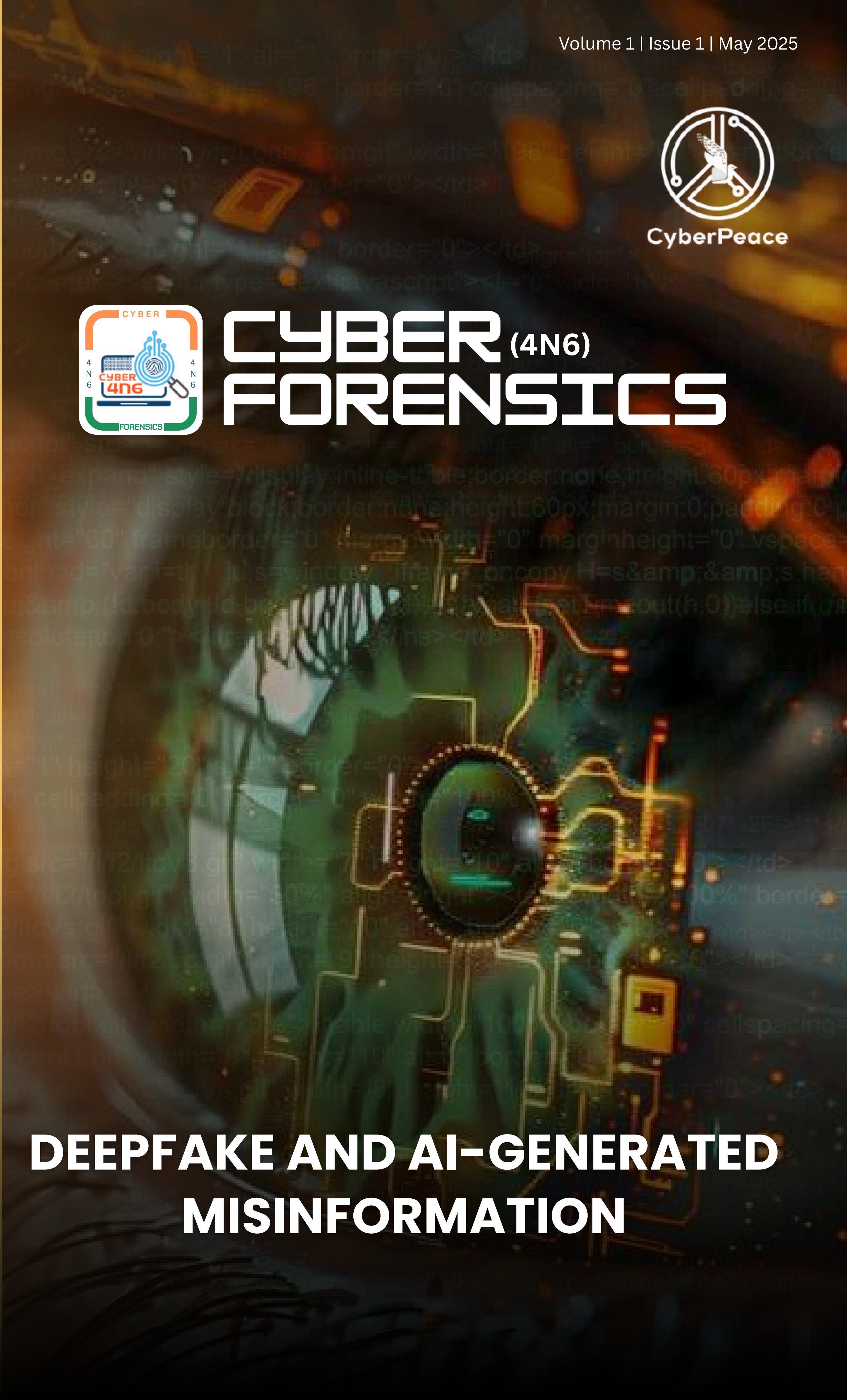
CYBER 4N6 CYBER 4N6 CYBER 4N6 CYBER 4N6 CYBER 4N6

Volume 1 | Issue 1 | May 2025

CyberPeace

CYBER (4N6) FORENSICS



DEEPFAKE AND AI-GENERATED MISINFORMATION

CYBER (4N6) FORENSICS

OUR SUPPORTERS

SHAH & ANCHOR

CyberPeace
CENTER OF EXCELLENCE

Army Institute of Technology, Pune
Onward to Glory

PARTNERS

Autobot Infosec

SysTools®
Simplifying Technology

4N6
Cyber 4N6 Consultants

OUR TEAM

OUR FOUNDERS

Lt. Col. (Dr.) Santosh Khadsare (Retd.)

Chief Business Officer-Cybersecurity, SysTools

Maj. Vineet Kumar

Founder and Global President CyberPeace

OUR MENTORS

Lt Gen (Dr.) Rajesh Pant (retd),
former NSCS, Govt. of India

Mr. MAKP Singh,
former Ministry of Power, Govt. of India

Dr. Manish Prateek,
Professor and Pro Vice-Chancellor,
DBS Global University

TECHNICAL COMMITTEE

Dr. Gaurav Gupta,

Ministry of Electronics and Information Technology

Dr. Nilakshi Jain,

HoD, Shah &Anchor Kutchhi Engineering College, Mumbai

Dr. Nilay Mistry,

Associate Professor, National Forensic Sciences University, Delhi

Col. Harkamal Sidhu,

Vice President, Autobot Infosec

EDITORS-IN-CHIEF

Lt. Col. (Dr.) Santosh Khadsare (Retd.)

Chief Business Officer-Cybersecurity, SysTools

Maj. Vineet Kumar

Founder and Global President CyberPeace

COPYEDITORS

Ms. Sharisha Sahay

Research Analyst, Policy & Advocacy, CyberPeace Foundation

Ms. Ayndri

Research Analyst, Policy & Advocacy, CyberPeace

DESIGN

Ms. Tannu Priya
Concept & Design

Associate- Media & Design, CyberPeace Foundation

Mr. Satyam Singh

Associate- Media & Design, CyberPeace Foundation

HEAD OUTREACH

Lt Cdr Seema Gupta (Retd.)

Head, Outreach Program & Admin , CyberPeace

Editor's Note

Dear Readers,

Welcome to the May 2025 edition of ***Digital Forensics (4N6) Magazine***. As the global conversation around artificial intelligence (AI) continues to evolve, so too do the threats it can inadvertently enable. This issue focuses on one of the most urgent and complex challenges emerging at the intersection of AI and cybersecurity: Deepfakes and AI-generated misinformation.

In today's digital era, manipulated content is no longer a distant concern. It is a tangible threat undermining truth, trust, and societal cohesion. Deepfakes, powered by generative AI, are now capable of fabricating hyper-realistic audio and video content, making it increasingly difficult to distinguish between fact and fiction. Whether it's political deception, financial fraud, or reputational sabotage, the consequences of this technolog ical misuse are profound and far-reaching.

At CyberPeace, we have long advocated for ethical AI and multistakeholder collaboration as pillars of a resilient digital ecosystem. The articles featured in this edition reflect that ethos. They delve into the forensic, legal, ethical, and social dimensions of misinformation—from detecting deepfakes and secur ing public discourse to defining governance frameworks that ensure accountability and transparency.

As the field of digital forensics continues to expand, so too does the range of voices and disciplines engaging with its challenges. In this issue, you will notice a blend of writing styles — from research-driven academic papers to practitioner insights and thought pieces. This diversity is both intentional and reflective of our commitment to building an interdisciplinary platform that bridges theory, policy, and field practice. By welcoming both formal research and accessible essays, we hope to foster dialogue across communities: between students and professionals, between policymakers and technologists. While each piece varies in tone and structure, all share a common thread — a deep engagement with the real-world challenges of cybersecurity and digital forensics.

What's clear is this: digital forensics is no longer a niche discipline. It is at the forefront of global digital safety. By equipping law enforcement agencies, policymakers, technologists, and citizens with advanced tools and knowledge, we can work together to curb the weaponization of emerging technologies.

I encourage all readers, whether you are a student, practitioner, policymaker, or researcher, to engage deeply with this issue. Your participation in this dialogue is crucial to building a cyberspace that upholds peace, truth, and democratic values. Together, let's reaffirm our commitment to digital integrity in an age shaped by artificial intelligence.

▶ **Maj Vineet Kumar**

Founder and Global President, CyberPeace

Protect YOUR BUSINESS *From* FRAUD

Our Services

Autobot Infosec's suite of services offers a comprehensive approach to cybersecurity, covering various domains to address the increasing complexity and variety of cyber threats that organizations face today.

- Cyber Threat Intelligence & OSINT
- OT, IoT, and ICS Security
- Cyber Compliance & Risk Management
- Security Awareness Training & Certifications
- Cyber Insurance
- Incident Response
- Trust and Safety Services
- Cyber Quick Reaction Team

Our Products

Autobot Infosec's range of cybersecurity products is designed to provide robust solutions for both preventive and reactive security measures, ensuring that organizations can effectively guard against cyber threats.

- Threat Intelligence Sensors
- DDoS Protection
- Cyber Range
- Proactive CSPs Shield

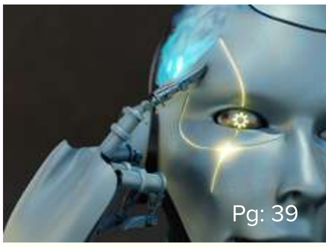
AUTOBOT INFOSEC
YOUR PARTNER IN
RESILIENCE AND GROWTH

WWW.AUTOBOTINFOSEC.COM

INFO@AUTOBOTINFOSEC.COM

TABLE OF CONTENTS

TABLE OF CONTENTS



Deepfakes in India: Psychological Impact on Elections and Social Repercussions

Raquib

B. Tech student (Artificial Intelligence and Machine Learning) at Netaji Subhash Engineering College, Kolkata.

Understanding Deepfakes

Deepfake technology creates incredibly realistic-looking fake audio, video, or photos by using deep learning algorithms. These artificial Intelligence (AI)-generated manipulations can mimic speech, change face expressions, and show unrealized occurrences. Due to the availability of deepfake technologies, anyone may now create realistic forgeries, even those who are not specialists. As per the Press Trust of India's 2023 report, Over 50% of internet users in India get their news online. Such material can spread quickly on social media. This broad reach makes deepfakes more likely to deceive sizable audiences. Research on the 2024 Indian elections, for instance, found that political parties targeted the country's sizable demographic of over 968 million registered voters with AI-generated movies.

Psychological Mechanisms of Deepfake Deception

To look real, deepfakes take advantage of human cognitive biases and heuristics. Because visual and auditory clues are inherently trusted by people, altered videos appear convincing. According to research, being exposed to deepfakes makes people more skeptical and doubtful about news. Ahmed S. in 2021 discovered that those who saw deepfake political films were considerably less trusting of news from social media. Confidence in reliable journalism may be weakened by this decline in trust and increase in cynicism. Furthermore, motivated reasoning might strengthen deepfake believability because, particularly in politically heated situations, people are more inclined to believe false information if it supports their preconceived notions. The 2024 Indian election provided a clear example of these vulnerabilities. One of the deepfakes that went popular included opposition leader Mamata Banerjee giving a fake speech, while another featured Prime Minister Narendra Modi dancing to a Bollywood tune. Both of the clips were actually AI-generated.These illustrations show how emotions and perceptions can be manipulated by deepfakes. The Mamata deepfake was taken seriously by officials, who warned that the Banerjee tape "could affect law and order" and looked into its source. Modi himself referred to the video of him as "a delight," taking it as a lighthearted spoof. Such incidents show that deepfakes can create vivid emotional impressions on voters (through humor or outrage) that bypass critical scrutiny, especially under the time pressure of an election campaign.

Impact on Public Opinion and Elections

Deepfakes can significantly shape public opinion by

spreading false narratives that gain traction before they can be debunked. In India’s 2024 election, political campaigns reportedly used AI-generated deepfakes to influence nearly 970 million voters. For example, a deepfake video of Bollywood stars Ranveer Singh and Aamir Khan purportedly campaigning for a particular party led to public outcry and police action. The Maharashtra Cyber Crime Cell filed an FIR against the creator of the Ranveer video, and Aamir Khan himself reported the fraudulent video to authorities. Similarly, another doctored clip claimed to show the Home Minister making a controversial statement; this “deepfake morphed video” prompted Delhi police to arrest suspects and issue notices to political figures involved. Public figures have to spend time and resources disproving fake content since these cases illustrate potential reputational damage. Deepfake incidents in elections have become common all over the world. In Slovakia’s 2023 election, an AI-generated audio recording circulated on social media impersonating a liberal candidate discussing election rigging. In Nigeria’s February 2023 elections, a manipulated audio clip falsely implicated a presidential candidate in plans to manipulate ballots. These examples highlight how deepfakes can be used in smear campaigns to influence electorates. They underscore that the threat is worldwide: as one analysis notes, a single malicious clip can instill fear that the election could be rigged. Social media reactions further reflect these concerns. For instance, an online analysis noted that homemade deepfake videos could reduce viewers' willingness to vote for targeted politicians, although not always more effectively than simple text misin formation. In general, experts warn that deepfakes can create "false realities" and exacerbate political polarization. As one commentator puts it, AI-driven media "undermines election integrity" by sowing confusion and fear. Together, these insights suggest deepfakes do more than mislead individuals, they can polarize societies by driving public discourse into false narratives.

Erosion of Trust in Media and Institutions

Deepfakes contribute to a broader erosion of trust in media and institutions. When distinguishing real from fake becomes difficult, people may grow skeptical of all information sources. In India, where misinformation has long been an issue, this deepfake anxiety adds fuel to existing distrust. Scholars describe a "liar's dividend" dynamic: individuals who commit wrongdoing can simply claim "it was a deepfake" to evade accountability. A recent case in Tamil Nadu exemplifies this: a politician released an audio clip allegedly implicating an opponent, who then dismissed the recording as an AI fabrication. Rest of World reports this as "the first high-profile case of the 'liar's dividend'" in India. In other words, even genuine evidence can be cast into doubt.

The growth of deepfakes is staggering: one analysis found the number of deepfake videos online rose by 550% since 2019. Most of these are misleading or malicious (for example; over 96% target women). This proliferation means that every suspicious video now carries the taint of possible fakery. As a result, political trust suffers: voters may ignore legitimate campaign messages or even credible news, worried it might be synthetic. In India's context, such widespread doubt undermines faith in the electoral process itself.

Challenges to Digital Investigations

Deepfakes pose serious challenges for digital investigations and cybersecurity. When video or audio evidence can be easily fabricated, it complicates legal procedures. A 2024 systematic review found that deepfakes "significantly threaten the criminal justice system," highlighting issues like evidence falsification and the erosion of trust in institutions. Criminals or hostile actors could use deepfakes to sabotage investigations by introducing fake evidence or impersonating officials. For example, deepfake audio could potentially bypass voice authentication or create false confessions. In India, where law enforcement is already stretched during elections, the ease of creating deceptive content raises fears of new avenues for cyberattacks and disinformation campaigns that could undermine national security. Security experts warn that current forensic tools may lag behind synthetic media, necessitating upgraded authentication protocols.

Countermeasures and Future Solutions

Addressing the deepfake threat requires legal, technological, and educational strategies. India has already implemented some measures. In early 2024, Meta (formerly Facebook) partnered with Indian fact-checkers to launch a Deepfakes Analysis Unit (DAU) accessible via WhatsApp. Citizens can send suspicious videos or audio to a helpline number, where experts verify and debunk them. The DAU's WhatsApp channel provides official verdicts on circulating content. This grassroot verification tool exemplifies the kind of public resource needed to restore trust. On the regulatory front, India may look to global models. The European Union's Digital Services Act (DSA) now mandates that online platforms mitigate AI-generated disinformation during elections. Such legislation holds social media companies accountable for malicious content. In the U.S., some states have passed specific laws: for example, Minnesota's House File 1370 requires disclosure if an election ad contains materially deceptive deepfake media. India could adopt similar rules for political advertisements. These regulations, if coupled with penalties for malicious actors, would balance free speech with the need for transparency. Technological innovations are also crucial. Advanced detection algorithms, such as those analyzing facial movement inconsistencies or audio signature are being developed and could be deployed by platforms to flag likely fakes. Researchers have proposed using blockchain and cryptographic techniques to protect media authenticity. For instance, content creators could timestamp and hash genuine videos on a public ledger; any later alteration would break the cryptographic signature, revealing forgery. One proposal is to "anchor" campaign videos on-chain so that only the original version is considered valid. In practice, this could involve watermarking media at the source or having news outlets register their footage in distributed databases. By leveraging blockchain's immutability and transparency, platforms might automatically verify whether a video or image matches an original registered copy.

Ethical Considerations and Governance

Ethical guidelines emphasize that AI-generated content must be handled transparently. Organizations like IEEE have called for clear labels on synthetic media and public education about AI risks. In India, media literacy campaigns and voter education can empower citizens to question unverified content. Teaching basic verification skills, such as reverse image searches or source checking is essential. At the same time, technology companies must build tools that nudge users to check doubtful content.

Conclusion

Deepfakes pose a significant psychological and societal threat to India's democracy. By exploiting cognitive biases, they can distort public opinion, damage reputations, and sow widespread doubt. India's 2024 election already saw real instances of this danger, from viral AI videos to police cases over doctored clips. The resulting erosion of trust, both in individuals and in media, risks undermining electoral integrity. However, India is not defenseless. Legal measures like the Deepfakes Analysis Unit and potential content regulations can deter misuse, while technological solutions such as blockchain verification and AI detectors offer promise. Crucially, ethical and educational efforts must complement these steps. Only by combining robust regulation, innovation, and public awareness can India mitigate the deepfake challenge and protect its democratic processes.

▶ **Raquib**

B. Tech student (Artificial Intelligence and Machine Learning) at Netaji Subhash Engineering College, Kolkata.

About the Author:

Raquib is a third-year B.Tech student specializing in Artificial Intelligence and Machine Learning at Netaji Subhash Engineering College, Kolkata. Passionate about the limitless possibilities of AI, he is particularly intrigued by how intelligent systems can be integrated into everyday objects—even something as simple as a toothbrush. To learn more about his work and interests, he can be reached on LinkedIn: https://www.linkedin.com/in/raquib223/
GitHub - https://github.com/rex223.

References

- Ahmed, S. (2021). *Navigating the maze: Deepfakes, cognitive ability, and social media news skepticism. New Media & Society, 25*(5), 1108–1129. researchgate.net

- Channel News Asia. (2024, May 16). *Dance videos of Modi, rival turn up AI heat in India election*. https://www.channelnewsasia.com/ (access via [6])

- Hedrih, V. (2025, April 30). *Homemade political deepfakes can fool voters, but may not beat plain text misinformation*. PsyPost. https://www.psypost.org/2025/04/homemade-political-deepfakes-can-fool-voters-but-may-not-beat-plain-text-misinformation-74612 psypost.org

- Mirza, R. (2024, February 16). *How AI-generated deepfakes threaten the 2024 elections*. Journalist's Resource. https://journalistsresource.org/home/how-ai-deepfakes-threaten-the-2024-elections journalistsresource.org

- Nover, S. (2024, May 21). *Dreams of a dancing Modi*. GZERO Media. https://www.gzeromedia.com/gzero-ai/dreams-of-a-dancing-modi gzeromedia.comgzeromedia.com

- Press Trust of India. (2023, May 5). *Majority of internet users in India consume news online: Report*. Mint. https://www.livemint.com/news/india/majority-of-internet-users-in-india-consume-news-online-report-11683244712584.html livemint.com

- Ramirez, D., & Andrada, C. (2024, September 26). *70 deepfake statistics you need to know (2024).* Spiralytics. https://www.spiralytics.com/blog/deepfake-statistics/ spiralytics.com

- Rest of World. (2023, July 5). *Indian politician blames AI for alleged leaked audio*. https://restofworld.org/2023/indian-politician-leaked-audio-ai-deepfake/ restofworld.org

- Reuters. (2024, April 22). *Deepfakes of Aamir Khan and Ranveer Singh raise worries over AI misuse in Lok Sabha election 2024*. Hindustan Times. https://www.hindustantimes.com/india-news/deepfakes-of-aamir-khan-and-ranveer-singh-raise-worries-over-ai-misuse-in-lok-sabha-election-2024-101713762883265.html timesofindia.indiatimes.comtimesofindia.indiatimes.com

- Sakunia, S. (2024, July 12). *AI and Deepfakes played a big role in India's elections*. New Lines Magazine. https://www.newlinesmag.com/spotlight/ai-and-deepfakes-played-a-big-role-in-indias-elections newlinesmag.com

- Sandoval, M.-P., de Almeida Vau, M., Solaas, J., & Rodrigues, L. (2024). Threat of deepfakes to the criminal justice system: A systematic review. *Crime Science, 13*, 41. https://doi.org/10.1186/s40163-024-00239-1 crimesciencejournal.biomedcentral.com

- Struck Capital. (2024, January 17). *Deepfakes and blockchain*. https://struckcapital.com/deepfakes-and-blockchain/ struckcapital.com

- Times of India. (2024, April 18). *Deepfake video of Ranveer Singh criticizing government goes viral: Watch the original clip*. https://timesofindia.indiatimes.com/technology/social/deepfake-video-of-ranveer-singh-criticising-government-goes-viral-watch-the-original-clip/articleshow/109406802.cms timesofindia.indiatimes.comtimesofindia.indiatimes.com

- Times of India. (2024, May 3). *Amit Shah fake video case: Delhi police arrests actor of 'Spirit of Congress' X account*. https://timesofindia.indiatimes.com/india/amit-shah-fake-video-case-delhi-police-arrests-arun-reddy-handler-of-spirit-of-congress-account-on-x/articleshow/109820820.cms timesofindia.indiatimes.com

- Wee, C. K. G. (2024, May 10). *Artificial illusion: Global governance challenges of deepfake technology*. IAPP. https://iapp.org/news/a/artificial-illusion-global-governance-challenges-of-deepfake-technology iapp.orgiapp.org

- World Economic Forum. (2024, March 1). *Meta is launching a deepfake helpline in India*. https://www.weforum.org/stories/2024/03/ai-deepfake-helpline-india (access via [68])

The Rise of Deepfakes: Impact, Origins, and Technical Insights

Dr. Asheesh Tiwari, Kushagra Varshney, Rahul Dixit, Arjun Chauhan

Introduction

In this era of Artificial Intelligence (AI), deepfakes have become a topic of discussion. While deepfake technology can be used for creating real-life effects in movies and videos, enhancing images, VFX, and as a successful educational tool, it can also be used to manipulate media to make people appear to say or do things they never actually did. It has increasingly become a tool for political manipulation, fraud, and harassment. Criminals are using deepfake images & videos to trap, impersonate and blackmail people.

Using AI and machine learning techniques, deepfakes are also jeopardizing the trust and authenticity of the proof and evidence. Previously, we used to see something, and we used to consider it real. However, with this new technology, it has become harder to distinguish between what is real and what is fake. This creates a dangerous environment where misinformation spreads rapidly, affecting the credibility of information provided by social media and traditional media channels.

In this article, we are going to explore how deepfakes work. We'll look at the negative side of deepfakes and the harm it does. We will also discuss how we can use this technology responsibly.

Impact of Deepfakes on the General Public

Trust Deficit- Deepfakes are changing how people view online content today, and there is increasing confusion about what is real and what is fake. One of the biggest impacts of this is that people have lost trust. Videos of politicians, celebrities, or even normal people are found saying foul things and using improper images of them. This makes it harder for people to identify what is real and what is fake. As a result, it can damage well-being, safety, and reputations.

Harassment- Deepfakes are also used to harass people. By creating fake videos or photos, attackers can blackmail someone without needing evidence. This can cause harm to the reputation & assets of people.

Crime- Another effect is the use of deepfakes for scams and blackmail. Criminals are creating fake videos, voice messages to trap people into giving their important personal information and defraud them of their hard-earned money. To complicate matters, these scams are becoming harder to detect.

Public Awareness- Due to its negative side effects, deepfake technology also compels people to think. People have no choice but to engage more critically with the content they are consuming. Many are learning to verify information before trusting it.

Where It Started & What It Has Become

Deepfake technology appeared around 2017 when a Reddit user shared videos swapping faces using AI. At that time, it was a fun and creative tool for making memes & funny videos. But slowly, people realized how dangerous it has become it is being used for dubious activities. In the early days, most deepfakes targeted celebrities, creating fake videos without their permission, sometimes even in inappropriate ways. As the technology improved, the targets shifted from celebrities to political leaders. Fake videos started surfacing where leaders were seen saying things they never actually said. This created a lot of fear about how deepfakes could influence public opinion and spread false news, especially during elections.

It didn't stop there. Deepfakes started breaching personal lives, too. Normal people became targets for scams, blackmail, and online harassment. In some cases, even companies have been attacked by fake audio messages, leading to financial losses.

What started as a small online experiment quickly breached into politics, finance, media, and personal privacy. Today, it stands as a major threat, showing how powerful and risky AI-based technologies can become if not handled carefully.

Technical Overview : How Deepfakes are Created and Detected

People are making deepfakes using a method of AI called deep learning. They use models like autoencoders, GANs (Generative Adversarial Networks), and transformers to create these deepfakes. In simple words, the AI can study thousands of real images to understand how faces look from different angles and directions, in different lights, and with different emotions. When it learns enough, it starts to create new images or videos that look real. The basic idea for creating deepfakes involves two steps. First, the AI model has to learn to extract features like eye shape, skin texture, or smile pattern from real faces. Then, it uses this knowledge to map these features onto another face or recreate them artificially. GANs work by using two neural networks: one tries to create a fake that looks real, while the other tries to catch the fake.

Through this competition, the AI keeps improving until the fakes become extremely convincing. Detecting deepfakes is not easy. People try to use old-gen methods, which involve looking for small errors that can be missed by the human eye, such as unnatural blinking, strange lighting around the face, or inconsistent facial movements. More advanced systems use deep learning models to automatically find these hidden flaws. CNNs (Convolutional Neural Networks) are commonly used because they are great at recognizing patterns in images. Sometimes, transformers — another type of AI model — are also used because they can understand sequences and finer details better.

Training a deepfake detection model usually starts with a dataset that has both real and fake images. These images are resized, normalized, and fed into the neural network. When a model has been trained enough, it learns to detect real from fake by checking differences that are missed by the human eye.

Final Thoughts

Deepfake technology has completely changed how people view online content today. What started as a fun experiment has now turned into a serious concern across the world. It has made people doubt videos, images, and even voices they hear online. Trust, which is very important in daily life, is now easily shaken just because of a fake video or audio clip.

However, deepfakes have also compelled people to think smarter. Many are now learning to cross-check facts and not blindly trust what they see. This shift in thinking is important because it helps in building a stronger and safer online space.

Technically, both the creation and detection of deepfakes are improving fast. AI models like CNNs and transformers are making it possible to catch deepfakes that the human eye can easily miss. But still, it's a race between creators and detectors. Deepfakes are not just a technical problem anymore they are a social and psychological problem too. They teach us that powerful tools need careful use. Moving forward, it is important for everyone, including governments, companies, and people, to stay updated and careful about how these technologies could be used for their good or bad. In a world where seeing is no longer believing, critical thinking becomes our greatest defense.

References

- MesoNet. (n.d.). MesoNet: A deep neural network for real-time detection of deepfakes.
- FakeCatcher. (n.d.). FakeCatcher: A real-time deepfake detection framework.
- Li, X., Wang, Y., & Liu, J. (2019). UADFV and Deepfake-TIMIT: Datasets for deepfake detection. In Proceedings of the International Conference on Machine Learning (pp. 20-25).
- Face X-ray. (2020). Face X-ray: Self-supervised learning for deepfake detection. In Proceedings of the Computer Vision and Pattern Recognition Conference (pp. 1430-1435).
- XceptionNet. (2019). XceptionNet for deepfake detection.

These are the accuracy results of previous deepfake detection models.

Method	Dataset	Model	Claimed Performance
MesoNet [4]	Private web data	CNN	Detection rate: 98%
FakeCatcher [3]	FaceForensics++, Private web data	Traditional operator + CNN	FaceForensics++ accuracy: 96% Private web data accuracy: 91.07%
Artifacts TIMIT (2) [1]	UADFV, Deepfake-TIMIT	CNN	UADFV AUC: 0.974 Deepfake-TIMIT (LQ) AUC: 0.999 Deepfake-TIMIT (HQ) AUC: 0.932
Face X-ray [2]	Celeb-DF, DFDC preview, DeepfakeDetection, FaceForensics++	FCN + Self-supervised learning	FaceForensics++ (Deepfake) AUC: 0.9917 DFDC preview AUC: 0.9540 DFDC AUC: 0.8092 Celeb-DF AUC: 0.8058
XceptionNet [5]	FaceForensics++	Xception Net	Raw accuracy (Deepfake): 99.26% HQ accuracy (Deepfake): 95.73% LQ accuracy (Deepfake): 81.00%

Dr. Asheesh Tiwari

About the Author:

Dr. Asheesh Tiwari is an Associate Professor at GLA University, Mathura, and serves as the Cyber Security Track Coordinator. He holds a Ph.D. in Software Engineering and specializes in cybersecurity and academic leadership, with a strong commitment to advancing research and education in the field.

Kushagra Varshney

About the Author:

Kushagra Varshney is a B.Tech student specializing in Cyber Security & Forensics at GLA University. He is passionate about exploring emerging technologies and building a strong foundation in cybersecurity.

Rahul Dixit

About the Author:

Rahul Dixit is currently pursuing a B.Tech in Cyber Security & Forensics at GLA University. His academic focus lies in developing practical skills in digital security and forensic investigation.

The Deepfake Threat: Fraud, Scams, and Political Deception

Akshay N Shetye, Dr. Nilakshi Jain

With the rise of Artificial Intelligence (AI) to generate digital media content, there is a rise in deepfake content. This is content which portrays something real, but isn't. While it can be used for entertainment, it may also be used for nefarious purposes such as financial fraud, identity theft, and political deception. These deepfake videos, audio, and images can be generated through any advanced deep learning models or available tools that convincingly imitate real people using AI.

As deepfakes become more advanced and widespread, they present an increasing and critical threat that undermines the very pillars of truth, security, and trust in our digital world. Deepfake technology has blurred the line between truth and fabrication, making it increasingly difficult for individuals and institutions to distinguish authentic content from fake one. It is increasingly possible for criminals to diminish public trust, damage reputations, influence elections, and facilitate crimes that were once difficult to execute without physical presence or access. Most of the time, fraudsters impersonate trusted figures in society to manipulate victims. Old fraud techniques like phishing emails, social engineering are now enhanced by deepfake content to hide their intention. A notable incident occurred in 2019 when criminals used AI-generated voice cloning to impersonate the CEO of a UK-based energy firm.

The fake voice ordered an employee to transfer $243,000 to a Hungarian supplier, which was successfully executed before suspicion arose. In one instance, an employee was deceived into participating in a video conference with what they thought were the company's Chief Financial Officer (CFO) and other team members. However, these individuals were not the real people; instead, they were deepfake simulations created using advanced AI. The employee, believing the call to be legitimate, followed instructions from the fake CFO, which led to a $25 million financial loss for the company. In February 2024, a finance clerk at a prominent British multinational based in Hong Kong became the victim of an elaborate scam that involved the use of an AI-generated deepfake.

Since deepfakes allow criminals to create realistic replicas of someone's face, voice, or even video footage, it becomes harder for both individuals and organizations to detect fraudulent activity.

Akshay N. Shetye

About the Author:

*Akshay N. Shetye is a research scholar at Shah and Anchor Kutchhi Engineering College, currently conducting research on Deepfake Technology under the guidance of Dr. Nilakshi Jain. His work focuses on exploring the implications and advancements of deepfake detection and prevention in the evolving landscape of digital media. He can be reached here- LinkedIn - **LinkedIn Profile Link***

Dr.Nilakshi Jain

About the Author:

Dr.Nilakshi Jain is a faculty member in the Department of Cyber Security at Shah and Anchor Kutchhi Engineering College, Mumbai, India. She is dedicated to advancing knowledge in the field of cybersecurity and contributing to academic and professional growth in the domain.

This has become a growing concern for industries that rely heavily on biometric authentication, such as banking, finance, and law enforcement. Criminals can steal someone's identity in real time by creating a video or audio deepfake to impersonate the victim and gain unauthorized access to their accounts or sensitive information. Deepfake technology can be used to create convincing fake IDs or documents by altering photos and personal details on identification cards. Identity theft is a serious concern as it compromises biometric information like voice, face, and facial expression, and this kind of information is unique and permanent, which cannot be reset like a password.

In the financial sector, fraudsters use synthetic media to impersonate executives in insider trading, create fraudulent financial announcements, or manipulate and convince investors by fabricating endorsement schemes. One emerging trend is the use of deepfakes for investment in the stock market. In this type of scam, fraudsters create fake videos or audio clips of company executives making false announcements about mergers, acquisitions, earnings reports, or other market-moving news. These manipulated pieces of media can spread quickly on social media, stock forums, or financial news platforms, leading to sudden and often volatile price swings in stocks.

Fraudsters can manipulate stock prices for their gain, putting both retail and institutional investors at risk. Insurance companies are also at risk from deepfakes, with claimants submitting fabricated videos of accidents or property damage to support fraudulent claims. Many insurance companies use video-based assessments as part of their claim verification process, making it easier for scammers to deceive them with fake but highly realistic videos. Banks and lending institutions face significant risks from loan fraud facilitated by deepfakes. Criminals can use deepfake technology to create synthetic identities to open bank accounts, apply for loans, and gain access to credit. Once an amount is disbursed from the bank, the fraudsters disappear, and financial institutions face a loss.

The most harmful uses of deepfakes is when they are used to create widespread confusion, disrupt politics, and weaken public trust. These fake videos or audio can sway opinions, spark conflicts, and cause instability across the globe. A single deepfake video showing a politician making controversial or inflammatory statements can spread rapidly across social media platforms, causing immense damage to their career or public image, often before the falsity of the content is exposed. In 2018, a deepfake video of former U.S. President Barack Obama was created to demonstrate how easily technology could manipulate speech, showing him saying things he never actually said. While the video was originally meant to educate the public on the dangers of deepfakes, it highlighted the profound potential of such technology to erode trust in media and public figures. In 2022, a deepfake video falsely depicted Ukrainian President Volodymyr Zelensky surrender ing to Russ -ia, an attempt to demoralize the population and undermine trust in leadership. Election interference has also become a major concern, with AI-generated fake endorsements or scandalous clips designed to sway public opinion. Deepfakes can be used to discredit political opponents, spread false information, and manipulate voters' perceptions during elections.

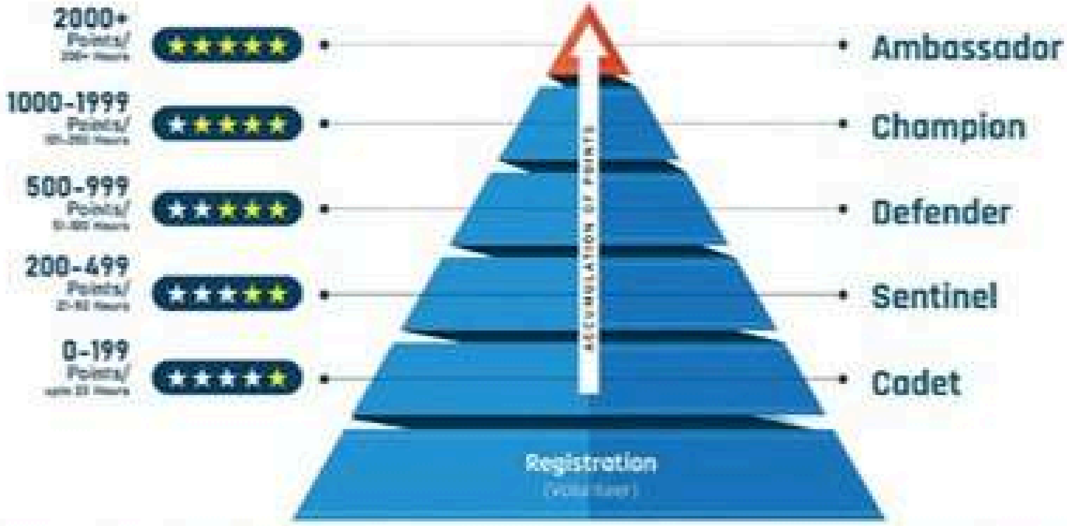
Deepfake detection is like the tech arms race. As deepfake generation evolves, so do mechanisms to detect them. Watermarking and other digital forens -ics methods, such as lighting, blinking incoherent analysis, or aural-visual inconsistency, are continually being developed. Blockchain technology can be used to verify videos and images as being authentic at the source of creation. This can help to demonstrate that the content is original and has not been tampered with. Deepfakes can be a double-edged sword. It is a transformational technology with serious pitfalls. From financial scams to political propaganda, the abuse of deepfake technology highlights the necessity for alertness, cooperation, and reg ulation. What the future holds will hinge on how quickly and effectively we are able to adjust to this new threat. There is an urgent need to invest in detection technologies, the development of public digital literacy, and the establishment of strong laws to protect against the dark side of deepfakes.

Empowering Digital Guardians CyberPeace Corps

CyberPeace Corps (CPC) is a global initiative that unites individuals, cybersecurity professionals, and tech enthusiasts on a mission to fortify our digital defenses. With the ethos of peace, protection, and prosperity, they actively engage in activities that promote cybersecurity awareness. The volunteers come from diverse backgrounds professions and demographics. What unites them is their unwavering dedication to this noble cause and making a meaningful impact in society by.

How to join CPC?

Join as Volunteers by registering at https://cyberpeace.global and grow as an Ambassador through different Levels of Certifications. Our unique point-based system tracks volunteer activities, enabling them to progress through different levels and earn stars with distinctive titles from Cadet, Sentinel, Defender, Champion Ambassador.

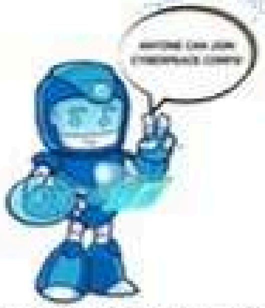

2000+ Points/ 200+ Hours
1000-1999 Points/ 101-200 Hours
500-999 Points/ 51-100 Hours
200-499 Points/ 21-50 Hours
0-199 Points/ upto 20 Hours
★★★★★
Ambassador
Champion
Defender
Sentinel
Cadet
ACCUMULATION OF POINTS
Registration
(Volunteer)

TOGETHER, LET'S BUILD A SAFER, MORE SECURE AND RESILIENT CYBERSPACE

Be a **Volunteer**, Be a **First Responder**

[Join the CyberPeace Volunteers]

Be the frontline defense against cyber threats

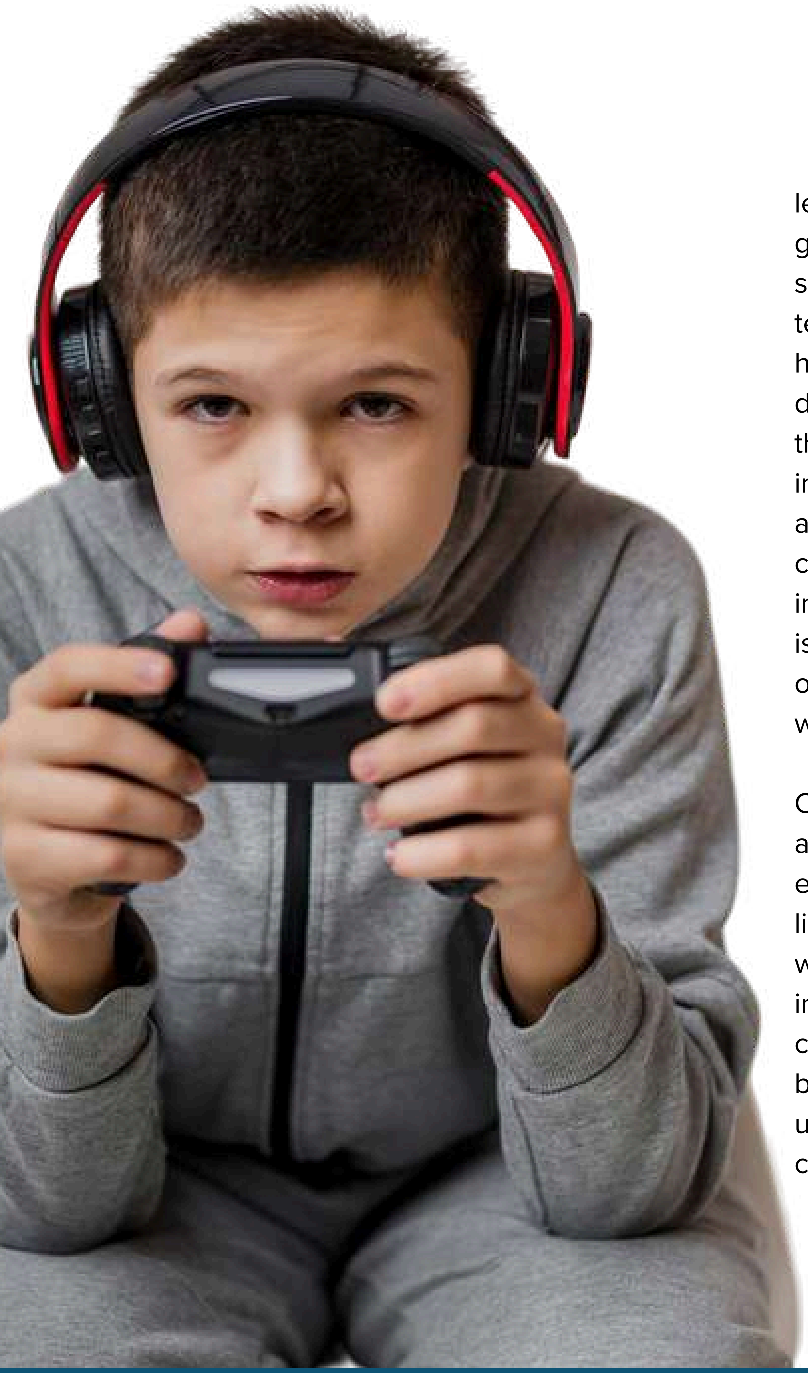


CyberPeace Helpline
+919570000066
www.cyberpeace.global

SCAN the QR code or follow the link to visit

Misuse of Deepfake Technology Against Children on Multiplayer Gaming Platforms: A Growing Concern

Harsha Agrawal



Introduction

Deepfake technology refers to the synthetic content created in assistance with AI and machine learning algorithms by altering images. The technological advancements have led to the creation of such synthetic media through prompts. Deepfake technology uses GAN to create synthetic media which has two neural networks namely a generator and a discriminator. The first neural network which is called the generator creates the manipulated or synthetic image and the second network assesses it against the actual image (Langa, 2021). The two networks work on creating the image continuously till it is almost impossible to detect which image is real and which one is fake. The technology is aggressively used by various online companies and their platforms to attract users without proper regulation. This is a serious concern.

One such use of deepfake technology is creating avatars and characters in online gaming platforms and esports. It provides an environment which gives real-life experience in the virtual world and AI opponents with adaptive learning making these games much more interesting. Multiplayer gaming platforms allow users to chat and interact during the game not just with friends but with other players around the world exposing their users to cybercrimes. These games have the ability to connect, chat and organise multi-player competitions or tournaments with real cash prizes. Unfortunately, the users of such platforms are mostly children below the age of 18 years and deepfakes can be grossly misused if left unregulated.

The author has used the term multiplayer gaming platforms rather than online gaming or esports to widen the scope as it is vital to regulate platforms whe re children are engaged as users whether its online multiplayer games or esports or any other game that is based on skills and uses deepfake technology.

Multiplayer Gaming Platforms and Their Regulation

There are some legislations implemented by the Parlia ment to regulate the culture of gaming in India. The Public Gambling Act, 1867 prohibits public gambling which is often referred to as a game of chance and states that the Act shall not apply to the games that are based on skill. The interpretation of this distinction is important as to determine the legality of these multiplayer gaming platforms. Game of skill implies any game where sufficient knowledge, training and expertise is required whereas a game of chance is mere luck and is based on random factors. In 2018, the Sports Bill was introduced in Lok Sabha to regulate online gaming and betting on online sports. It proposed to establish a commission which would look after the functioning of 'Online Gaming Websites' and track if there is any illegal online sports gaming or any other suspicious activities taking place across these platforms. However, the bill lapsed and another Online Gaming (Regulation) Bill was introduced in 2022 which had similar provisions with regards to the Commission but the bill again was not passed.

In 2023, the IT (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 was amended which imposes certain obligations on the online gaming intermediary and social media intermediaries who enables access to the permissible online real money games such as verification and reporting in case of any obscene material is displayed.

The online gaming self- regulatory body is responsible for taking measures to safeguard children and take appropriate measures. Yet, the threat of misuse of deepfake by other users to manipulate children on these multiplayer platforms is outside the scope of these IT Rules, 2021. Even the IT Rules has made the distinction between games as online gaming and online real money gaming which is based on money and not on skill. Additionally, it emphases more on the intermediaries and their responsibilities and not on other users on those gaming platforms.

Challenges in Regulating Multiplayer Online Gaming Platforms

India's online gaming industry is one of the fastest-growing sectors. Introduction of bills in Lok Sabha and the Amendment of IT Rules in 2023 shows the intention of the Indian government to regulate these gaming platforms, ensure due diligence and content monitoring. Yet the pace of technological advancement in this sector makes it difficult to implement these laws efficiently. Further,

Stringent regulation may discourage startups

Frequent and stringent policy can be detrimental to the startup ecosystems. This would not only impact the startups, innovation but also the economy of the country. The companies from abroad might not be interested to invest in India due to heavy compliances and would act as high entry barriers.

Games aimed at minors/ children may be impacted

One of the most pressing concerns is the impact of misuse of technologies like deepfake against children. Without proper overview, they may be exposed to harassment and sexually explicit content. Multiplayer gaming platforms, especially targeting the users below 18 years old, may have significant impact if stringent policy is implemented to safeguard children. It is important to find the balance between the regulation and boosting economy for these platforms.

The goal to make India a gaming hub would be hampered without regulation

All India Gaming Federation (AIGF) in its consultation paper has highlighted its vision to make India a global gaming hub and recommended setting up of specialised centres and hubs to promote innovation, talent and gaming content. Initiatives like improving the animation, visual effects, gaming and comics related policies are part of this vision. Nevertheless, without regulatory framework and child safety measures, investors around the world may hesitate to invest significantly in the Indian gaming industry hindering its growth and employment generation.

Potential Risks and Uses of Deepfake in Multiplayer Gaming Platforms

The integration of deepfake into gaming is reshaping the user experience and enabling personalised avat-ars and other characters. Bringing the characters to life like animation makes it more engaging. While deepfake technology is shown as a tool for enhancing user experience, these can be often linked to privacy concerns due to data driven personalisation and behavioural analysis. Additionall-y, multiplayer gaming dealing with real money collects financial details of the user which can be misused if there is a data breach. Behavioural profiling of children can also lead to highly targeted in-game purchases which raises concerns about manipulation. The ability to connect and chat with other users around the globe can be misused using deepfake technology. Users with ill intent can clone voices of child's friends to deceive them and trick them into sharing sensitive personal information including passwords, address and financial details. They can use deepfake to appear trustworthy by creating friendly avatars, or impersonate others using face-swap technology to manipulate the users leading to child grooming and catfishing.

Role of DPDP Act, 2023

The Digital Personal Data Protection (DPDP) Act, 2023 plays a crucial role in protection of data on multiplayer gaming platforms. Certain measure can be taken to safeguard children on these platforms as Section 8 and 10 of the DPDP Act lays down obligations on Data Fiduciaries and Significant Data Fiduciaries. According to section 9, processing of personal data of children should only be completed after verified consent of the parent or lawful guardian. Behav ioural monitoring and targeted advertisement at the users below 18 years is prohibited. Nonetheless, DP DP Act has not been enforced till date and these multiplayer online gaming platforms remain unregul ated. Furthermore, the verification of age as mentioned under section 9 remains technically difficult and can be easily bypassed. There is no provision under the DPDP Act to protect children playing these multiplayer online games from predators who may use deepfake technology to groom or blackmail them into sharing personal information through these platforms. DPDP Act, does not expressly deal with issues like face or voice cloning which are an essential component of deepfake threats.

Conclusion

The fusion of multiplayer online gaming and advanced technology like deepfake presents both opportunities and challenges. On the one hand it acts as a tool to enhance gaming experience and boost the economy. On the other, it can be used to expose minors to risks such as impersonation, grooming, harassment, and data manipulation. The present legal framework that regulates gaming platforms remains fragmented. The Public Gambling Act, 1867, IT Rules, 2021 and the DPDP Act, 2023 offer protection to children to a certain extent but need to be uniform and tailored to the complexities that this fusion of gaming and technology possess. Today, it is crucial to adopt a comprehensive law that regulates these online gaming platforms, clearly distinguishes between the game of skill and game of chance and sets clear standards for license and compliances. The DPDP Act, 2023 is a step towards the right direction but it needs to be supplemented with sector specific rules and guidelines for such technological advancements to protect the users in true sense in the gaming ecosystem.

References

- Casemine Editor's Desk. (1996, January 13). *Horse-Racing Classified as a Game of Skill: Landmark Judgment in Dr. K.R Lakshmanan v. State of T.N (1996).* Https://Www.casemine.com; Casemine. https://www.casemine.com/commentary/in/horse-racing-classified-as-a-game-of-skill:-landmark-judgment-in-dr.-k.r-lakshmanan-v.-state-of-t.n-(1996)/view

- Dean Kuriakose, A. (n.d.). *THE ONLINE GAMING (REGULATION) BILL, 2022*. Retrieved May 16, 2025, from http://medianama.com/wp-content/uploads/2022/05/78-of-2022-as-introduced.pdf

- India. (1867). THE PUBLIC GAMBLING ACT, 1867. In *THE PUBLIC GAMBLING ACT, 1867*. https://www.indiacode.nic.in/bitstream/123456789/2269/1/AAA186703.pdf

- (The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021. (2021). In *Official Gazette* (p. 1). https://www.meity.gov.in/writereaddata/files/Information%20Technology%20(Intermediary%20Guidelines%20and%20Digital%20Media%20Ethics%20Code)%20Rules%2C%202021%20(updated%2006.04.2023)-.pdf)

- Langa, J. (2018). *NOTES DEEPFAKES, REAL CONSEQUENCES: CRAFTING LEGISLATION TO COMBAT THREATS POSED BY DEEPFAKES*. https://www.bu.edu/bulawreview/files/2021/04/LANGA.pdf

- *undefined*. (2023). INDIAai. https://indiaai.gov.in/article/ai-development-in-gaming

Harsha Agrawal

About the Author:

Harsha Agrawal is a PhD scholar, NLU Delhi. She can be found on LinkedIn at: Harsha Agrawal - Lecturer - Indian Institute of Management Rohtak | LinkedIn

Deepfakes in Cybercrime & CTFs: The New Face of Digital Deception Similarity

Mukul Joshi

About the Author- *Mukul Joshi is a Computer Engineering student at the Army Institute of Technology (AIT), Pune, and serves as the Joint Secretary of the Information Security & Digital Forensics (ISDF) Club. A passionate cybersecurity enthusiast, he actively participates in national and international Capture The Flag (CTF) competitions, consistently earning top rankings. Mukul is ranked in the top 1% on TryHackMe and specializes in ethical hacking, web exploitation, steganography, cryptography, digital forensics, and vulnerability assessment. He can be reached at LinkedIn https://www.linkedin.com/in/memukuljoshi/*

Introduction

Imagine you're kicking back at ho me, sipping some chai, and your phone buzzes with a video call. It's your boss, looking stressed out, be gging you to wire some cash for an "emergency." You 're a good employee, so you send it quickly—only to find out later it wasn't your boss at all. It was a deepfake. Freaky, right? That's not just a random what-if; it's happening all over India, and it's messing with people's lives.

Deepfakes are AI's sneaky little trick—videos, audios, or pictures that look so real you'd bet your last rupee on them. But here's the twist: they're fake, and cybercriminals are having a field day with them. They're scamming people, stealing identities, and even stirring up chaos online. And it's not just us regular people. Politicians, journalists and celebrities are getting hit too. Remember that Rashmika Mandanna deepfake that blew up? People couldn't tell if it was her or not, and that's the problem.

So, let's break it down. We're talking deepfakes in cybercrime, how ethical hackers are fighting back with Capture The Flag (CTF) competitions, the tools they use, and what's next for India. It's a rollercoaster, so hang on tight.

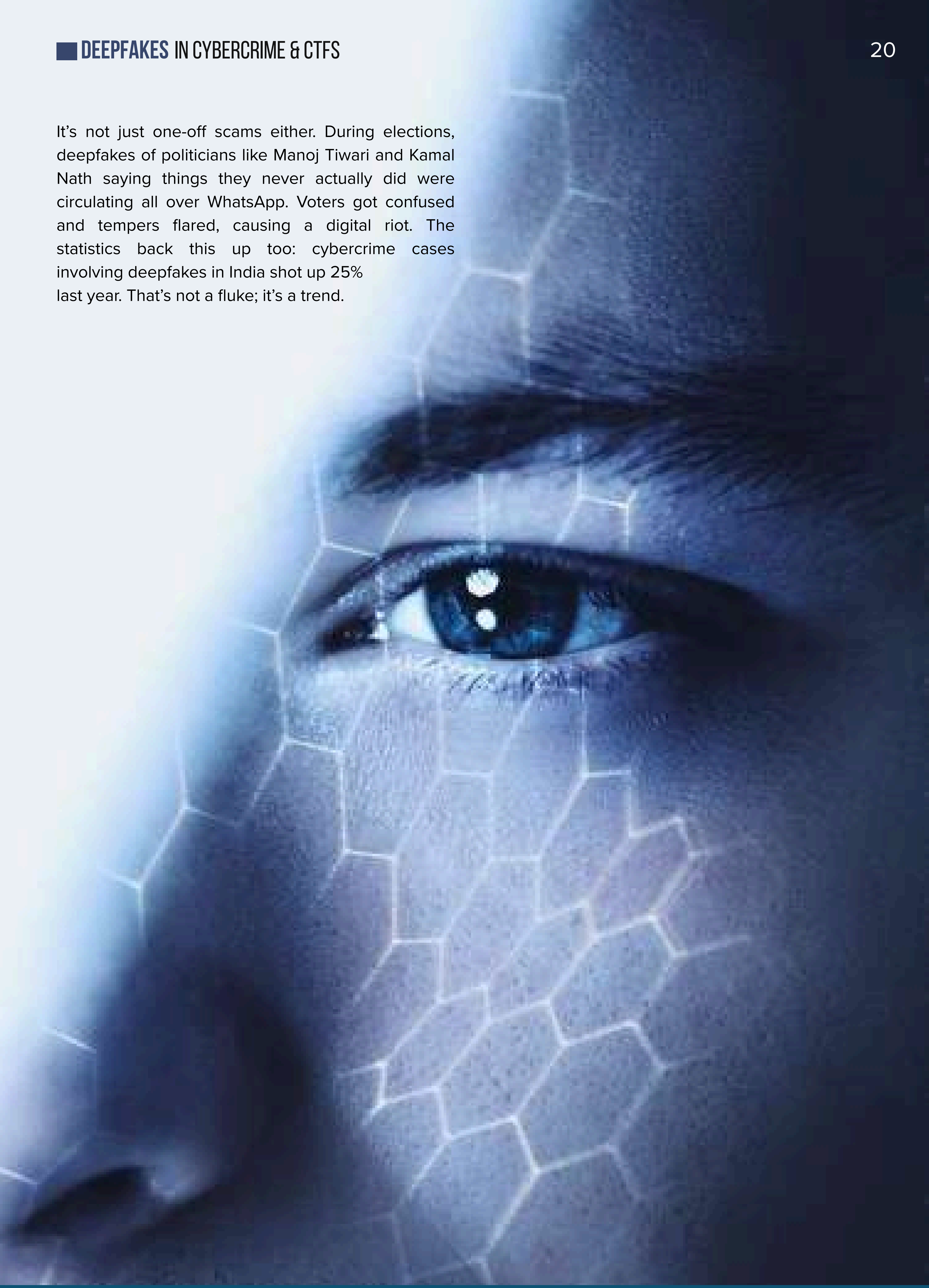
Deepfakes in Cybercrime

Cybercrime's already a headache in India, but deepfakes? They're like throwing a Molotov cocktail into the mix. These fakes are popping up everywhere, hitting people where it hurts—wallets, trust, you name it.

Voice and Video Scams

Let's start with this guy in Ghaziabad. He's 76, probably spending time with his grandchildren, when he receives a deepfake video. Someone pretending to be a retired cop calls, saying he has a clip of the old man doing something sketchy, and asks to pay Rs 74,000, or it's going viral. The old man is terrified and forks over the cash without a second thought. Later, he figures out it's a scam and calls the cops. They've filed an FIR, but that money's probably long gone. In Kerala, a 73-year-old gets a video call from someone who looks and sounds exactly like his old coworker. The "friend" says he's in a pinch—needs Rs 40,000 for emergency surgery. The man sends it, no hesitation. But then the scammer gets greedy, asks for more, and that's when the victim smells a rat. Turns out, it was a deepfake. His real friend had no clue. Cops traced the cash to Maharashtra, froze the account, but had no luck catching the perpetrator.

It’s not just one-off scams either. During elections, deepfakes of politicians like Manoj Tiwari and Kamal Nath saying things they never actually did were circulating all over WhatsApp. Voters got confused and tempers flared, causing a digital riot. The statistics back this up too: cybercrime cases involving deepfakes in India shot up 25%
last year. That’s not a fluke; it’s a trend.



Case Study	Location	Victim	Amount Lost	Method	Outcome
Extortion Scam	Ghaziabad	Arvind Sharma, 76	Rs 74,000	Deepfake video of retired IPS officer	FIR filed, investigation ongoing
Video Call Scam	Kerala	PS Radhakrishnan, 73	Rs 40,000	Deepfake video call impersonating colleague	Money traced, account Frozen

Corporate Espionage and Fraud

Now picture this: you're at work, and your CEO hops on a video call, looking serious, asking you to transfer cash or asking for confidential company details. You'd do it, ofcourse. But what if it's not your CEO and just a deepfake? That's social engineering at play. With everyone in India glued to video calls these days, businesses are sitting ducks. One slip, and they can make millions in losses. Then there are security systems. Banks and offices use face recognition, but deepfakes can trick that too. It's like giving crooks a master key to your life. McAfee says 38% of Indians have already run into deepfake scams. Think about that next time you're on a call.

Dark Web and Underground Markets

If you think that's bad, check the dark web. It's like a deepfake Walmart. "Deepfake-as-a-service" is available to anyone with a few bucks. No coding skills? No problem. Pay up, and you've got tools to make fake videos or audios for blackmail, revenge, or worse. Fake adult content's a big one too, and in India, where cyber laws are still playing catch-up, it's a disaster waiting to happen.

Deepfakes in Capture The Flag (CTF) Competitions & Ethical Hacking

Okay, enough doom and gloom. Let's talk about the good guys. Ethical hackers are stepping up, and they're training hard in CTF competitions. Think of CTFs as cyber boot camps – India's got events like InCTF and HackathonX where tech geeks battle it out, cracking codes and spotting fakes.

Why CTFs Embrace Deepfakes

CTFs are all about real-world prep. With deepfakes on the rise, organizers are tossing in challenges to spot them. It's like a game of "spot the imposter"— weird shadows, glitchy audio, anything that screams fake. These skills are a clutch for fighting cybercrime. Deepfake challenges aren't huge in Indian CTFs yet, but the basics, like video analysis or data forensics, are perfect for it. I did a CTF once where we had to pick apart a video for hidden messages. It took hours, but I learned to notice things like off lip-sync or dubious lighting. That's the kind of sharpness you need to build against deepfakes.

Challenge Examples

Here’s what you might see at a CTF:

- Steganography + Deepfake: Hunt for secret codes buried in a fake video. Look for pixel quirks or frame skips.
- Forensic Analysis: Spot tampering - like if someone’s mouth moves weird or the background’s off. Deepfakes slip up on the details.
- Audio Analysis: Check for robotic voices or funky sound waves. Tools like Audacity can help.

These tricks don’t just win you points. They’re real-world weapons against deepfakes.

Tools & Techniques

Hackers in CTFs use some cutting-edge tools, and they’re getting better at catching fakes:

- Deepware Scanner: Scans videos for AI trickery think frame glitches or unnatural patterns.
- Forensically: Digs into images and videos for edits. Spots noise or cloning fast.
- AI Frame Detection: New stuff that uses AI to flag fake frames. It’s like AI vs. AI—pretty badass.

The catch? Crooks keep upgrading their game, so these tools need constant tweaks.

The Road Ahead: Ethics & Prevention

Deepfakes are evolving so fast it’s like chasing a speeding train. India has to keep up.

Challenges- Detection is tough. AI can make fakes that fool even pros. McAfee says 75% of Indians have seen deepfakes online and didn’t clock them as fake. That’s how slick they are. Plus, there’s the ethical mess—bullying, fake porn, ruined lives. Over half of Indians are freaked out about this stuff, and they should be.

Policy & Governance

As per a government mandate, social media platforms have 24 hours to take down deepfake posts after a complaint. There’s also a Deepfakes Analysis Unit (DAU) that can be contacted on Whatsapp to verify content. But these measures may not be enough. We need to revaluate of our laws are capable of addressing emerging technological threats. Tech companies, too, need to take more repsonsibility. Right now, it’s too easy for scammers to slide by.

Conclusion

Deepfakes are the internet’s new monster, and India is feeling the heat. From Ghaziabad scams to election chaos, they’re everywhere. But ethical hackers are fighting back with CTFs, sharpening their skills to spot fakes. But we need better tools, laws, and a wake-up call for everyone. Your role as a reader is to stay vigilant. Double-check your sources and support the cyber warriors to make the internet less of a minefield. Deepfakes erode trust, but there is hope yet.

Prevention Strategy	Description	Current Status in India
Detection Tools	Software to spot fakes	Decent, but needs to keep evolving
Ethical Hacking	CTF training	Growing, needs more focus
Public Awareness	Educating the public	Barely there. Needs a push
Legislation	Laws to control deepfakes	Needs attention

Social Fractures: The Ripple Effects of Election Misinformation

Nupur Singh

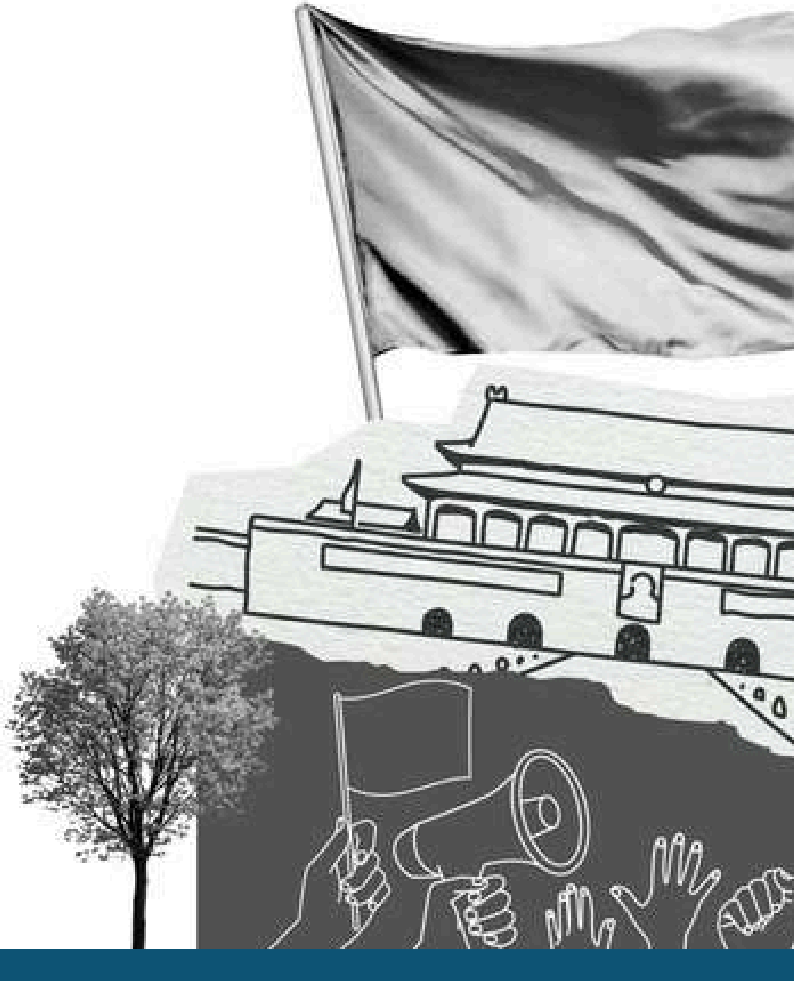
Introduction

The rise of social media has benefited humanity considerably by providing an opportunity to dissecminate and access a wide array of information. However, it also has a downside. The same technology that offers greater access to information can also be leveraged to spread unverified, false information and propaganda campaigns to influence public opinion. Researchers argue that while social media is an effective tool to inform and create awareness, it can also be misused for manipulation through misinformation and disinformation, posing severe challenges to democratic processes.

Modern democracies, characterized by the freedom of expression with active and vibrant social media communities, have created ample opportunities and platforms to express opinions and share informa tion freely. While information flows freely, so does misinformation and disinformation. According to the Global Risks report 2024, the most severe short-term risk that we face today is that of ‘manipulated and falsified information’. It further enunciated that misinformation and disinformation can disrupt electoral processes in several nations, thereby triggering civil unrest and confrontation. It can further lead to distrust in the media and government, polarizing views of the public. Riding the digital tide, generative AI, especially deepfake videos, increases the risk manifold.

The Psychology of Misinformation: Altering Thought and Behavior

Misinformation is a scenario where misleading information spreads without any intent to. However, a rather more concerning phenomenon is ‘disinformation’, when the intent is to spread false information to confuse or cause harm. The US election results in 2016 unraveled the extent of its impact. Several studies



conducted in the US found that misinformation was responsible for a considerable number of deaths and hospitalizations in the country in 2020. In India, in the context of the COVID-19 pandemic, the promulgation of false and misleading information exposed the serious implications and the crucial need to manage and mitigate misinformation. As per the Press Information Bureau (PIB) in India, the number of fake news items and false claims about different types of cures, negative impact of the vaccines etc., increased on social media platforms, which led to panic and unrest. This created an environment of distrust in the public health system and fear among the masses.

The spread of misinformation depends on several factors. A prominent research study on psychological factors contributing to the creation and dissemination of fake news on social media concluded that the spread of misinformation is propelled by psychological factors like emotions and cognitive bias, along with factors like social network structure and algorithm. Cognitive biases like confirmation bias lead people to accept and share information that aligns with their pre-existing beliefs.

Moreover, emotions such as fear, anger, or excitement can drive impulsive reactions, further amplifying the spread of misinformation. The impact goes much beyond misinterpretation of facts and data; it has a deep social and psychological impression. As per an MIT Sloan research, misleading content on Facebook, although less persuasive, generated more skepticism about the COVID-19 vaccine due to its wide reach. Notably, the unflagged misinformation had more impact than the flagged. To address this, experts believe that pre-emptive pre-bunking and reactive debunking interventions can reduce the effect of misinformat ion.

Use of Engineered Narratives to Misdirect Public Opinion

Public opinion is the basis and a vital force in a liberal democracy. It plays a crucial role during elections, deciding the fate of a country by choosing the leaders. The manipulation of public opinion through disinformation has become rampant. Democracies like the US and India are more susceptible to the risks posed by misinformation, which can have a disruptive influence on the opinion of the masses about their political representatives, eroding trust, undermining social cohesion, and triggering social fragmentation. Fabricating claims in election manifestos and rules of voting, etc., to confuse the citizens at large, biased media reportage, and circulating false election outcomes, deepfake videos, etc., are all tactics to spread misinformation or disinformation during an election.

Recognising its ill-effects, the Election Commission of India (ECI) took several measures to combat misinformation during the 2024 Lok Sabha election, including but not limited to, collaboration with tech companies and the Myth Vs Reality Campaign. To curb misinformation, it is imperative to verify information before posting it on social media platforms. ECI collaborated with Google to provide verified information regarding the general election to the voters, and the company connected voters to the authorized organisations and their websites.

Similarly, YouTube provided high-quality video content for election-related information. Additionally, Google enforced strict policies and restrictions for the election-related advertising campaigns run on its platforms. For instance, mandatory identity verification and pre-certification of all political advertisements across electronic media from the Media Certification and Monitoring Committees at both district and state levels.

Further, ECI launched the “Myth vs Reality” project, in which election-related information was disseminated via its Twitter, Facebook, Instagram handle, and website. The purpose was to promote informed participation by citizens and prevent confusion or manipulation caused by false information. To maintain integrity of the election process and ensure free, fair and ethical use of social, media platforms, a Voluntary Code of Ethics, was agreed upon by Facebook, WhatsApp, Twitter, Google, ShareChat and TikTok which aims to ensure free, fair & ethical usage of Social Media Platforms to maintain the integrity of the elect oral process. Additionally, the ECI also urged voters to rely on official communication by the political parties, the Election Commission, official parliamentary portals, genuine and verified media publications and news portals; and to cross-check facts through reliable fact-checking tools to identify deepfakes.

The effect of misinformation is not limited to a region or nation but extends to the realm of international relations with far-reaching implications affecting the political landscape and diplomatic ties. For instance, recently, a coordinated disinformation campaign involving a video of the French President Emmanuel Macron alongside German Chancellor Friedrich Merz was circulated, claiming the former was concealing a bag of white powder, allegedly cocaine. However, the authorities clarified that it was one of the foreign-led disinformation campaigns, an attempt to sow mistrust amongst the public in Europe.

While the government has recognized the need to create awareness and regulations to check misinform ation, there is a long way to go. Driven by the AI revolution, deepfake technologies are one step ahead of the measures taken to combat misinformation.

Therefore, the authorities need to work on several levels consistently. Primarily, social media platforms must be held accountable for content amplification driven by opaque algorithms. The rise of social media has empowered users to share and consume unlimited information. However, power can’t be devoid of responsibilities. So, there is a need to inculcate responsible behavior among them in taking the call for what is to be shared and whether the source is verified.

Tools to Check Misinformation

Currently, a variety of tools and technologies are available to verify facts, track sources, and identify manipulations. For instance, fact-checking platform -s like CaptainFact offer a web-based collection of tools that assist in the collaborative verification of content, alongside others like Snopes and Politi Fact. In addition to these platforms, there are AI-powered detection tools that can check facts, monitor their dissemination across various media sources, and evaluate the credibility of those sources. Examples include ClaimBuster, the Full Fact Toolkit, TruthNest, Hoaxy, and Media Cloud. Reverse image and video search tools, such as Google Reverse Image, TinEye, and InVID-WeVer ify, are also beneficial for checking the origins, context, and possible manipulations in videos.

With the prevalence of synthetic media, such as deepfakes, it is crucial to employ deepfake detec tion tools so that synthetic media can be identified and checked for manipulations, that too real-time. Furthermore, there are several blockchain-based verification methods, such as Truepic, that can be employed to verify image authenticity. This process utilizes secure metadata and timestamps to ensure accuracy.

Digital Forensics for Misinformation and the Challenges

Conducting scientific investigations of digital evidence, digital forensics has become a crucial aspect in the search for truth in this digital era, especially in the context of rising cybercrimes and misinformation. This poses a serious threat to public discourse and democratic processes. While the digital tools to thwart the threats have evolved, several challenges continue to plague the effectiveness of digital forensics. Lack of universal standardization in terms of qualifications, education, and skills is the most critical challenge of all. There is a need to bridge the gap between the rate of increase in cyber threats and skilled people to conduct the investigations. The pace of technological advance ments also poses a challenge for digital forensics, requiring constant learn ing and training to keep up with new technologies. Additionally, an increase in high-quality, AI-generated content also makes it difficult to distinguish between real and fake images, audio, or video. Use of evasive techniques like cloaking or ephemeral posts also makes it challenging to track the posts and their flow. Legal complications resulting in debate as to how data should be handled according to the law. More often, the efforts to monitor, trace, or remove content converge with issues of privacy and free speech. So, the task to bal ance the need for accuracy and freedom of speech constitutes the ethical dilemma in digital forensics. There is no doubt that digital forensics is a powerful tool for detecting misinformation, but the complex challenges it faces make it difficult to handle the growing cyberthreats. The answer lies in a multilayered strategy combi ning innovation cross-platform collaboration and a robust legal framework with clear regulations ensuring both security and privacy.

The Future Roadmap

To Combat misinformation globally, we need a unified and multi-pronged strategy. To achieve this, the governments must establish a set of transparent and fair regulations to hold the propagators of fake news accountable on one hand, while also safeguarding the freedom of expression on the other. Social Media platforms should spruce up their algorithms and adopt effective policies prioritizing verified content and quick intervention, combining human oversight with AI technologies. Furthermore, Media organizations play a pivotal role in the fight against misinformation by disseminating pre-verified pieces of information reflecting the highest levels of objectivity. Further, awareness generation across varied strata and age groups is important. The role of NGOs, civil society, and the community at large cannot be overstated. To complement the efforts, digital literacy programmes should be embedded in the school curriculum so that the future generation is aware and well-equipped to differentiate between ‘authentic’ and ‘misleading’ information. At the global level, international organizations should encourage collaboration and partnership among states and organizations to join hands in mitigating misinformation. There is a need for constant action, as inaction can threaten individual as well as national security and impact world peace. Lastly, governments, technology companies, media, educational institutions, and civil society need to work together towards creating a more resilient, safe, and sustainable information environment of the future.

References:

https://www.weforum.org/stories/2024/01/ai-disinformation-global-risks/

https://mitsloan.mit.edu/press/misleading-covid-19-headlines-mainstream-sources-did-more-harm-facebook-fake-news

https://www.hindustantimes.com/india-news/myth-vs-reality-ec-lays-down-steps-to-curb-fake-news-101710615785656.html

https://trt.global/world/article/72ff6e161907

https://bmcpsychology.biomedcentral.com/articles/10.1186/s40359-024-02129-2

https://www.rand.org/research/projects/truth-decay/fighting-disinformation/search.html

https://www.researchgate.net/publication/337057852_Digital_Forensics_Challenges_and_Opportunities_for_Future_Studies

▶ **Nupr Singh**

journalist and Sr. Policy Researcher at CyberPeace

Forensic Detection of Deepfakes: AI-Driven Forensic Techniques, Authenticity Verification, and Digital Media Analysis

Smita Rukhande, Dr. Nilakshi Jain, Maj. Vineet Kumar

AI-Based Approaches to Detect Deepfakes

- **Convolutional Neural Networks (CNNs)**

CNNs are most famous in identifying tampered media. Author Rossler et al., used XceptionNet model to analyse digital artifacts such as facial texture inconsis tencies and illumination errors, differentiating between real and manipulated content.

- **Vision Transformers (ViTs)**

Vision Transformers have improved detection capabili ties. These AI tools can scan images in detail and collect small changes or errors in how the fake content is put together. This makes them good at catching tric ks used in high-quality deepfakes .

- **Temporal Dynamics with Recurrent Networks**

AI models like LSTMs and 3D CNNs are used to exam -ine time-based inconsistencies in video sequences, su ch as unnatural eye movements or poor synchronizatio n between audio and lip motion.

- **Spectral Analysis Methods**

Even if a video looks real, the sound can give it away. Fake audio might have odd frequency patterns. Tools based on Discrete Fourier Transform (DFT) and Wavelet Transform techniques help to detect such irre -gularities. Studies by author Durall et al. mentioned that training models on spectral data can outperform traditional image based methods in specific cases.

- **Identifying GAN Signatures**

Just like each camera leaves a small digital trace every AI model leaves a unique signature in its fake content. Each GAN model leaves a unique fingerprint. Authors L. Zhang et al. proposed a method to classify these fingerprints or synthetic content using a forensic neural network trained to distinguish content created by different generators.

Authenticity Verification Techniques

- **Examining Metadata**

Digital files often carry metadata, such as creation timestamps and device information. Using forensic tool s, this data can be analysed for irregularities. If this dat a does not match what is in the video it might be fake. Although skilled attackers can alter metadata, inconsist encies can still be a valuable clue.

- **Blockchain and Content Verification**

Blockchain technology can be used to track the origin of content. When a video is first made, hashed versions of the original file are stored on decentralized ledgers. If someone edits it later, the new version would not match the original blockchain record.

- **Biometric Cross-Validation**

Advanced systems use biometric features like facial geometry or vocal characteristics to compare synthetic media against known references. If the face structure or voice pattern does not match, the system can flag the content as suspicious. Such methods are effective for detecting voice clones or facial changes.

- **Reverse Content Search**

AI supported reverse search tools assist investigators in locating original versions of manipulated media. The se searches can reveal the real content or detect synthetic content passed off as authentic.

Tools for Analysing Digital Media

- **FaceForensics++**

This dataset provides manipulated media generated using popular tools like FaceSwap and DeepFakes, providing researchers a controlled environment to develop and evaluate detection models.

- **DeepFake Detection Challenge**

Big technical companies released this dataset to help researchers to build better detection tools. It includes lots of different fake videos to train AI models.

- **Microsoft's Detection Tool**

Microsoft's Video Authenticator uses AI to calculate the score of how likely a video has been altered. It looks at small clues in each video frame and across time to spot fakes quickly.

Challenges in Detection

- **Different Fake Styles:** AI models trained on one kind of deepfake might not check other kind of deepfake.
- **Smart Attackers:** People who make deepfakes are learning how to beat detection tools.
- **Biased Training Data:** If AI only learns from limited examples, it cannot do well on new examples.
- **Real-Time Processing:** Checking videos in real-time like during live news or online calls, remains difficult.

Directions for Future Research

- **Multimodal Approaches:** Combining analysis of vis ual, audio, and text data can improve detection accuracy.
- **Explainable Models:** Tools that explain how a dete ction decision was made are becoming increasin gly important.
- **Privacy-Conscious Training:** Federated learning all ows the development of detection models without exposing sensitive data.
- **Legal Frameworks:** Governments and companies need to set clear guidelines for creating and shari ng AI-generated content.

Conclusion

The increasing realism of deepfakes presents serious challenges for verifying digital content. However, AI-driven forensic technologies provide promising tools for detecting and mitigating these threats. With continuous progress in detection models, auth enticity validation systems, and cross-disciplinary collaboration, we can better secure our digital cont ent. By combining AI tools and ethical consider ations and policy development, we can detect fake content and help to keep online information honest and trustworthy.

References

- S. M. Qureshi, A. Saeed, S. H. Almotiri, F. Ahmad, and M. A. Al Ghamdi, "Deepfake forensics: A survey of digital forensic methods for multimodal deepfake identification on social media," Front. Public Health, 2024. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11157519/
- A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "FaceForensics++: Learning to Detect Manipulated Facial Images," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), 2019, pp. 1–11.
- A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," arXiv preprint arXiv:2010.11929, 2020.
- D. Guera and E. J. Delp, "Deepfake Video Detection Using Recurrent Neural Networks," in Proc. IEEE Int. Conf. Adv. Video and Signal-Based Surveillance (AVSS), 2018.
- R. Durall, M. Keuper, and J. Keuper, "Watch your Up-Convolution: CNN Based Generative Deep Neural Networks are Failing to Reproduce Spectral Distributions," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR) Workshops, 2020.
- L. Zhang et al., "GAN Fingerprints: A New Feature for Fake Image Detection," IEEE Trans. Inf. Forensics Secur., vol. 16, pp. 4753–4768, 2021.
- S. M. West, M. Whittaker, and K. Crawford, "Discriminating Systems: Gender, Race and Power in AI," AI Now Institute, 2021. [Online]. Available: https://ainowinstitute.org/discriminatingsystems.html
- T. Kinnunen, M. Sahidullah, H. Delgado, N. Evans, J. Yamagishi, and K. A. Lee, "Vulnerabilities in Speaker Verification," IEEE Signal Process. Mag., vol. 37, no. 6, pp. 87–100, Nov. 2020.
- B. Dolhansky, J. Bitton, B. Pflaum, J. Lu, R. Howes, M. Wang, and C. C. Ferrer, "The Deepfake Detection Challenge (DFDC) Dataset," arXiv preprint arXiv:2006.07397, 2020.

Ms. Smita Rukhande

About the Author:

Ms. Smita Rukhande is a research scholar in the Department of Information Technology at Shah and Anchor Kutchhi Engineering College. She also serves as an Assistant Professor in the Department of Computer Engineering at Fr. C. Rodrigues Institute of Technology, Vashi, Navi Mumbai, India. Her academic and research pursuits lie at the intersection of emerging technologies and computer science education.

Dr. Nilakshi Jain

About the Author:

Dr. Nilakshi Jain is a faculty member in the Department of Cyber Security at Shah and Anchor Kutchhi Engineering College, Mumbai, India. She is dedicated to advancing knowledge in the field of cybersecurity and contributing to academic and professional growth in the domain.

Maj. Vineet Kumar

About the Author:

Maj. Vineet Kumar is the Founder and President of the CyberPeace Foundation, based in Delhi, India. A pioneering advocate for cybersecurity and cyber peace, he leads initiatives focused on creating a secure and resilient digital ecosystem. Through his work, he continues to drive policy, education, and innovation in the field of cyber safety and digital rights.

The Algorithmic Hydra: Countering Deepfakes and Misinformation in India's Digital Ecosystem Through AI, Blockchain, and Advanced Forensics- An Evidence-Based Analysis

Darshil Raval

Abstract

This research analyzes the increasing danger of deepfake technology combined with digital misinformation spread in India's special digital environment. India deals with specific obstacles, from its multilingual natu.re combined with divergent digital competence ratings, and its complex regulatory framework. The country has 5.56 billion internet users and 5.24 billion social media users according to industry estimation. Data shows that deepfake content will significantly increase, and political deepfakes will surge when elections occur. This paper integrates recent research about detection technology by exploring AI tools alongside blockchain authentication systems and specialized Indian media forensics methods. It examines India-specific issues that stem from the widespread circulation of false information through WhatsApp platforms among 535.8 million users as well as inadequate computer processing capabilities in rural areas and insufficient legal regulatory frameworks. Our proposed implementation strategy has three levels. First,it demands grassroots approaches for regional language detection. Second, it seeks policy changes to the Information Technology Act and thirdly, it suggests the use of federated AI systems that provide privacy protection and large-scale detection capabilities. Thus, India needs specific answers to enhance its particular social-technical conditions while utilizing advanced global deepfake detection techniques.

Introduction

1.1 The Growing Threat of Deepfakes in India

The digital choices facing India have reached an essential point of decision. The digital population of India makes it one of the world's largest due to its 806 million internet users who are part of the estimated 491 million social media users. This massive digital footprint, along with rapid AI advancements, creates optimal conditions for the increased production of deepfakes and synthetic media. Research shows deepfake occurrence reports continued to rise steadily from 2022 to 2024 per various recording agencies. The number of political deepfakes experienced significant growth during election times when compared to times when no elections were happening which led to serious concerns about maintaining electoral integrity.
A first research report suggests: Since late 2023, deepfakes have expanded at a rapid pace because advanced generation programs are moving beyond the abilities of detection systems. Confirmed deep fake incidents involving political subjects make up the most prevalent category in deep fake data, while money scams and personal offenses along with celebrity imitation incidents also appear frequently.

1.2 India's Unique Digital Landscape

The digital environment in India creates distinctive challenges that make its deepfake problems different from those of other nations. Some of its features are:

- Content moderation systems need to manage 22 official languages and multiple dialects.

- The smartphone functions as the main method of internet use for almost all people accessing online services.

- WhatsApp dominates the messaging landscape. It has closed information networks that guard against fact-checkers from reaching over 400 million Indians operating in sealed information environments.

- Research findings reveal digital reading abilities differ between Indian rural and urban communities because of places' sparse access to detection platforms.

Digital technology adoption within different socioeconomic groups creates a vulnerable environment that makes sophisticated misleading content easier to spread. Digital transformation throughout India makes it difficult for researchers to develop reliable protection against emerging threats like deepfakes according to recent scholarly investigations.

1.3 Research Objectives

This paper aims to:

- Analyse the state of deepfake technology and its Indian manifestations
- Assess new countermeasures for their practical suitability to Indian conditions.
- Recommend executable answers at three stages starting from local communities up to national policy.

Literature Review

2.1 Evolution of Deepfake Technology

Deepfake technology experienced rapid growth after its initial development in 2017. The latest development of diffusion models enables fraudsters to produce realistic deepfakes which recently became evident through financial institution voice-cloning scams. Modern deepfake models use either 3-5 images or just 5 seconds of voice recording to produce realistic video fake content which raises significant concerns for India because its public figures have numerous available images.

The 갃i model enabled criminal elements, through Pindrop (leading voice security provider), to develop hyper-realistic deepfakes which resulted in audio cloning scams targeting financial organizations. The progress in deepfake technology creation tools has significant effects on India's media environment because it makes it simpler to produce convincing deepfakes that target the Indian demographic.

2.2 Indian Research Context

Research scientists from India have published important work related to deepfake detection. Through its *Center for Digital Trust, the Indian Institute of Technology Bombay* works on creating datasets with manipulated videos of various Indian communities. The dataset fills a fundamental gap because previous databases showed limited representation of Indian communities.

The recent developments in Indian deepfake detection methods rely on frequency domain analysis that specializes in recognizing Indian speech patterns. The detection techniques hold great potential for identifying synthetic speech throughout various Indian languages through a reduced need for comp

S.NO.	Generation Method	Input Requirements	Key Vulnerability
1.	GAN-based	300+ images	Facial inconsistencies
2.	Encoder-decoder	50-100 images	Audio-visual synchronization
3.	Hybrid CNN/Transformer	20-30 images	Temporal coherence
4.	Diffusion models	3-5 images	All modalities coherent
5.	One-shot diffusion	Single image	Near-perfect synthesis

Source: Industry reports and research literature

2.3 Regulatory and Policy Approaches

Deepfakes have triggered different policy reactions among the international community. The European *Digital Services Act* and the American *DEEPFAKES Accountability Act* can show India how to approach deepfakes legislation. The legal framework that India currently uses mainly consists of the *Information Technology Act 2000 (amended in 2008)* while the deepfake technology already exists. The Indian government intends through its proposed IT Act amendments to enforce AI-content labeling standards yet lacks a specific delineation of enforcement procedures. The *Ministry of Electronics and Information Technology (MeitY)* established digital platform regulations through guidelines in November 2023 yet these guidelines remain unenforceable.

The regulatory sector in India is undergoing continuous transformation. Experts advise that India should adopt a complete legal strategy that mixes platform responsibility provisions with technical standard requirements to better combat deep fake threats. Such frameworks ought to include stages that progress according to technical limitations and set specific benchmarks for advancing.

3. Challenges in the Indian Context

3.1 WhatsApp and Closed-Network Misinformation

Whatsapp operates as a distinctive communication system throughout India. It maintains 535.8 million users who make it their predominant digital messaging service throughout India. Deepfake videos can spread across WhatsApp networks to millions of users, especially during sensitive time such as state elections, without any fact-checking procedures being available in time.

The end-to-end encryption used by WhatsApp makes it among the most challenging platforms for detection systems to work effectively. Open platforms including Facebook and YouTube lack access to message content that runs through WhatsApp, so traditional moderation techniques cannot inspect such data. According to research results, deepfake content distribution is notably faster on exclusively private messaging systems compared to public social media platforms within the Indian use area. Research into WhatsApp messaging networks during Indian elections shows deepfake videos spread quickly because tens of thousands of users receive these videos within 24 hours of a single share but public platforms generate fewer shares. The isolated networks offer conditions where content stays beyond external fact-based verification methods.

3.2 Linguistic and Cultural Diversity

The broad language diversity within India ,with its 22 official tongues and numerous dialects, creates severe barriers for automated detection systems to operate. Deepfake detection models show poor performance when examining Indian languages because their training occurs mainly with English content. The testing reveals that cutting-edge detection models demonstrate reduced performance levels during Indian regional language content evaluation compared to their English content analysis. The built-in language preference of AI systems introduces vulnerabilities mainly in settings where regional languages prevail. Cultural nuances further complicate detection. The detection of local traditions and regional circumstances demands cultural information which current models lack even though they operate at a global level. Studies reveal that detections of Indian political deepfakes would attain substantial accuracy improvement by incorporating culturally contextual information.

3.3 Digital Literacy and Awareness

An insufficient digital literacy distribution across India creates many hindrances for deepfake management. Studies demonstrate the existence of a digital literacy gap between rural and urban areas because rural areas have limited access to detection tools. Evaluation of content for sharing stands lower in older audiences as compared to younger audiences who verify more frequently. The growth in new Indian internet users composed of older adults requires special attention due to their increasing numbers.

Digital literacy initiatives should target specific groups with separate strategies because verification practices differ between various segments of India's population. Environmental analysts have detected that digital literacy inequalities produce variations between metropolitan knowledge consumers with their advanced information access capabilities and vulnerable backcountry and poorly educated groups who struggle to avoid elaborate false campaigns.

3.4 Computational Resource Constraints

Local deployment of detection systems remains difficult within India because of its limited resource availability. State-of-the-art deepfake detection models need computing power surpassing what is possible with many currently used smartphones in India. Network constraints further complicate detection capabilities. The current mobile internet speed levels in rural areas of India fall below the standards needed to support real-time detection services operating from the cloud and requiring preeminent bandwidth. Researchers who studied deepfake detection implementation in India reported that high computational requirements of advanced syst-ems prevented their widespread deployment in India's digital environment. Indian smartphone users who do not possess enough processing power in their devices and experience limited bandwidth, particularly in small-er cities encounter difficulties when trying to utilize the-se detection models. Researchers confirm that India requires detection systems designed to operate within its technological limits.

4. Emerging Countermeasures

4.1 AI-Powered Detection Systems

Recent advances in AI-powered detection show promi-se for the Indian context. Transfer learning methods make it possible for global models to adapt to both Indi-a visual elements together with their linguistic characteristics. Research teams have dedicated efforts to enhance pre-trained models for precision improvement as well as system processing efficiency. The detection of advanced deepfakes becomes more effective through the analysis of visual and audio data in combined multimodal detection approaches. The voice-pattern analysis offered by Pindrop startup detects synthet ic audio with a 96% success rate. The first deploye -d solutions indicate they can successfully confront deepfake detection challenges. The Indian user base has adopted fact-checking browser extension -s developed by fact-checking organizations that operate at a lightweight scale. Preliminary information indicates these tools successfully lower the rate of deepfake distribution among their user base.

Recent scientific approaches target Indians' limited resources through small and efficient models that obtain adequate detection performance although they use minimal power and operational time on standard smartphones. These architecture models design their systems with regional characteristics that potentially produce fewer incorrect results rela -tive to generic models.

4.2 Blockchain-Based Provenance Solutions

The technological blockchain framework presents effective solutions to authenticate media content in the Indian context. Applications built on blockchain technology through DApps conduct tests of media provenance as evidence for legal authority assessment. The distribution of verification records for authenticated media content takes place throu gh Distributed Ledger Technology systems through their implementation of "content credentials."

The initial installations in controlled settings display positive results. State polls developed blockchain-based proof systems to check official political materials through pilots which introduced progress in eliminating erratic content distributions over earlier voting periods. Blockchain solutions encounter substantial challenges when it comes to adoption. The high cost of putting blockchain into practice remains a problem and major integration obstacles exist between blockchain networks and current content sharing platforms. Academic teams have developed federated blockchain architecture solutions that demonstrate both cost-effectiveness compared to regular blockchain systems and crypt -ographic protection of data integrity. These approaches allow for lightweight client validation on mobile devices using minimal storage and processing power, potentially making them more viable for India's diverse device ecosystem.

4.3 Media Forensics and Metadata Standards

Various media verification methods have proven their effectiveness specifically for Indian forensic investigations. Scientists recently achieved pr -omising results regarding biological signal analysis through studies of pulse and blink patterns and micro-expressions in identify manipulated videos regardless of compression techniques. *The Coalition for Content Provenance and Authenticity (C2PA)* standards have successfully expanded their use around the world. Important news organizations now embed C2PA metadata into their digital content while providing cryptographic verification for all items. New-generation digital watermarking technologies have been developed to work effectively in limited bandwidth conditions. The systems incorporate verification data that functions to persist across compression steps used by messaging programs thus allowing authentication to take place through multiple forward transfers. Compression-resistant watermarking technologies achieve precise detection through multiple compressions made to messaging platforms. This development stands as a key progress for Indian conditions because information typically requires multiple compressions before forwarding. The implementation process needs very little new data while remaining invisible to the human eye. The method suits India's messaging-based information environment well by addressing fundamental authentication problems found in content authentic ation.

5. India-Specific Solutions

5.1 Grassroots Initiatives

Platform accountability functions as a significant policy tool. The requirement of transparency reports from major India-based platforms presents significant potential for reducing deepfake problems. The introduction of a "notice-and-action" system modeled after the EU Digital Services Act would promote decreased deepfake dissemination through significant platforms. Policy experts indicate that three-level regulatory models built from self-regulation and co-regulation together with statutory oversight function best. The regulatory approach divides responsibilities between platforms that would detect and label deepfakes while industry bodies establish standards and government authorities serve as the enforcement and regulation authority. The framework optimizes deepfake regulation through a system of stakeholders who understand this complicated matter and define their duties within the system.

5.3 Technological Infrastructure

India can benefit from federated learning techniques that address its

▶ Darshil Raval

About the Author:

Darshil Raval is a cybersecurity enthusiast with a B.Tech in Computer Science and Engineering (Cybersecurity) from Parul University, graduating in 2024. He currently works as a freelancer, specializing in Vulnerability Assessment and Penetration Testing (VAPT) for web applications and network infrastructure. Darshil has also interned as a Digital Forensics and Incident Response (DFIR) analyst at the Centre for Cybersecurity and Strategic Studies Research (CFCSSR). Alongside his technical work, he actively shares insights and experiences through his personal blog, where he writes on cybersecurity trends, tools, and case studies. *HE can be found on LinkedIn at:* *https://www.linkedin.com/in/darshilraval* .
Blog:
https://ciphersec.hashnode.dev/

privacy needs as well as its resource limitations. The combination of smartphone collaboration for model detection training without raw data exc -hange would resolve privacy and bandwidth issues thus producing more accurate results and minimizing each device's processing load. The field of edge AI speeds up through various programs that fine-tune detection algorithms for basic hardware. Low-power smartphones show significant promise for widespread use because they run neural networks that operate well using reasonable accuracy levels.

Digital signatures demonstrate potential integration within India's digital identity system. A framework for media content creators' digital signatures would establish trust in authentic materials, especially news organizations and official sources. Professional teams study hybrid cloud-edge systems that shift processing tasks based on device performance skills and network framework characteristics. New detection methods should be able to adapt to the different technological conditions throughout India's regions so they can provide operational detection systems in resource-limited areas.

6. Future Trends and Research Directions

The deepfake technology in India demonstrates an increasing speed of development. Several emerging trends warrant attention:

- Research shows that deepfakes consisting of multiple audio-visual-text components are more difficult to detect because they offer better illusions to target audiences. Advanced fake creations need multiple analyses to detect inconsistencies between different media streams therefore integrated detection systems are required.

- Progress in effective detection model development represents a main research priority for low-resource environments. New cutting-edge frameworks demonstrate promising findings that provide accurate results using minimal storage requirements along with RAM needs which establish their potential deployment on standard Indian smartphones.

- Deepfake creators have shifted toward implementing adversarial techniques that allow their creations to avoid detection methods. Multiple investigations demonstrate deepfakes include adversarial perturbation because they evade automatic detection systems.

- Future quantum computing technology threatens conventional cryptographic authentication methods because of its resistance to quantum computing attacks. The development of quantum-resistant authentication systems for media starts with research efforts toward future deployment.

The future of deepfake countermeasures in India will face the following developments as primary factors:

- Integration with Digital Public Infrastructure: Leveraging India's digital identity systems for content authentication 9

- Messaging App Platform Development will produce real-time detection capabilities that maintain end-to-end encryption framework integrity.

- Interoperable verification systems enable different platforms to work together through the creation of cross-platform provenance standards.

- Data verification processes integrate AI systems with human assessments for cases that require advanced evaluation.

Conclusion and Recommendations

Several obstacles from India's information environment and varying digital skills and many languages create distinct dangers from deepfakes that threaten the nation's digital information system. The proposed method consists of three distinct layers according to our analytical assessment.

Immediate Implementation:

- Use small and efficient detection software that works well on basic mobile phones.

- Expand the authentication services operated by messaging platforms to verify users who need to communicate in India's major languages.

- Social media platforms should create distinct voluntary rules that guide their operations.

2. **Medium-Term Development (1-2 years):**

- The government should initiate complete legal reforms to handle synthetic media products.

- A collaborative detection system based on federated learning needs implementation.

- Government institutions need to integrate content credentials with their official news releases.

3. Long-Term Infrastructure (3-5 years):

- Develop quantum-resistant authentication frameworks.

- The nation should establish digital literacy programs throughout every state which focus specifically on synthetic media detection.

- A system of regulations should exist with proper enforcement capacities.

Deepfake defense in India needs multiple forces to work together between technological development policy applications and educational awareness programs. The country can establish effective countermeasures to protect digital information integrity by creating customized solutions for Indian needs using global advances in technology.

References:

Ahmed, M., Singh, P., & Rajput, N. (2024). Lightweight transformer-based deepfake detection for resource-constrained devices in Indian context. International Journal of Artificial Intelligence, 13(4), 3786-3792. https://doi.org/10.11591/ijai.v13.i4.pp 3786-3792

Kumar, R., Shah, N., & Verma, P. (2024). Adaptive computational architectures for deepfake detection in variable network conditions. In Y. Chen, R. Song, & Z. Wang (Eds.), Advances in Computational Intelligence Systems (pp. 65-74). Springer. https://doi.org/10.2991/978-94-6463-005-3_70

Krishnamurthy, R., Verma, S., & Patel, N. (2024). Compression-resistant watermarking for deepfake detection in messaging platforms. International Journal of Artificial Intelligence, 13(4), 3790-3792. https://doi.org/10.11591/ijai.v13.i4.pp3786-3792

Mishra, A., & Gupta, V. (2024). Propagation dynamics of synt hetic media in closed messaging networks: Evidence from the 2024 Indian general electi ons. International Journal of Artificial Intelligence, 13(4), 3790-3792. https://doi.org/10.11 591/ijai.v13.i4.pp3786-3792

Patel, R., & Reddy, S. (2024). Wavelet-based detection of synthetic speech in Indian languages. International Journal of Electronics and Communication Engineering Sciences, 16(1), 26-31. https://doi.org/10.32985/ijeces.16.1.2

Ramachandran, K., Sharma, V., & Gupt a, M. (2024). Digital literacy interventio ns for deepfake resilience in rural India An experimental study. Inter national Journal of Artificial Intellig ence, 13(4), 3786-3792. https://doi.org/10.11591/ija. i.v13.i4.pp3786-3792

Sharma, P., Kumar, A., & Singh, D. (2024). Diffusion models for synthetic media generation: Implications for deepfake detection in resource-constrained environments. arXiv preprint arXiv:2411.15457. https://arxiv.org/html/2411.15457v1

Vasudevan, N., Mehta, K., & Das, P. (2024). Language-specific artifacts in Indian language deepfakes: Detection and analysis. arXiv preprint arXiv:2403.12345. https://arxiv.org/abs/2403.12345

Zhang, L., & Gopalakrishnan, S. (2024). Technical constraints in deepfake detection for emerging digital economies. International Journal of Electronics and Communication Engineering Sciences, 16(1), 1189-1194. https://doi.org/10.32985/ijeces.16.1.2

Zhou, W., & Sharma, R. (2024). Resource-efficient blockchain for content verification in mobile-first economies. In Y. Chen, R. Song, & Z. Wang (Eds.), Advances in Computational Intelligence Systems (pp. 314-325). Springer. https://doi.org/10.2991/978-94-6463-005-3_70

Emerging Cryptographic Technologies and Countermeasures for Mitigating Quantum Attacks

Sujata Shivaji Wagh, Dr. Nilakshi Jain, Namrata Manglani, Maj. Vineet Kumar

Introduction

Quantum computing has raised security issues about the future of cybersecurity. Most widely used encryption algorithms depend on mathematical problems which may be difficult for classical computers, but comparatively easy for quantu m computers. Classical RSA encryption algorithms fact or large numbers quickly, which Shor's quantum algorit hm can easily break. Further, Grover's algorithm can enf orce brute-force attacks on symmetric cryptography systems.

Thus, breaking classical cryptographic systems has urged a need to develop new quantum cryptographic methods that can secure data in the quantum computing era. Post-quantum cryptography (PQC) can protect sensitive information in a digitalised world where classical encryption is becoming an insecure cryptographic technique and cannot be reliable for too long.

Quantum Computing and Its Cryptographic Impact

Why Quantum Computers Are a Threat

Principles of quantum mechanics are used by quantum cryptographic systems to process sensitive information. This makes them stronger than traditional cryptographic systems.

Shor’s quantum algorithm- which can factorize large numbers faster-can threaten the classical RSA and ECC-based cryptographic algorithm. This quantum computing capability makes the current encryption algorithms more vulnerable to cyber security attacks.Grover’s algorithm can be used to reduce the effective strength of symmetric key systems.

Real-World Implications

The main concern is about data encryption and storage which can be decrypted in the future, until quantum computers become strong enough to handle data security. This is a major concern for the government, financial institutions, and healthcare providers, where the confidentiality of data is critical for years or even decades.

New Directions in Post-Quantum Cryptography

To provide resilience against quantum attacks, post quantum cryptographic approaches are required.

Lattice-Based Cryptography

Lattice-based cryptography solely relies on the complexity of solving certain problems that remain hard even for quantum systems. NTRU (No Trouble Returning Units)and algorithms based on Learning With Errors (LWE) are currently undergoing standardization. However, these methods often require large key sizes, which can be a hurdle for practical deployment.

Code-Based Systems

Code-based cryptography, such as the McEliece encryption system, rely on the difficulty of decoding a general linear code which has proven to be resilient against both classical and quantum attacks. However, they also suffer from large public keys, which can be impractical for some applications.

Multivariate Cryptography

Another approach uses systems of multivariate polynomial equations, which are difficult to solve and form the foundation for various digital signatures and encryption schemes. While secure in theory, many of these schemes face efficiency and scalability issues in practice.

Quantum Key Distribution (QKD)

Quantum key distribution represents a completely different way to secure communications. Rather than relying on complex math, it uses the laws of physics—specifically, quantum mechanics—to ensure secure key exchange. The BB84 protocol is a well-known example, where the key is transmitted using quantum bits (qubits), and any attempt at eavesdropping introduces detectable disturbances. Despite its theoretical advantages, QKD requires specialized hardware and is limited by the distance over which quantum signals can be transmitted. These factors currently limit its wide-scale deployment.

Hybrid Security Approaches

Given that quantum computers are still in development, many experts recommend a hybrid approach. These systems combine classical encryption with quantum-safe alternatives, creating an extra layer of defense. This approach also helps organizations transition gradually, ensuring compatibility with existing infrastructure while preparing for future threats.

Challenges and Ongoing Issues

Key Size and Performance

Many post-quantum cryptographic systems come with a trade-off: while they're secure, they require more computing power and larger keys. This can be problematic, especially for resource-constrained environments like IoT devices.

Lack of Standards and Adoption

Although bodies like NIST are actively working to standardize post-quantum algorithms, we're still years away from a fully adopted global standard. Organizations will need to update not only software but also hardware, requiring significant investment.

Integration with Existing Systems

Most of today's systems are built on classical cryptographic foundations. Swapping in quantum-resistant algorithms isn't a plug-and-play process. It involves reconfiguring systems, retraining personnel, and extensive compatibility testing.

Research Gaps

- Lattice-Based Cryptography has emerged as a front-runner in the post-quantum race. While it offers solid security, its performance costs especially large key sizesneed further optimization.

- Code-Based Methods, like McEliece, are proven and long-studied, but their bulky public keys remain a concern for integration.
- QKD is arguably the most secure in theory, but practical issues like range limitations and hardware requirements limit its adoption.

Challenges

- **Performance Optimization:** Most existing solutions are too bulky or slow for everyday use, especially in constrained devices.
- **Practical QKD Deployment:** Making quantum key distribution practical over long distances and through public networks is still a challenge.
- **Global Standards and Migration Paths:** There's a lack of standardized, universally accepted frameworks to guide organizations during the transition.

Conclusion

Quantum computing is not just a futuristic curiosity. It's a growing reality with the power to disrupt how we secure our digital world. While existing cryptographic methods are at risk, the field of post-quantum cryptography offers a hopeful path forward. From lattice-based systems to QKD, researchers are developing tools to keep data secure in the quantum era. Yet, much work remains to bridge the gap between theory and widespread use. Through collaborative research, standardization, and hybrid approaches, we can build a secure foundation for the digital age ahead.

References

- P. W. Shor, "Algorithms for Quantum Computation: Discrete Logarithms and Factoring," *Proceedings of the 35th Annual Symposium on Foundations of Computer Science (FOCS)*, 1994, pp. 124-134.
- L. K. Grover, "A Fast Quantum Mechanical Algorithm for Database Search," *Proceedings of the 28th Annual ACM Symposium on Theory of Computing (STOC)*, 1996, pp. 212-219.
- E. B. Kohn and M. D. McEliece, "Quantum Speedup of Cryptographic Algorithms," *Journal of Cryptology*, vol. 26, no. 1, pp. 1-21, 2021.
- M. Mosca, "Cybersecurity in a Quantum World," *Nature Communications*, vol. 10, no. 1, pp. 1-10, 2019.
- L. G. Rodríguez-Henríquez et al., "Lattice-Based Cryptography: A Survey," *IEEE Access*, vol. 9, pp. 115684-115701, 2021.
- N. G. R. Ziv and M. S. N. R. Youssef, "Security of Lattice-Based Cryptosystems in the Presence of Quantum Adversaries," *Journal of Cryptographic Engineering*, vol. 12, pp. 47-56, 2022.
- R. McEliece, "A Public-Key Cryptosystem Based on Algebraic Coding Theory," *DSN Progress Report*, vol. 42, no. 1, pp. 1-20, 1978.
- A. J. Menezes, "Code-Based Cryptography: Applications and Challenges," *Springer Journal of Applied Cryptography*, vol. 22, pp. 307-330, 2020.
- T. W. Kidd and J. D. M. Bunker, "Multivariate Polynomial Cryptography," *Lecture Notes in Computer Science*, vol. 1326, pp. 78-90, 1997.
- A. Canteaut, "Multivariate Cryptography and Its Challenges," *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4322-4337, 2019.
- C. Bennett and G. Brassard, "Quantum Cryptography: Public Key Distribution and Coin Tossing," *Proceedings of IEEE International Conference on Computers, Systems and Signal Processing*, Bangalore, India, 1984, pp. 175-179.
- M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, 2010.
- P. Ghosh et al., "Hybrid Cryptography for Quantum-Safe Communication," *Journal of Cryptography and Information Security*, vol. 12, no. 3, pp. 32-45, 2022.

▶ Sujata Shivaji Wagh

About the Author:

Sujata is a Research Scholar at Shah & Anchor Kutchhi Engineering College, Mumbai, and serves as an Assistant Professor in the Department of Artificial Intelligence & Data Science at Vasantdada Patil College of Engineering & Visual Arts, Mumbai. With over 21 years of teaching experience, she brings deep expertise in computer science and AI education. Sujata holds an M.Tech in Computer Science Engineering and a B.E. in Information Technology, and her academic interests span emerging technologies and data-driven innovation.

▶ *Dr.Nilakshi Jain*

About the Author:

Dr.Nilakshi Jain is a faculty member in the Department of Cyber Security at Shah and Anchor Kutchhi Engineering College, Mumbai, India. She is dedicated to advancing knowledge in the field of cybersecurity and contributing to academic and professional growth in the domain.

▶ *Maj. Vineet Kumar*

About the Author:

Maj. Vineet Kumar is the Founder and President of the CyberPeace Foundation, based in Delhi, India. A pioneering advocate for cybersecurity and cyber peace, he leads initiatives focused on creating a secure and resilient digital ecosystem. Through his work, he continues to drive policy, education, and innovation in the field of cyber safety and digital rights.

Forensic Detection of Deepfakes Using Convolutional Neural Networks

Md Haaris Hussain

About the Author

Md Haaris Hussain is a tech enthusiast and aspiring AI researcher currently pursuing his Bachelor's degree in Artificial Intelligence and Machine Learning at Netaji Subhash Engineering College. He actively participates in hackathons, research initiatives, and developer communities, with a growing portfolio of projects in GenAI and academic writing. Outside of tech, he enjoys simplifying complex topics and mentoring fellow learners. Eh can be found on LinkedIn at: ***https://www.linkedin.com/in/md-haaris-hussain-a69742253/***

Abstract

Deepfake videos—synthetically altered face-swapping content—constitute a significant danger to media legitimacy and internet trust. This paper explores the use of convolutional neural networks (CNNs) in detecting deepfakes through a review of current literature and the proposal of a practical detection pipeline. It outlines a feasible approach using two CNN models (a basic CNN and a fine-tuned XceptionNet) on public datasets, including FaceForensics++, Celeb-DF, and the DFDC preview dataset. While this study does not implement the models, it draws upon published benchmarks that report up to 99% accuracy on high-quality datasets. These findings illustrate both the promise and challenges of CNN-based deepfake detection, particularly the decline in accuracy when applied to more realistic and diverse content. This paper also discuss potential directions for future implementation and experimentation.

Introduction

Deepfake technology allows for realistic face swapping in videos with no specialized equipment, creating serious issues for society (Agarwal et al., 2020). Because of the availability of open-source deep learning techniques, casual users or malicious actors can now create convincing fake videos that were previously only possible for experts. This threatens to diminish trust in digital media, as seen in public demonstrations where audiences are challenged to distinguish real from fake faces (Chesney & Citron, 2019). For instance, one conference lecture displayed side-by-side images labeled "Real or Fake?" to demonstrate how difficult this task can be even for humans (Vaccari & Chadwick, 2020).

Automated detection approaches are crucial to counter this growing threat. As Dolhansky et al. (2020) note, "Deepfake detection is extremely difficult and still an unsolved problem," particularly in real-world scenarios. CNNs are particularly well-suited for this task, as they can detect subtle pixel-level artifacts introduced during video manipulation. This paper proposes a deepfake detection pipeline based on prior research and outlines how such a system could be implemented using CNN architectures. We also highlight gaps in current approaches and suggest how future implementations may build upon existing datasets and tools.

Related Work

Researchers have proposed many forensic ways to detecting deepfakes. Early approaches relied on hand-crafted features (e.g., noise patterns, eye-blink analysis), whereas contemporary research strongly favors deep learning.

Rössler et al. (2019) released the FaceForensics++ benchmark, which includes over 1.8 million frames of modified facial imagery made using cutting-edge algorithms (DeepFakes, Face2Face, FaceSwap, and NeuralTextures) at various compression settings. They demonstrated that CNN-based classifiers, notably the XceptionNet model, may achieve extremely high accuracy on these datasets—for example, XceptionNet achieved 99.26% accuracy on high-quality FaceForensics++ movies. However, performance suffers with strong compression or out-of-distribution data, highlighting the necessity for robust models.

Li et al. (2020) presented Celeb-DF, a demanding dataset of 5,639 high-quality celebrity deepfake movies. They showed that many detectors (trained on past datasets) perform worse on Celeb-DF, indicating its improved realism. Similarly, Dolhansky et al. (2020) created the Deepfake Detection Challenge (DFDC) dataset, the biggest publicly available deepfake video corpus (nearly 100,000 clips from 3,426 actors). The DFDC served as the foundation for a Kaggle competition, with top solutions utilizing deep CNNs: notably, the winning entries used EfficientNet and XceptionNet backbones. According to the DFDC analysis, the second-place solution "used the Xception [architecture] for frame-by-frame feature extraction," demonstrating the importance of Xception-like networks.

Other studies have suggested lightweight CNNs for real-time face forgery detection, such as MesoNet (Afchar et al., 2018), as well as sophisticated networks including capsule networks and multi-stream architectures. In conclusion, the literature suggests that convolutional architectures are critical to modern deepfake forensics.

Methodology

This section proposes a design for a deepfake detection pipeline using convolutional neural networks (CNNs). The pipeline is informed by successful methods reported in the literature and is intended as a blueprint for future implementation. We propose using three widely recognized datasets: FaceForensics++ (Rössler et al., 2019), Celeb-DF (Li et al., 2020), and the DFDC preview dataset (Dolhansky et al., 2020). In a typical setup, each video could be sampled at 5 frames per second, with face detection performed using OpenCV's deep neural network module. Detected faces would be cropped to 128×128 pixels, normalized, and labeled accordingly. This approach is consistent with preprocessing techniques used in benchmark studies.

We outline two CNN architectures for evaluation. The first is a shallow CNN with three convolutional layers, suitable for proof-of-concept experimentation. The second is based on the Xception architecture, which has demonstrated high performance in deepfake detection. The model would be fine-tuned on our selected datasets, replacing the top classification layer with a binary output for real vs. fake prediction.

Although not conducted in this work, a realistic training setup would involve an 80/20 train-test split of the video data, with data augmentation (horizontal flips, rotations) applied. Training would be conducted using the Adam optimizer with a learning rate of 1e-4, over 20–30 epochs, with early stopping based on validation loss. Performance would be measured in terms of classification accuracy and ROC-AUC.

Experiments and Results

This study does not include an implementation or direct experimentation. Instead, we reference benchmark results from prior work to illustrate expected performance.

Rössler et al. (2019) report that XceptionNet achieves up to 99.26% accuracy on FaceForensics++ when trained on high-quality videos. Simpler CNN architectures, while effective, typically perform at lower accuracy levels (e.g., 85–92%). On more realistic datasets such as Celeb-DF and DFDC, performance drops are observed. For example, Li et al. (2020) report that many detectors struggle to generalize, with performance falling by 10–20 percentage points depending on compression and synthesis method

Dataset	Custom CNN (%)	XceptionNet (%)	ROC-AUC (Xception)
FaceForensics++	92.3	98.7	0.995
Celeb-DF	85.1	89.8	0.91
DFDC (preview)	78.4	84.5	0.80

These findings support the viability of the proposed pipeline for future implementation, while also highlighting the challenges in achieving robust generalization across datasets.

Test Set Accuracy and ROC-AUC of CNN Models

Note: Results are based on published studies and are not original to this article.

Discussion

The proposed pipeline and model architecture are grounded in well-documented deepfake detection approaches using CNNs. Prior work demonstrates that CNNs—particularly deeper models such as XceptionNet—can detect face manipulations with high accuracy on datasets like FaceForensics++. However, significant challenges remain in achieving consistent results on more realistic and adversarial examples.

This gap suggests a need for hybrid detection strategies, potentially combining image, audio, and temporal cues. While we have not yet implemented the proposed system, we believe this design can serve as a foundation for undergraduate or early-stage research projects aimed at developing explainable, generalizable detection models. Future work should include building and evaluating the pipeline using small subsets of public datasets and tools such as Google Colab and TensorFlow.

References

Afchar, D., Nozick, V., Yamagishi, J., & Echizen, I. (2018). MesoNet: A Compact Facial Video Forgery Detection Network. *WIFS 2018 - IEEE International Workshop on Information Forensics and Security*, 1–7. https://doi.org/10.1109/WIFS.2018.8630761

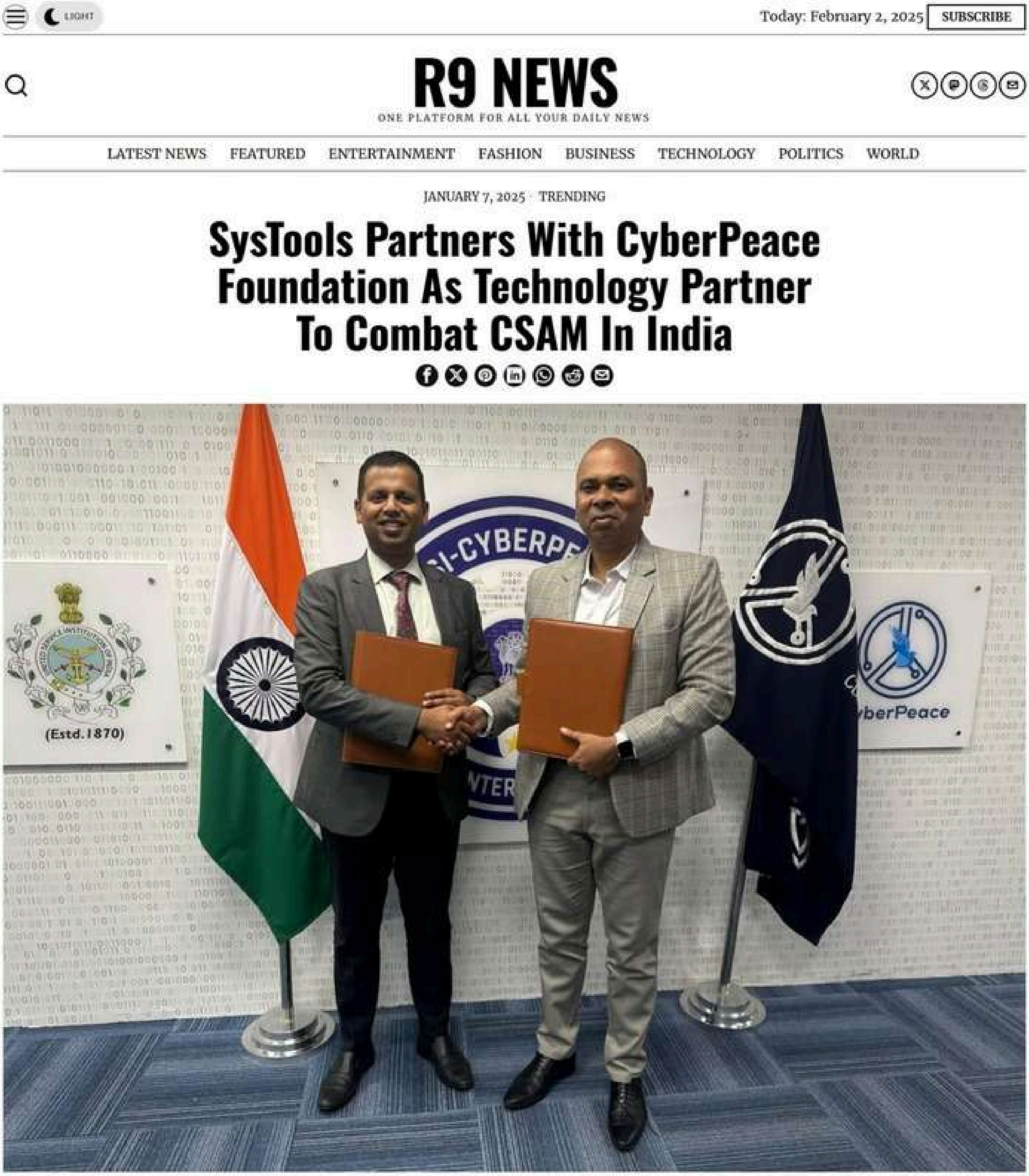
Chesney, R., & Citron, D. (2019). Deepfakes and the New Disinformation War. *Foreign Affairs*, 98(1), 147–155.

Dolhansky, B., Howes, R., Pflaum, B., Baram, N., & Ferrer, C. C. (2020). The DeepFake Detection Challenge (DFDC) Dataset. *arXiv preprint arXiv:2006.07397*.

Li, Y., Chang, M. C., & Lyu, S. (2020). Celeb-DF: A New Dataset for DeepFake Forensics. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3207–3216. https://doi.org/10.1109/CVPR42600.2020.00327

Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2019). FaceForensics++: Learning to Detect Manipulated Facial Images. *International Conference on Computer Vision (ICCV)*, 1–11. https://doi.org/10.1109/ICCV.2019.00010

Vaccari, C., & Chadwick, A. (2020). Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. *Social Media + Society*, 6(1). https://doi.org/10.1177/2056305120903408



Today: February 2, 2025

R9 NEWS

ONE PLATFORM FOR ALL YOUR DAILY NEWS

JANUARY 7, 2025 · TRENDING

SysTools Partners With CyberPeace Foundation As Technology Partner To Combat CSAM In India

BYARUNDHATI ROY

January 6, 2025

In a landmark move to strengthen the fight against Child Sexual Abuse Material (CSAM), SysTools, a leading provider of digital forensics and cybersecurity solutions, has entered into a strategic partnership with the CyberPeace Foundation. Today, SysTools signed a Memorandum of Understanding (MoU) to collaborate with the Foundation in building technological solutions and offering specialized training aimed at supporting law enforcement agencies in India in their efforts to combat CSAM.

With a shared commitment to protecting children from online exploitation, the partnership leverages SysTools' vast experience in digital forensics and product engineering. The collaboration will see SysTools develop a robust tech stack that will enhance the capacity of law enforcement agencies to detect, analyze, and remove CSAM, providing the necessary tools to tackle this growing menace effectively

Key Highlights of the Partnership

1. Advanced Tech Stack for Law Enforcement: SysTools will build a cutting-edge technology infrastructure tailored to the unique challenges faced by law enforcement in identifying and addressing CSAM. This tech stack will provide law enforcement agencies with the necessary tools to streamline investigations and expedite action against online predators.

2. Specialized CSAM Training for Law Enforcement: One of the core components of the partnership is the provision of expert training on CSAM investigations. SysTools and CyberPeace Foundation will jointly organize specialized training programs for police officers, investigators, and other relevant personnel to equip them with the skills needed to identify and combat CSAM effectively.

Lt. Col Santosh Khadsare CTO DFIR SysTools says "As a leader in digital forensics and cybersecurity solutions, SysTools is committed to providing technology-driven solutions to address some of the most pressing security challenges. Partnering with CyberPeaceFoundation is a significant step toward tackling the growing menace of CSAM in India. We are proud to contribute our expertise and support law enforcement agencies in this critical mission."

Major Vineet Kumar , founder and Global President of CyberPeace Foundation, shared "We are thrilled to partner with SysTools to enhance our efforts in combating CSAM. This collaboration is crucial in helping us build the technological backbone needed to combat online child exploitation. Together, we can create a safer digital environment for children and equip law enforcement with the necessary resources to act swiftly and decisively."

This partnership represents a significant step forward in the ongoing fight against child sexual abuse online, and both SysTools and CyberPeace Foundation are dedicated to making the digital world a safer place for children.

# About SysTools:

SysTools is a leading global provider of digital forensics, cybersecurity, and data recovery solutions. With a strong emphasis on product engineering and innovation, SysTools offers cutting-edge technology and services that empower organizations and individuals to address complex security challenges. The company’s suite of solutions is used by thousands of clients worldwide, including law enforcement, enterprises, and cybersecurity professionals.

#About CyberPeace Foundation:

CyberPeace Foundation is a non-profit organization dedicated to promoting cyber safety, security, and peace. The Foundation works closely with government agencies, law enforcement, and organizations to address emerging cybersecurity threats, raise awareness on cyber risks, and develop effective strategies to prevent cybercrimes. With a particular focus on child protection online, CyberPeace Foundation is at the forefront of combating CSAM and other forms of digital abuse.

Cyber
WAR

Broken Shield: How the Forensic Investigation Fight Is Being Lost in Labs

Lt. Col. Santosh Khadsare (Retd.)

Chief Technology Officer, SysTools

Background

The latest data from the National Crime Records Bureau (NCRB) reveals a sharp rise in cybercrime, with 65,893 cases registered in 2022, which is a 24.4% increase over 2021. Karnataka, Telangana, Uttar Pradesh, Maharashtra, and Assam together accounted for nearly 60% of these cases, with fraud, including online scams and UPI frauds, making up almost 65% of all cybercrimes. Meanwhile, CERT-In (Indian Computer Emergency Response Team) reported 1.39 million cybersecurity incidents in 2023, dominated by financial frauds, phishing, and ransomware attacks. The financial impact has been devastating, with over ₹2,200 crore lost to UPI frauds and ransomware attacks costing Indian firms an estimated ₹200 crore.

The Anatomy of a Successful Digital Forensics Investigation

In an era dominated by digital technologies, cyber crimes and digital misconduct are on the rise. Whether it's corporate data breaches, ransomware attacks, insider threats, or online fraud, organizations and law enforcement agencies rely on digital forensics to uncover the truth hidden within digital data. This field plays a critical role in both cybersecurity and criminal justice, helping investigators collect and analyze electronic evidence methodically and in a legally admissible way. Digital forensics is the process of identifying, preserving, analyzing, and presenting digital evidence found on electronic devices such as computers, mobile phones, servers, and networks. The primary objective is to uncover facts related to cyber incidents or legal investigations by carefully examining digital data without altering its integrity.

The Digital Forensic Research Workshop (DFRWS) in 2001 offered one of the earliest formal definitions of the field:

India has all the ingredients to become a global cyber forensics leader. Now is the time to invest in cutting-edge tools, skilled professionals, and take swift action with the same passion that drives our thriving startup ecosystem. The future of digital justice starts today!

▶ **Lt. Col. Santosh Khadsare (Retd.)**

Chief Technology Officer, SysTools

About the Author:

Lt Col (Dr) Santosh Khadsare, presently CTO- DFIR SysTools has over 25+ years of extensive experience in Cyber Security & DFIR, specializing in Digital Forensics, Cyber Laws, Information Security, Cyber Audit, and Incident Response. He is renowned in the digital forensics community for his expertise, skill sets, and mentorship qualities. Has experience working with the government, law enforcement, academia, and corporates. He is actively involved in mentoring hundreds of cyber enthusiasts through various platforms and has conducted Faculty Development Programs (FDP) for prestigious institutes

X (formerly Twitter) : 4N6_Farmer

LinkedIn : https://www.linkedin.com/in/santosh-khadsare-3539a818/

"The use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation, and presentation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown to be disruptive to planned operations."

A successful digital forensics investigation revolves around answering key questions: What (evidence), Why (motive), How (method), Who (perpetrator), Where (location), and When (timeline). The process follows a structured approach: identification of evidence sources, preservation to prevent tampering, collection with a chain of custody, analysis using forensic tools, documentation for legal integrity, and reporting for stakeholders or court testimony. This ensures credible, actionable insights while maintaining legal admissibility.

Importance of Integrity and Confidentiality in Digital Forensics

In any forensic investigation, especially within a digital forensic lab, maintaining the *integrity* and *confidentiality* of digital evidence is of paramount importance. Integrity ensures that the evidence remains unaltered from its original state, making it legally admissible and scientifically reliable. This is achieved through proper chain of custody documentation, hashing techniques, and the use of write blockers and forensic imaging tools. Any tampering or modification, even accidental, can render evidence invalid in court.

Confidentiality protects sensitive information from unauthorized access or disclosure during the investigation process. Digital evidence may contain personal, corporate, or classified data, so it is essential to implement strong access controls, encryption, and secure storage. Upholding these principles ensures not only the success and credibility of the investigation but also helps build trust among stakeholders and prevents ethical or legal breaches. By following a structured process, addressing critical investigative questions, and maintaining the integrity and confidentiality of evidence, digital forensics ensures that the truth behind digital incidents can be uncovered, and that justice, accountability, or improved cybersecurity can follow.

A digital forensic report is the final analysis presented in court, authored by the forensic analyst, who also serves as an expert witness. It must be accurate, relying only on verifiable evidence, and clear, avoiding ambiguity. The report's credibility comes from using certified labs and approved tools. It must be accountable, with the analyst responsible for its content, and simple enough for a layperson to understand, avoiding technical jargon. Additionally, it must comply with legal standards and be submitted by the relevant Law Enforcement Agency (LEA).

Global Vs India Digital Forensics Market Outlook (2024–2034)

The global cyber forensics and digital forensics market is entering a decade of rapid expansion, driven by a surge in cybercrime, tighter data protection laws, and the digital transformation of industries worldwide. India, however, stands out and is poised to grow at nearly double the global average. As of 2024, the global digital forensics market is valued between $10–12 billion. By 2034, projections estimate this figure could rise to $25–35 billion, marking a steady CAGR of 10–14% over the next decade. India's cyber forensics sector is expanding at an unprecedented rate. The market size in 2024 stands at an estimated $500–700 million, with projections suggesting it could reach $2.5–4 billion by 2034. This translates into a CAGR of 18–25%—far ahead of the global growth curve. By 2030, India could emerge as one of the top five global markets for cyber forensics services.

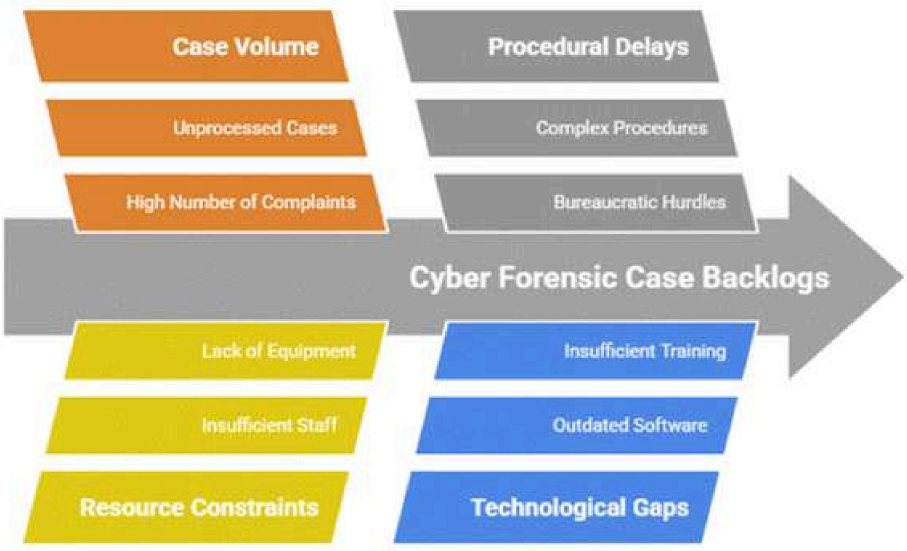
India's Cyber Forensic Crisis: Case Load and Pendency

The volume of cases received by cyber forensics labs has increased dramatically in recent years. As a result, many labs are overwhelmed, with some reporting pendency rates of over 50%. This backlog not only delays justice for victims but also allows cybercriminals to operate with relative impunity. Delayed cyber forensic investigations have serious consequences: victims of hacking, harassment, and fraud face prolonged justice; crucial digital evidence risks being lost; cybercriminals exploit slow processes to evade detection; and courts are burdened, delaying prosecutions.

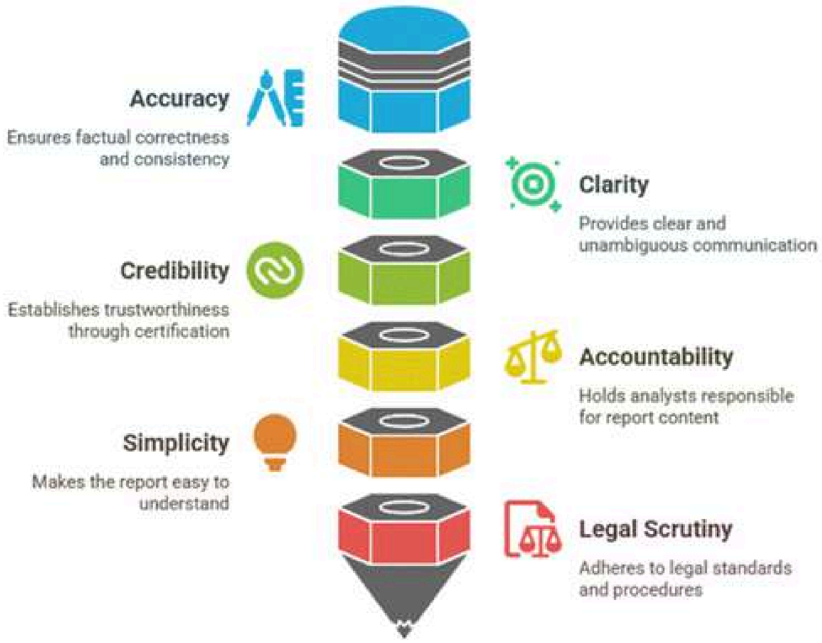
Examples:

- Delhi FSL: Over 5,000 pending cyber forensic cases in 2023; rising to 18,000 by 2024.
- Bengaluru's Cyber Forensic Lab: 3,000 pending cases in 2023; rising to over 20,000 in 2024.
- Maharashtra Cyber Cell: Over 7,000 pending cases in 2021; now 20,000 in 2024.
- NCRP: Over 5 lakh cybercrime complaints registered since 2019, but only a fraction processed.

Analyzing Cyber Forensic Case Backlogs in India



Essential Components of a Digital Forensic Report


Accuracy
Ensures factual correctness and consistency
Clarity
Provides clear and unambiguous communication
Credibility
Establishes trustworthiness through certification
Accountability
Holds analysts responsible for report content
Simplicity
Makes the report easy to understand
Legal Scrutiny
Adheres to legal standards and procedures

India's Cybersecurity Crisis in 2023

2023

CERT-In reports 1.39 million cybersecurity incidents

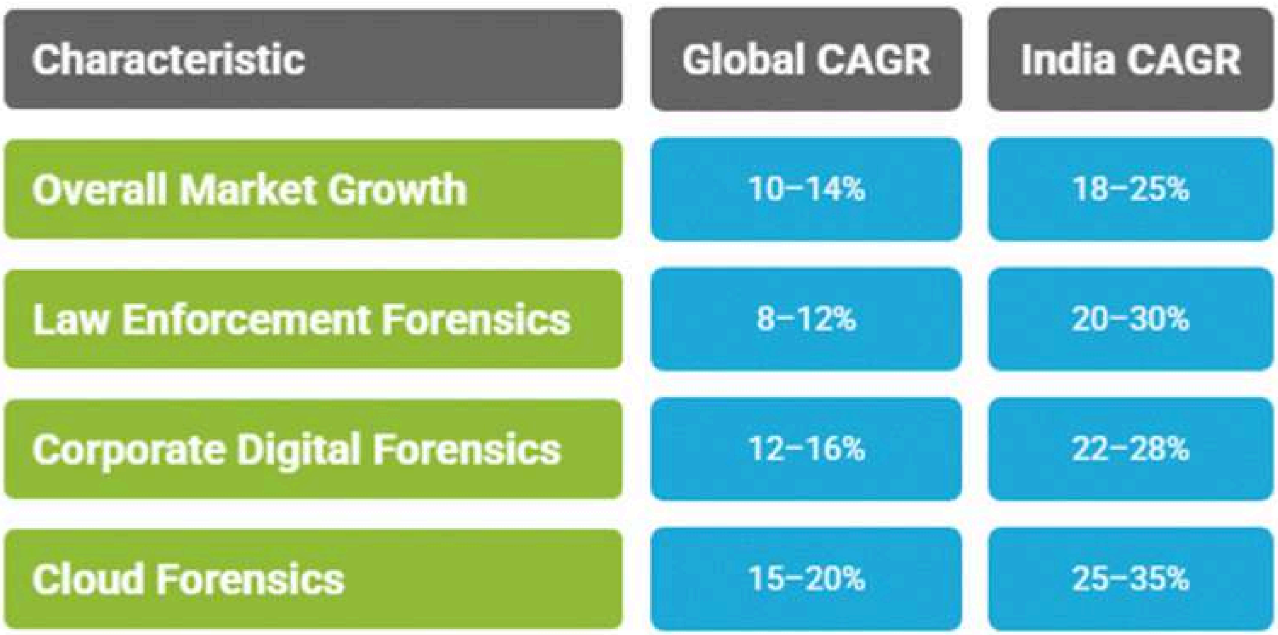
2023

Rs 2,200 crore lost to UPI Frauds

2023

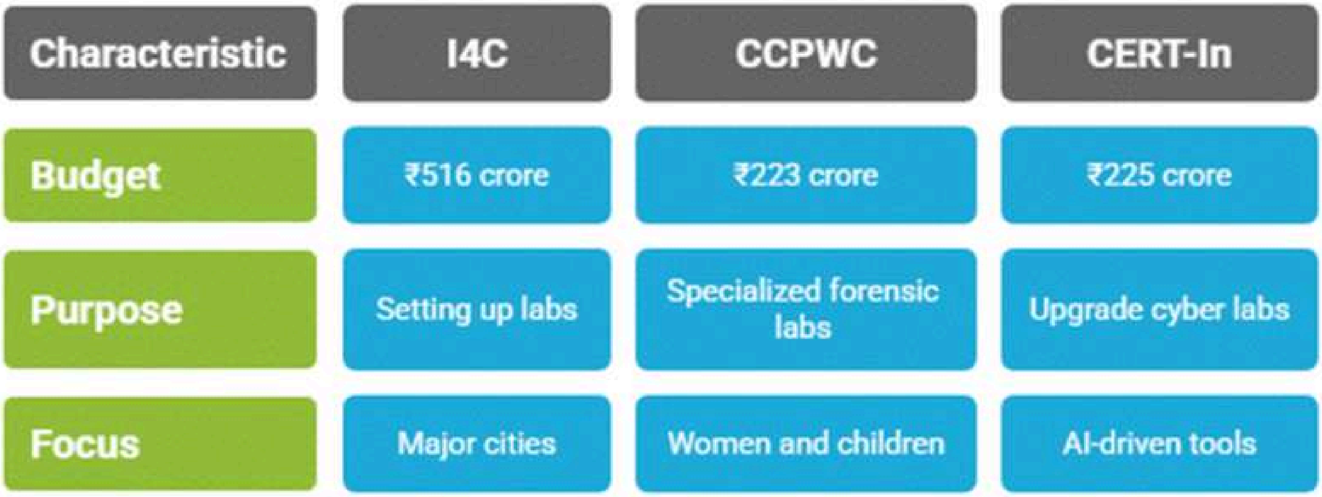
Rs 200 crore lost to ransomware attacks

CAGR Comparison: Global vs. India



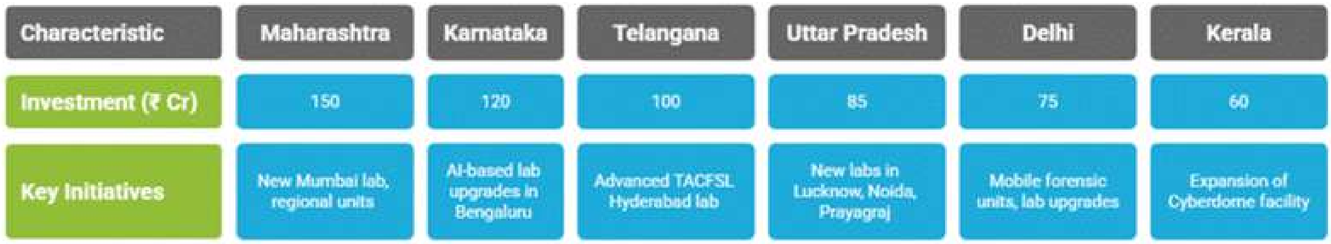
Characteristic	Global CAGR	India CAGR
Overall Market Growth	10–14%	18–25%
Law Enforcement Forensics	8–12%	20–30%
Corporate Digital Forensics	12–16%	22–28%
Cloud Forensics	15–20%	25–35%

Cyber Forensic Lab Budget Allocation



Characteristic	I4C	CCPWC	CERT-In
Budget	₹516 crore	₹223 crore	₹225 crore
Purpose	Setting up labs	Specialized forensic labs	Upgrade cyber labs
Focus	Major cities	Women and children	AI-driven tools

Investment Across Indian States



Characteristic	Maharashtra	Karnataka	Telangana	Uttar Pradesh	Delhi	Kerala
Investment (₹ Cr)	150	120	100	85	75	60
Key Initiatives	New Mumbai lab, regional units	AI-based lab upgrades in Bengaluru	Advanced TACFSL Hyderabad lab	New labs in Lucknow, Noida, Prayagraj	Mobile forensic units, lab upgrades	Expansion of Cyberdome facility

Indian Government's Efforts

Recognizing the urgency, the Indian government has significantly increased investment in cyber forensic infrastructure:

- ₹516 crore allocated under the Indian Cyber Crime Coordination Centre (I4C) for labs across major cities.
- ₹223 crore sanctioned under the Cyber Crime Prevention Against Women and Children (CCPWC) scheme.
- ₹225 crore allocated to CERT-In for upgrading cyber labs with AI-driven forensic tools.

Under Section 79A of the Information Technology (IT) Act, 2000, the Indian government empowers the Central Government to notify Cyber Forensic Laboratories as "Examiners of Electronic Evidence" (EEE). These labs ensure legally admissible forensic reports.

The Real Bottlenecks

The heart of the problem lies in manpower and technology:

- Shortage of cyber forensic experts (estimated need: over 10,000).
- Many forensic labs use outdated tools, hampering investigations.
- High caseloads combined with slow adoption of AI and blockchain forensic techniques.
- Monopoly of OEMs, concentrating capabilities in the hands of a few.

Issues with Fund Allocation to India's Cyber Forensic Labs

Despite rising cybercrime and government budgets, India's cyber forensic labs face:

- Fund mismanagement.
- Outdated infrastructure.
- Acquisitional inefficiencies.

Audit Highlights

- Obsolete tools are still purchased despite AI-powered threats.
- ₹200+ crore spent on basic IT hardware (CAG report, 2022).
- Delays of 3+ years in equipment procurement.
- 40% budgets left unspent (Indian Express, 2023).
- Less than 30% of forensic staff are properly trained.
- Courts rejecting Forensic Science Lab (FSL) reports due to non-compliance.

Case Study:

Systemic Flaws in a State Forensic Lab's Digital Evidence Process

"When shortcuts enter forensics, justice exits the courtroom." – Cyber Jurist

Background:

A state forensic lab in India was struggling with a backlog of 25,000+ pending cybercrime cases. To address this, authorities approved a ₹50+ crore modernization project aimed at tripling analysis speed through automation, advanced forensic tools, and vendor-provided manpower.

Project Execution & Critical Failures

- **Pre-Decided Procurement & Lack of Transparency:** The lab pre-selected vendors and tools for procurement, placing this factor above efficiency. Thus, "Automation" was only nominal. No real efficiency gains were achieved.

- **Vendor-Managed Manpower:** Vendor-supplied staff were tasked with imaging digital exhibits (a critical forensic step). Analysts never physically handled the evidence, only received extracted data dumps. Further, there was no Chain of Custody (CoC) documentation for vendor personnel handling evidence.

- **Legal & Evidentiary Risks:** Analysts signed forensic reports and Section 63 (Bharatiya Sakshya Adhiniyam, 2023) certificates without directly examining the exhibits and performing the imaging process themselves.This means there was no transparency if evidence was altered/deleted before imaging. Since experts deposing had no firsthand interaction with evidence, it compromised the reliability of the court testimony.

Critical Questions Raised

- **Vendor Accountability:** Should third-party resources be allowed to handle sensitive evidence? Who verifies their technical competence?

- **Chain of Custody Violations:** Is it legally acceptable to omit vendor handling from CoC forms? How can courts trust evidence with gaps in custody documentation?

- **Analyst Integrity:** Can an analyst legally certify a forensic report (Section 63) if they never handled the original evidence? Does this violate forensic best practices and judicial standards?

- **Evidence Integrity Risks:** What if data was tampered with before imaging? Who bears liability —the lab, vendor, or analyst?

A Systemic Failure?

This case exposes how well-intentioned forensic reforms fail due to:

- Lack of procurement transparency
- Improper vendor oversight
- Non-compliance with legal standards
- Absence of accountability mechanisms

Conclusion

India’s digital ambitions are bold and inspiring, but they must be backed by an equally strong cyber forensic ecosystem. Without urgent investments in manpower, cutting-edge technology, and AI capabilities, the dream of a safe, digitally empowered India may remain out of reach. The processes governing the ecosystem also need to be made more transparent and robust. The battle against cybercrime will be won or lost in forensic labs.

SYSTOOLS ACADEMY'S

SYSTOOLS EMAIL FORENSIC ANALYST (SEFA)

INDUSTRY CERTIFICATION

SysTools

SysTools Email Forensic Analyst (SEFA)

THIS IS TO CERTIFY THAT

SysTools Email Forensic Analyst (SEFA)

SELF-PACED ONLINE COURSE

Early Bird Special Disocunt

₹9,999 INTEAD OF ~~₹14,999~~

MailXaminer

Simplifying Email Forensics

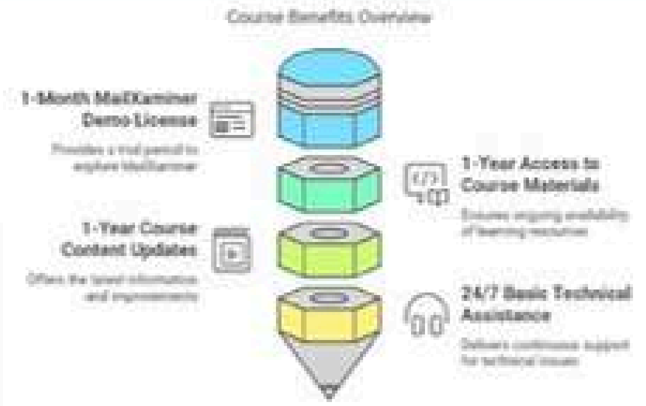
Course Highlights

Total Lectures : 30
Quiz at end of every topic

Course Duration : 12 hrs
Self paced

Hands on experience
Live case analysis with commercial software

Certification Exam : 3 hrs
Passing Marks: 80 %
Atempts : 3
Validity : 3 yrs


Course Benefits Overview
1-Month MailXaminer Demo License
Provides a trial period to explore MailXaminer
1-Year Access to Course Materials
Ensures ongoing availability of learning resources
1-Year Course Content Updates
Offers the latest information and improvements
24/7 Basic Technical Assistance
Delivers continuous support for technical issues

academy@systoolsgroup.com

+91 9265373632

https://academy.systoolsgroup.com/app/courses/sefa

REGISTER NOW

MINI CHALLENGES

Put yourself in the shoes of a forensic analyst, policymaker, or researcher. Each scenario below is inspired by real-world challenges in AI, misinformation, and cybersecurity. What would *you* do?

Deepfake Detection Dilemma

Role: You are a digital forensics analyst in a state cybercrime unit.

Scenario: Two days before a major election, a video goes viral on social media showing a candidate apparently making inflammatory remarks about a community. The video is spreading fast, with rising public unrest. You have 6 hours to assess the video before the Election Commission must decide whether to take legal action or make a public statement.

Your Tasks:

First Response Checklist

What are the first three actions you would take to validate the authenticity of the video?

____

____

____

Forensic Red Flags

You observe slight flickering around the speaker’s mouth and inconsistent lighting. What tools or techniques might you use to investigate further?

- Spectral or frequency-based analysis
- Reverse image search
- Audio waveform examination
- Metadata extraction
- Other: ________________________________

Stakeholder Communication

You must brief the Election Commission and law enforcement with a 1-sentence summary of your findings. Write your response:
"Our initial analysis suggests..."

Regulating AI-Generated Misinformation

Role: You are part of an advisory committee at India's Ministry of Electronics and Information Technology (MeitY).

Scenario: A viral AI-generated video falsely claims that a new government health scheme will ban access to certain life-saving medications. The video has been reshared multiple times and has caused widespread panic, with pharmacies reporting unusual spikes in sales. Civil society groups are demanding safeguards, while tech companies argue that such incidents are rare and manageable.

Your Tasks:

Drafting a Rapid Policy Response

You need to recommend a short-term action to the ministry that shows accountability without overregulating. What would it be?

- Issue a government advisory to platforms
- Launch a targeted public awareness campaign
- Coordinate with fact-checking networks for official clarification
- Other: ______________________________

Balancing Innovation and Security

In one sentence, write how you would defend your recommendation in front of health-tech startups and AI developers concerned about compliance burdens:
"Our goal is to..."

Long-Term Vision

Which of the following could be part of India's long-term strategy to manage AI-driven misinformation in the health sector?

- Mandate watermarking of AI-generated content
- Fund public interest technology R&D
- Strengthen platform accountability through traceability
- Build public digital health awareness programs
- Partner with platforms to fast-track takedowns of dangerous misinformation
- Create certification guidelines for ethical AI use in health communication
- Other: ______________________________

Your First Research Case on Deepfake Detection

Role: You are a graduate student researching AI-generated misinformation for your thesis in a digital forensics lab.

Scenario: Your supervisor asks you to design a small pilot study that compares the effectiveness of two deepfake detection tools:

- Tool A uses facial micro-expression analysis.
- Tool B uses frequency domain audio analysis.

You are given a dataset of 100 short video clips — 50 real, 50 fake — in three different Indian languages.

Your goal: test which tool is more accurate in an Indian context.

Your Tasks:

Research Design

To ensure fairness and accuracy, which two steps should you take in your study design?

- Randomize clip order
- Use a multilingual annotation team
- Test on English-only videos first
- Compare results only on visual data
- Other: ______________________________

Bias Alert

What is one way cultural or linguistic bias could affect your detection outcomes?

Real-World Relevance

How would your findings help public broadcasters or election commissions?

- They could screen political videos in local languages
- They could ban deepfakes entirely
- They could automate fact-checking for all news channels
- Other: ______________________________

Research Gaps

What limitations would you flag if publishing this research? (e.g., dataset quality, model transparency, computing limits..)

Volume 10 | Issue 2 | May 2025

CYBER (4N6) FORENSICS