IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, June 2026
Impact Factor: 8.2

Deepfakes, Public Trust in Digital Media, and Political Propaganda: The Indian Perspective

Ramu Rampreet Prajapati[1] and Sandhya Kaprawan[2]
[1]MS Cyber Security & [2]Assistant Professor, Department of Information Technology,
Mumbai University, Kalina, Mumbai, India

Abstract: *The Advancement in artificial intelligence now allows anyone to create highly realistic synthetic audio and video, commonly known as deepfakes [1, 2]. In the political point, deepfakes can be used to manipulate public opinion, spread misinformation in specific voter groups, and potentially influence electoral decision-making and outcomes [3, 4]. This review investigates how deepfake technology impacting democratic processes, with a specific focus on the Indian election system [5]. Recent election in India have noticed the use of AI-generated voice fakes and regional language adaptations to increase political outreach, creating concerns of their potential role or misuse in shaping public opinion and voter behaviour [5,6]. The continuous circulation of highly realistic deepfake contents is creating confusion and reducing public trust in digital media [7,8]. It is now creating a situation like "liar's dividend," where anyone accused of misconduct can challenge the authenticity of legitimate audiovisual evidence by saying that such material is artificially generated or manipulated [9,10].*

Exiting research highlights the ongoing race between the tools or techniques to create ultra realistic deepfake contents and the detection methods to identify manipulated content. Deep learning detection models are achieving high accuracy in controlled lab settings, but consistently they are unable to identify new, unseen deepfakes in real-world environment due to cross-dataset generalization and variations in content generation techniques [11, 12]. Legal regulations also faces limited effectiveness to handle the rapidly evolving deepfake technologies. India currently attempts to regulate modern AI manipulation using older frameworks like the Information Technology Act of 2000 [13, 14]. Which creates huge gaps, when compared to proactive international frameworks like the European Union's Artificial Intelligence Act [15, 16]. To secure electoral integrity, this paper suggests for a multi-layered defence framework. The framework should integrate robust detection systems capable of generalizing to previously unseen data, specific law and regulations that mandating transparent labelling of AI-generated content, and development of public media literacy initiatives against manipulated media [17].

Keywords: *Deepfakes, Artificial Intelligence, Political Propaganda, Digital Forensics, Media Trust, Electoral Integrity, Cyber Law, Misinformation, Election commission of India, Deep learning.*

I. INTRODUCTION

A. Background and Motivation

Artificial intelligence and deep learning algorithms now allow anyone to create highly realistic fake audio, video, and text [1,2]. These manipulated digital files are known as deepfakes. They depend on complex computational models, specifically Generative Adversarial Networks (GANs), to generate or manipulate events that never happened [3,4]. Political campaigns regularly using this technology to spread false information and manipulate public opinion during critical voting periods [5,6]. The ultra realistic contents create confusion in individuals to differentiate between authentic reporting and artificially generated or manipulated media [7,8].

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 1, June 2026 Impact Factor: 8.2

The Indian democratic system faces serious vulnerabilities to this technological threat. India has over 500 million internet users with highly different levels of digital media literacy [9,10]. Political parties are increasing the use of AI-generated content to target specific demographic groups [11]. They use voice cloning tools to translate political messages into multiple regional languages, instead of traditional media they use social media platforms for spreading these videos directly on encrypted messaging platforms like WhatsApp [12,13]. This rapid circulation of localized, synthetic media creates a challenge to free and fair elections [14,15].

B. Research Objectives

This Systematic Literature Review achieves four primary research objectives.

First, it categorizes the main generative models used to create synthetic media and evaluate how the complexity of these models affects the ability to detect deepfakes in real time [16, 17].

Second, it examines how exposure to AI-generated political propaganda increases uncertainty and reducing public trust in digital journalism [18,19].

Third, review highlights cross-dataset generalization limitations in current deep learning-based detection models, especially when they face real-world data containing previously unseen manipulation patterns. [16,20].

Fourth, it checks key gaps in India's existing cyber law, the Information Technology Act, 2000, and suggests regulatory improvements comparing by global best practices [21,22,23].

C. Scope of the Review

This review focuses on empirical studies, digital forensic analyses of previous papers, and Research Papers published between 2018 and 2025 [16,24]. This study investigates synthetic media manipulation techniques used for visual, audio, and textual content [2,25]. The primary focus of this research is the socio-political influence of deepfakes on India's democratic ecosystem [5,26]. A comparative analysis between the Indian laws and international regulatory models, like the European Union's Artificial Intelligence Act and election-related regulations implemented at the state level in the United States [27, 28].

D. Methodology and Organization

This paper follows a Systematic Literature Review (SLR) methodology. It extracts and analyses peer-reviewed studies across three distinct domains: digital forensics, political communication, and cyber law [16, 25, 29].

This paper covers the technological foundations of synthetic media generation, current gaps in existing literature, weaponization of deepfakes in India and their psychological impact on public trust, current digital detection models and their limitations, compare global legislative frameworks against Indian legal gaps and propose multi-dimensional countermeasures.

II. TECHNOLOGICAL FOUNDATIONS OF SYNTHETIC MEDIA

A. Generative Architectures

Synthetic media creation depends on deep learning algorithms to generate events that never occurred [1]. Generative Adversarial Networks (GANs) act as core computational architecture for manipulation [2,3]. A GAN is a continuous optimization mechanism that takes place between two separate neural networks [2]. Within a GAN framework, the generator creates artificial digital content, while the discriminator checks its authenticity by comparing it against genuine data samples [2,4]. Iterative feedback enables the generator to enhance the quality of the synthesized content, ultimately producing highly realistic media capable of misleading the discriminator [4,5].

Autoencoders act as a critical function in feature manipulation, as primary mechanism for facial replacement [6,7]. The encoder transforms the original facial data into a compact, lower-dimensional latent representation [2]. Along with, a target-specific decoder reconstructs that encoded information, enabling the source face to be integrated seamlessly with the target body [2]. Unlike conventional autoencoders, Variational Autoencoders (VAEs) with probabilistic modelling

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, June 2026
Impact Factor: 8.2

enabling the generation of more diverse and variety outputs [2]. The advancement in generative modeling such as diffusion models act an effective framework for creating high-quality media through progressive denoising [8]. These models form the foundation of modern text-to-image generation systems [8].

B. Modalities of Manipulation

Synthetic manipulation consists three primary modalities: visual, audio, and textual [9]. Visual deepfake techniques includes complex modifications, such as face swapping, lip-sync generation, and full body simulation [10,11]. Face-swapping algorithms enable the smooth transfer of facial attributes from a source face to a target face [12,13]. Reenactment techniques work as replicating the movements and facial expressions of a source object to a target subject for generating synchronized and realistic actions [11].

Audio manipulation uses advanced Text-to-Speech (TTS) systems and voice cloning networks [14]. With the help of acoustic datasets, modern architectures effectively reproduce an individual's unique vocal features, including intonation, pitch, and speech patterns [14,15]. Speech-to-speech conversion extends this capability by modifying speaker's voice to match a target identity without changing the original message [16].

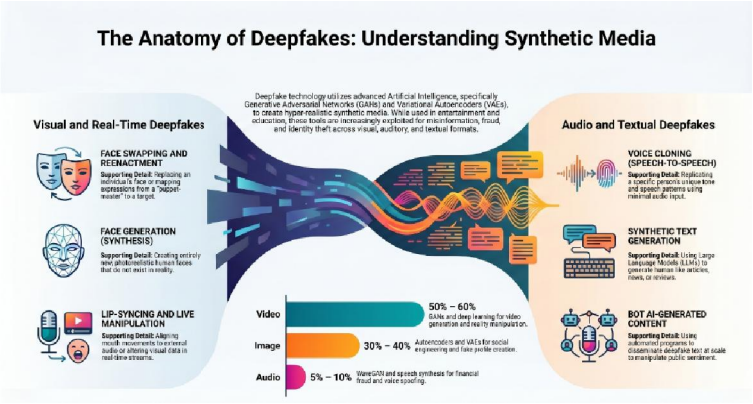
Textual deepfakes uses Large Language Models (LLMs) and transformer architectures to generate highly convincing textual content to influence opinions [17,18]. Natural language processing algorithms can generate human styles writing, this enables automated systems to generate large volumes of synthetic news articles, product reviews, and highly targeted social media content [17,19].

C. Deepfakes vs. Cheapfakes

Evaluating digital threats requires differentiation between complex deepfakes and low-tech cheapfakes [20,21]. Deepfakes uses significant computational resources, extensive training datasets, and advanced deep learning methods to generate synthetic media that did not previously exist [22]. On other hand Cheapfakes require relatively little technical expertise and typically uses basic editing tools to modify, distort, or present authentic content out of context [20,21]. Common approaches include cropping video segments, altering audio recordings, and misrepresenting the location or context of events to influence audience interpretation [20,21].

Due to technical simplicity, cheapfakes are used easily in political campaigns [23]. Experimental studies indicate that such content is also plays equally credible as more sophisticated deepfakes, particularly when it combined with pre-existing political beliefs [23]. Because they consist genuine footage part of its original context, cheapfakes easily bypass detection systems and exploit the human opinions just as effectively as perfect GAN-generated media [24,25].

Picture No 1: - The Anatomy of Deepfakes: Understanding Synthetic Media



IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, June 2026
Impact Factor: 8.2

III. GAPS IN PREVIOUS RESEARCH (LITERATURE GAPS)

A. The Cross-Dataset Generalization Failure

Many researchers build highly complex neural networks that easily achieve over 99% accuracy when tested in controlled lab environments [1]. However, when these models are deployed in real-world scenarios, working with previously unseen deepfake samples, their performance typically declines by approximately 10%–15% [1]. This generalization failure happens because deep learning detectors generates the results on basis of specific dataset they trained on [1]. Instead of learning the universal characteristics of a manipulated video, the software simply memorizes specific compression artifacts or isolated blending boundaries unique to that single dataset [1]. As a result, there remains a lack of lightweight yet robust detection approaches capable of maintaining consistent performance across heterogeneous datasets and evolving manipulation strategies [1].

B. Empirical Limitations in Attitude Measurement

In psychological and political science literature, it is often assumed that highly advanced deepfakes automatically brainwash voters. However experimental data challenging this assumption, the existing literature noticed less systematically compare for psychological impact of high-quality AI-generated deepfakes with low-tech manipulations technologies as cheapfakes [2]. Recent studies indicate that cheapfakes authentic videos simply cut out of original context can equally impact as expensive deepfakes, especially when the fake narrative confirms a voter's pre-existing political beliefs [2]. However, most of the political communication research focuses primarily on short-term reactions immediately following exposure to manipulated media [2]. This creates a significant gap in investigations examining the “illusory truth effect,” whereby repeated exposure to synthetic or misleading content over time may increase credibility across an election cycle [2].

C. Under-Researched Modalities

Existing academic research highlights a strong bias toward the analysis of visual deepfakes, specifically focusing on face-swapping and facial reenactment videos. This concentration led to insufficient attention over other consequential modalities. Text-based deepfakes generated by Large Language Models (LLMs) rarely receive the same forensic attention or analysis, even though these models automate the mass production of highly misleading fake news articles and microtargeted social media posts [3]. Also, the growing risk of audio-based manipulations has received limited attention [3]. During recent Indian elections, political groups successfully used AI translation tools to mass-produce cloned audio messages in various regional languages [4]. This approach bypassed traditional visual detection systems entirely, highlighting the necessity for expanded research into multimodal synthetic media threats [3,4].

D. Lack of Specialized Legal regulations in India

Global legal authorities have concentrated on regulatory responses to deepfake like in United States and European Union [5]. Comparatively gap exists in India's specific legal and constitutional framework addressing synthetic media [6]. Existing research also limited examines the intersection of artificial intelligence and Indian intellectual property law, which rises important questions concerning unauthorized deepfakes of political figures and their implications for individual’s rights [6]. Although Indian courts have begun to acknowledge aspects of personality rights, the applicability of the Copyright Act, 1957 to AI-generated political content remains insufficiently explored [6]. The research points outed India’s legacy cyber law framework does not offer practical, jurisdiction-specific recommendations for institutions such as the Election Commission of India [5]. Current status, including the Information Technology Act, 2000 and the Bharatiya Nyaya Sanhita, still not defined regulations for synthetic media on that level which needed [5].

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, June 2026
Impact Factor: 8.2

IV. DEEPFAKES IN POLITICAL PROPAGANDA: THE INDIAN PERSPECTIVE

A. Social Media Virality and the WhatsApp Ecosystem

Political campaigns in India now heavily uses social media and encrypted messaging platforms like WhatsApp to distribute information to millions of voters. Political parties hire dedicated IT cells to manage these digital networks and push specific narratives directly to voters' phones [1]. When campaigns used deepfakes into this communication ecosystem, the end-to-end encryption of the platforms prevents regulators and election monitors from tracking the original source or stopping the spread [2]. Because the manipulated content circulates privately among trusted family and friends, it easily bypasses traditional fact checking systems [3]. A notable early example occurred during the 2020 Delhi Legislative Assembly elections. Political IT cells distributed manipulated videos of an impactful political leader speaking fluently in multiple regional languages [2]. This strategy increased the reach and virality of the disinformation, as the synthetic videos spread rapidly through closed WhatsApp groups long before authorities responses to correct the record [4].

B. Linguistic Microtargeting and Demographic Divides

India’s language and geographic diversity create a unique opportunity for AI-oriented political manipulation. Political campaigns now uses advanced text-to-speech algorithms and AI translation platforms, such as the Bhashini tool, to instantly translate a politician's speech into various regional languages with perfectly cloning their original voice [3]. This specific capability allows political groups to target rural demographics and linguistic minorities who traditionally not directly linked to mainstream political messaging. Researchers noticed this mass production of localized, emotional messaging as "synthetic intimacy" [3]. While this approach makes politicians more reachable to the common citizen, it sometimes misleads voters. Rural populations and first-time voters, who have low digital literacy, not easily identify these high-quality voice clones as artificial [3]. They generally accept manipulated audio as authentic communication. This allows political groups to exploit demographic diversity and manipulate voting behaviour without facing accountability [3].

C. AI in the 2024 Indian General Elections

The 2024 Indian general elections in India marked a noticeable shift in how synthetic media dominate political strategy. Financial reports indicate that political organizations spent an estimated $50 million specifically on developing AI-generated campaign materials [3]. Major parties deployed over 500,000 individual voice clones to reach voters at low cost instead of traditional physical rallies [3]. Campaigns also used deepfake technology to digitally increase their political figures. Unauthorized deepfakes of famous Bollywood actor’s fake videos also circulated widely online. These videos falsely showing the celebrities criticizing ruling parties and promoting opposition candidates [3]. These aggressive deployments suggests that synthetic propaganda is no longer limited, instead they now act as prominent component in mainstream electoral strategy in India [2,3].

V. EROSION OF PUBLIC TRUST AND EPISTEMIC CERTAINTY

A. The "Liar’s Dividend" in the Indian Context

Deepfake technology creates a significant challenge to the share verified data necessary for the functioning of democratic systems [1,2]. As synthetic media heavily seen on digital platforms, it contributes to a phenomenon known as the "liar’s dividend" which can weaker political accountability [3,4]. This effect arises when public figures exploit increasing public uncertainty about digital content to dismiss authentic video or audio evidence as fabricated or artificially generated [4,5]. In Indian political scenario, this tactic may enable groups to escape legal and social consequences for genuine misconduct [6,7]. Digital manipulation can lead voters to consistently question the authenticity of all media, if voters begin to assume that any video may be fabricated, the evidentiary value of legitimate journalism is correspondingly weakened [4,9].

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 1, June 2026 Impact Factor: 8.2

B. Psychological Impact on Media Trust

Psychological research reveals that deepfakes not always immediately brainwash voters into believing false narratives [10]. Citizens struggle to verify the authenticity of the information they encounter, which can weak their baseline trust in digital journalism [12,13]. Experimental data demonstrates that when individuals experience lack of understand reality, their overall confidence decline in news consumed via social media platforms [14,15]. Repeated seen of synthetic propaganda in media, social platforms, creates the "illusory truth effect," where voters gradually accept false political claims simply due to constant repetition across their digital feeds [5,16]. This continuous unverified, hyper-realistic information manipulates the public mind, it may reduce political engagement among voter groups [17,18].

C. Societal Disruptions and Polarization

The societal impact of synthetic media extends far beyond individual misleading and contributes to broader patterns of structural polarization [19,20]. Political IT cells actively using deepfakes to create false realities that properly align with pre-existing biases of specific demographic groups [2,21]. In India, such synthetic narratives are commonly designed to exploit regional, linguistic, and religious divisions [6,22]. Distributing microtargeted deepfakes through encrypted networks like WhatsApp, malicious groups can create communal tensions and, in extreme cases, contribute to real-world violence [22,23]. Until rural populations get necessary digital literacy to identify synthetic content, their capacity to make right electoral choices is basically compromised [24].

VI. DIGITAL FORENSICS AND DETECTION METHODOLOGIES

A. Current Deep Learning Approaches

Detecting synthetic media requires the use of artificial intelligence-based approaches. Current digital forensics depends on complex neural networks to detect the microscopic manipulations that human eyes naturally miss. Convolutional Neural Networks (CNNs) analyse individual image frames to extract spatial features, including inconsistencies in skin texture, irregular facial boundaries, and unnatural colour blending around facial regions [1,2]. Because videos contain moving elements, researchers frequently combine CNNs with Long Short-Term Memory (LSTM) networks. LSTMs track time-based errors across a continuous sequence of video frames. This enables the detection of temporal inconsistencies such as abnormal blinking patterns, disruptions in facial expressions, and minor lip-sync mismatches [3,4]. In controlled laboratory environments, these hybrid CNN-LSTM models perform very well, achieving up to 97% accuracy in distinguishing real from synthetic facial content [4]. Recently, computer scientists have also introduced Vision Transformers and self-attention mechanisms. These architectures capture global contextual relationships within images, enabling more precise localization of manipulated regions and enhancing interpretability for forensic analysis [5].

B. Traditional Machine Learning vs. Architectural Complexity

A continuous debate in computer science concerns whether the detection of deepfakes requires large-scale, highly complex neural networks. Deep learning approaches often uses high computational resources, memory, and processing time, particularly when applied to high-resolution video data. In comparison, traditional machine learning and classical image processing methods challenge this dependency on deep architectures by achieving same level performance under significantly lower resource constraints [6]. For text-based manipulation detection, simple classifiers such as Support Vector Machines (SVMs) and Random Forests can analyse basic word frequencies and sentence structures to successfully catch AI-generated articles [7,8]. In visual domain, frequency-based image analysis techniques detect spectral irregularities and pixel-level features that are not readily observable in the spatial domain [1]. These traditional algorithms offer a critical operational advantage. Because they require significantly less computing power and energy, they process digital files much faster than deep neural networks. This speed makes traditional machine learning highly practical for real-time content moderation on massive social media platforms [6].

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 1, June 2026 Impact Factor: 8.2

C. Advancing Detection Generalization

A major limitation of current deepfake detection systems is their poor cross-dataset generalization. Multiple existing models uses for overfit to the characteristics of their training data, thereby limiting their effectiveness in real-world scenarios [9]. When programmers train a neural network on controlled benchmarks such as FaceForensics++, the model learn dataset-specific artifacts, including particular compression patterns or recurring blending inconsistencies, rather than generalized forgery cues. When researchers test that identical model on a brand new, unseen deepfake uploaded from the real world, their performance often degrades significantly [9]. This decline in accuracy reflects a reliance on dataset-specific signatures rather than robust, transferable features capable of identifying manipulated media across diverse domains.

This generalization gap presents a significant challenge for electoral integrity. In practice, Political IT cells continuously adapt new manipulation strategies including the use of heavy video compression to facilitate rapid dissemination of deepfakes across mobile messaging platforms such as WhatsApp [10]. Rigid laboratory models fail to detect these degraded, real-world videos, rather than optimizing performance on isolated benchmark datasets, future detection systems should prioritize robustness and cross-dataset generalization. Future digital defences must prioritize cross-dataset generalization, building lightweight, adaptable software capable of catching completely unseen deepfake tactics in real-time before they go viral [11,12].

VII. LEGISLATIVE FRAMEWORKS: GLOBAL BEST PRACTICES VS. INDIAN GAPS

A. The Global Regulatory Landscape

International regulatory strategies approached a sharp proactive, unified frameworks and fragmented, reactive rules. The European Union uses a structured approach through its Artificial Intelligence Act (AI Act) and Digital Services Act, which categorize AI systems by risk and explicitly mandate transparency and clear labelling for all synthetic media to ensure accountability [1,2]. United States uses highly decentralized system, rather than enacting a unified federal ban, individual U.S. states such as Texas, California, and Minnesota have passed specific legislation prohibiting the distribution of deceptive deepfakes immediately preceding an election [1,3]. This state-level approach prevents immediate electoral manipulation but leaves substantial regulatory loopholes at the national level, even though proposed federal bills like the NO FAKES Act and the DEFIANCE Act [1]. United Kingdom's Online Safety Act 2023 criminalizes non-consensual explicit deepfakes but leaves political misinformation to legacy defamation laws, while China enforces strict regulations mandating the explicit labelling of all AI-generated content [1,2].

B. Statutory Deficits in India

India currently lacks a dedicated legal framework specifically for counter deepfake technologies. Law enforcement and election officials still depends on legacy system and rules in emergence of modern deep learning systems to address AI-driven manipulation [1,4]. The Information Technology (IT) Act of 2000 serves as the primary legal tool, specifically utilizing Section 66D for identity theft through computer resources and Section 67 for publishing obscene content [5,6]. However, this framework still unable to deal with the speed, scale, and technical complexity of against algorithmic content manipulation, limiting its effectiveness in addressing deepfake-related harms.

The newly implemented Bharatiya Nyaya Sanhita (BNS) of 2023 contains provisions against forgery, criminal defamation, and cheating by impersonation, but it lacks the technical terminology required to impose a penalty on synthetic media creators directly [1,6]. The Representation of the People Act of 1951 addresses corrupt electoral practices and the spreading of mutual opposition among classes, yet it does not explicitly account for AI-generated propaganda or digital voter suppression [6,7]. The Digital Personal Data Protection (DPDP) Act of 2023 focuses on data privacy and consent but requires clearer guidelines regarding synthetic media [4,8]. The application of the Copyright Act of 1957 presents additional legal challenges, as Indian courts continue to solving with the extent to which traditional copyright principles can address unauthorized AI-generated voice replicas and the misuse of an individual's likeness, particularly in the context of public and political figures [6,7].

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 1, June 2026 Impact Factor: 8.2

C. Balancing Regulation with Freedom of Expression

Drafting new laws to address synthetic media requires careful balance between mitigating digital harmful AI-generated content and preserving the right to freedom of speech and expression guaranteed under Article 19(1)(a) of the Constitution of India [8]. Consequently, effective regulatory frameworks must be narrowly tailored to address demonstrable harms while ensuring adequate protections for lawful and constitutionally protected speech [1,5].

Effective legal frameworks should focus on the malicious use of synthetic media intended to influence voters or manipulate democratic processes, while avoiding undue restrictions on clearly identified synthetic content created for entertainment, education, satire, or social commentary [1]. A transparency-based mechanisms such as mandatory disclosure and labelling requirements for AI-generated political content have emerged as a promising regulatory approach. Such measures can enhance democratic accountability and informed decision-making while preserving the fundamental right to freedom of expression [1,2].

Table No 1: Legal and Regulatory Landscape of Deepfakes in Indian Elections

Legal Instrument or Framework	Key Provisions and Sections	Target Misuse or Harm	Regulatory Body or Entity	Proposed Reforms and Recommendations	Source
Information Technology Act, 2000 (IT Act)	Section 43A (negligence), Section 66D (cheating by personation), Section 66E (privacy violation), Section 66F (cyber terrorism), Section 67 (obscene material), Section 69A (blocking content), Section 72 (breach of privacy), and Section 79 (intermediary liability).	Identity theft, cheating by personation, violations of privacy via fabricated imagery, dissemination of obscene content, non-consensual deepfake pornography, threats to national security, and AI-generated misinformation.	Ministry of Electronics and Information Technology (MeitY); Government of India	Updating laws to explicitly cover AI-generated disinformation; drafting a dedicated Deepfake (Political Integrity) Bill; establishing Digital Election Courts; amending Section 79 to make safe-harbour conditional on a 60-minute service-level agreement for removal.	1, 2, 4-6, 10-12, Inferred
Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 (IT Rules 2021)	Rule 3(1)(b) (prohibited content), Rule 3(2) (removal of deepfakes within 24 hours), and Rule 4(2) (identification of first originator).	Spread of misinformation and deepfakes via social media intermediaries; lack of accountability for content originators.	Ministry of Electronics and Information Technology (MeitY); Grievance Appellate Committees	Directing intermediaries to proactively identify and remove misinformation; mandatory alignment of platform terms with prohibited content guidelines.	1
Representation of the People Act, 1951 (RPA) / Electoral Regulations	Section 123 (corrupt practices), Section 123(4) (false statements about candidates), Section 125 (promoting enmity), and Section 126 (48-hour campaign silence period).	Electoral manipulation, inciting societal divisions, influencing voter decisions through unauthorised media, discrediting political	Election Commission of India (ECI)	Empowering the ECI with capacity to crackdown on digital disinformation; updating the Model Code of Conduct for AI content; amending Section 123 to prohibit	1-7, 10-14, Inferred

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 1, June 2026 Impact Factor: 8.2

		candidates, and use of voice clones to sway voters.		synthetic endorsements; establishing a Digital Integrity Task Force.	
Bharatiya Nyaya Sanhita (BNS) / Indian Penal Code	Sections 73, 499 & 500 (defamation), Sections 78, 354A & 354D (harassment/modesty), Sections 83 & 145 (forgery), Sections 195(1), 292-293, 316(1), 336, and 420.	Criminal defamation, incitement against public order, virtual forgery, sale of obscene material, character assassination, financial fraud, and harassment through morphed content.	Law Enforcement Agencies; Indian Judiciary	Integration of provisions to address deepfake-related cybercrimes; establishment of special cybercrime divisions; fast-track tribunals; 5-10 years imprisonment for non-consensual deepfake pornography.	2, 6-8, 10-12, Inferred
Digital Personal Data Protection Act, 2023 (DPDP Act)	Section 4 (consent), Section 8 (prevention of likeness misuse), Section 22 (penalties), and Article 66 (prohibition on false personal data).	Unauthorised use of biometric data, face-swapping, voice cloning, privacy infringements, and harm to individuals through fabricated personal data.	Data Protection Board of India; Government of India	Explicit inclusion of AI-generated data misuse; strengthening digital rights and identity protection; development of robust enforcement mechanisms for synthetic media.	1, 4, 7, 10, 12, Inferred
Proposed Digital India Deepfake Bill / Deepfake Prohibition Act	Not in source (Draft stage)	Malicious use of AI for content manipulation; fraudulent or libelous dissemination; gaps in existing definitions of AI-generated media.	Proposed Deepfake Regulatory Authority; Ministry of Electronics and Information Technology (MeitY)	Criminalising bad-faith AI use; providing legal recourse for victims; mandatory visible disclaimers and machine-readable watermarking; strict liability for social media platforms.	1-5
Copyright Act, 1957 and Trade Marks Act, 1999	Section 51 (penalties for unauthorised copies); remedies for unauthorised use of proprietary symbols.	Unapproved use of copyrighted material to create deepfakes; intellectual property infringement; unauthorised use of campaign songs, logos, and slogans.	Registrar of Copyrights; Intellectual Property Office; Judiciary	Applying copyright law to provide legal grounds against unauthorised use of likeness; fast-track injunctive relief; amendments to address AI-generated misuse of symbols.	1, 6-8
Watermarking and Technical Regulatory Framework	Inclusion of metadata (creator, time, location); Blockchain-based image authentication.	Lack of provenance in digital media; inability of users to distinguish authentic	Ministry of Electronics and Information	Mandatory watermarking for authentication of sensitive media;	1, 5, 9, 10

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 1, June 2026 Impact Factor: 8.2

		from manipulated content; untraceable synthetic content.	Technology (MeitY); Industry Stakeholders	standardisation of blockchain authentication techniques; embedding immutable metadata to ensure traceability.	
General Election Guidelines / Voluntary Code of Ethics	Pre-certification of political advertisements; responsible use of social media.	Electoral manipulation, misinformation, and inflammatory content during election periods.	Election Commission of India (ECI)	Granting binding force to guidelines; technological empowerment with AI-based detection tools; real-time takedown mechanisms.	6, 11, 12

VIII. MULTI-DIMENSIONAL COUNTERMEASURES FOR INDIA

A. Strengthening ECI Oversight and Legal Reform

India requires a specialized legal framework, specifically a dedicated Deepfake Prohibition Act, to directly criminalize the malicious use of synthetic media in elections [1,2]. Parliament must amend Section 123 of the Representation of the People Act of 1951 to explicitly classify unauthorized AI-generated political endorsements and voice-cloned use as corrupt electoral Practices [3,4]. The Election Commission of India (ECI) currently lacks the technical infrastructure to enforce digital rules. The ECI must establish a dedicated Digital Integrity Task Force equipped with multi language AI-monitoring tools to receive real-time alerts when deceptive media circulates across regional networks [3,4]. Social media platforms must follow strict legal liability. If a technology company fails to remove flagged deepfake propaganda content within 24 hours, the government must impose required financial penalties on the platform to enforce accountability [2,4].'

B. Cryptographic Authentication and Watermarking

To increase public trust in digital content it requires mechanisms that enable the verification of authenticity before information reaches end users. Blockchain technology provides a highly secure, decentralized Proof of Authenticity (PoA) system to establish a verifiable history for digital files [5]. Integrating technologies such as Ethereum smart contracts and the InterPlanetary File System (IPFS), journalists and content creators can register digital assets on tamper-resistant register, creating auditable records of origin and modification history, because alterations to a file typically result in changes to its cryptographic hash, unauthorized modifications can be more readily detected and verified to the public [5]. Lawmakers must mandate digital watermarking for all commercially generated AI content [4,6]. Watermarking techniques embed machine-readable metadata within digital content, enabling information about its origin, generation process, or subsequent modifications to be traced and authenticated [7,8]. Such approaches can enhance transparency, support forensic investigations, and assist users in assessing the credibility of digital media [4].

C. Cultivating Digital Media Literacy

Algorithms and laws alone cannot fully protect a voter. Citizens must develop strong understanding against digital manipulations [6,9]. India must launch a National Digital Literacy Mission specifically targeting rural populations, first-time voters, and linguistic minorities who are primary target for localized deepfake propaganda [3,4]. Educational institutions must integrate media ethics and deepfake identification training directly into secondary school syllabus to build awareness from an early age [3,4]. Independent fact-checking organizations must collaborate with social media platforms to conduct public awareness campaigns, teaching users how to identify suspicious videos before sharing them [7,8].

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, June 2026
Impact Factor: 8.2

IX. CONCLUSION

A. Summary of Findings

Generative artificial intelligence has totally changed the way of modern political communication. Deepfake technology are used for creating highly realistic synthetic media that can be used to political content and propaganda [1]. These synthetic media files bypass natural human identification and can undermine public confidence in digital information and challenge the trust over democratic electoral processes [2]. The Indian political ecosystem noticed vulnerability of synthetic contents due the highly geographical and language diversity [3]. 2024 general elections in INDIA, political communication teams used AI tools for voice-cloning systems and regional AI translation tools, including platforms such as Bhashini, to enhance outreach strategies [4]. Political campaigns spent an estimated $50 million on these tools to address rural and low-literacy voters through encrypted messaging networks like WhatsApp [5]. Repeated exposure to synthetic kind of manipulated content can lead to mental overload and creates doubt about what information can actually be trusted, this psychological confusion known as the "liar's dividend," where even genuine audio or video evidence of misconduct can be dismissed as artificial or fabricated [7].

Current defensive models still faces some challenges, from a technological perspective, deep learning-based deepfake detection systems struggle with cross-dataset generalization [8]. In some cases, these models tend to overfit to the specific datasets used during training, which limits their effectiveness when deployed in real-world conditions where manipulation techniques continuously are evolving [9]. On the legal point, India still depends on relatively old rules and regulatory frameworks likes, the Information Technology Act of 2000 and the Bharatiya Nyaya Sanhita do not yet provide sufficiently precise technical definitions to exclusively address synthetic media manipulation or to impose clear accountability on social media platforms for the spread of such content [11,12].

B. Final Verdict

Securing the future of Indian electoral integrity requires an immediate, multi-dimensional defence strategy. The Indian government must introduce a dedicated Deepfake Prohibition Act [13]. This specialized legislation must mandate strict digital watermarking and algorithmic transparency while actively protecting constitutional free speech [14]. Digital forensic research must immediately shift toward prioritizing lightweight, generalizable detection models capable of real-time analysis on edge devices [16].

Finally, technological and legal barriers alone cannot fully protect a voter or public without an awareness in electorate about synthetic media [17]. Nationwide digital media literacy programs are must to teach citizens how to evaluate synthetic propaganda [18]. With this integrated approach: Adaptive computational forensics, Precise cyber legislation, and Sustained public education can modern democracies tackle the threat of artificial intelligence generated manipulation [19].

REFERENCES

1. Raza, A. Basit, A. Amin, Z. A. Arfeen, M. I. Masud, U. Fayyaz, and T. A. Jumani, "A Comprehensive Review of Deepfake Detection Techniques: From Traditional Machine Learning to Advanced Deep Learning Architectures," 2026.
2. R. Mubarak, T. Alsboui, O. Alshaikh, et al., "A Survey on the Detection and Impacts of Deepfakes in Visual, Audio, and Textual Formats," *IEEE Access*, 2023.
3. Yaohui Wang and Antitza Dantcheva Inria, "A video is worth more than 1000 lies: Comparing 3DCNN approaches for detecting deepfakes," FaceForensics++ Dataset Analysis, 2019.
4. Dr. S. M. Raza, "AI-Generated Misinformation and Deepfakes: Legal Challenges and Regulatory Responses in India," *The Infinite: An International Peer Reviewed Journal of Multidisciplinary Research*, vol. 3, no. 4, Apr. 2026.
5. Riya Chugh, "Artificial Intelligence, Deepfakes, and Electoral Integrity in India: Legal and Intellectual Property Challenges". Panjab University Law Magazine (MAGLAW), Volume IV Issue II, pp 69-83

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 1, June 2026 Impact Factor: 8.2

6. M. Hameleers, "Cheap Versus Deep Manipulation: The Effects of Cheapfakes Versus Deepfakes in a Political Setting," *International Journal of Public Opinion Research*, vol. 36, no. 1, 2024.
7. H. R. Hasan and K. Salah, "Combating Deepfake Videos Using Blockchain and Smart Contracts," *IEEE Access*, 2019.
8. F. Folorunsho and B. F. Boamah, "Deepfake Technology and its Impact: Ethical Considerations, Societal Disruptions, and Security Threats in AI-Generated Media," *International Journal of Information Technology and Management Information Systems*, vol. 16, no. 1, Jan. 2025.
9. "Deepfakes and Electoral Integrity: Legal Gaps in India and Global Best Practices (1)," *Journal of Legal Research and Juridical Sciences*, vol. 4, no. 4, 2025.
10. "Deepfakes and Electoral Integrity: Legal Gaps in India and Global Best Practices," *Journal of Legal Research and Juridical Sciences*, vol. 4, no. 4, 2025.
11. "Deep Fake in Indian Elections," Executive Summary and Case Background.
12. V. M. Patel and S. Degadwala, "Deepfake Detection Using Convolutional Neural Networks and LSTM Modelling," *International Journal of Scientific Research in Science and Technology*, vol. 12, no. 3, May 2025.
13. M. S. Rana, M. N. Nobi, B. Murali, and A. H. Sung, "Deepfake Detection: A Systematic Literature Review," *IEEE Access*, 2022.
14. S. Bhale, "Deepfake Laws in India: The Need for Legal Regulation in the AI Era", Manikchand Pahade Law College
15. R. A. Rahman and R. Anggriawan, "Deepfake and Election Crimes: Comparative Perspectives from Indonesia, India, Pakistan, and the U.S.," *Indonesian Comparative Law Review*, vol. 7, no. 2, 2025.
16. A. Heidari, N. J. Navimipour, H. Dag, and M. Unal, "Deepfake detection using deep learning methods: A systematic and comprehensive review," *WIREs Data Mining and Knowledge Discovery*, vol. 14, no. 2, 2024.
17. A. Vaccari and A. Chadwick, "Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News," *Social Media + Society*, vol. 6, no. 1, 2020.
18. "Deepfakes and the Crisis of Trust: Public Perception of Media Authenticity in the Age of Synthetic Content," *Nigerian Journal for Technical Education*, vol. 24, no. 2, 2025.
19. D. Kumar, "Deepfakes, Free Speech, and the Right to Truth: A Comparative Legal Study on Regulating Synthetic Media in the USA, UK, and India," *Advanced International Journal for Research*, 2024.
20. J. Hu, X. Liao, et al., "Detecting Compressed Deepfake Videos in Social Networks Using Frame-Temporality Two-Stream Convolutional Network," *IEEE Transactions on Circuits and Systems for Video Technology*, 2023.
21. G. P. Zachary, "Digital Manipulation and the Future of Electoral Democracy in the U.S.," *IEEE Technology and Society Magazine*, 2020.
22. T. Dobber, D. Trilling, et al., "Do (Microtargeted) Deepfakes Have Real Effects on Political Attitudes?," *The International Journal of Press/Politics*, vol. 26, no. 1, 2021.
23. R. Sunil, P. Mer, A. Diwan, R. Mahadeva, and A. Sharma, "Exploring autonomous methods for deepfake detection: A detailed survey on techniques and evaluation," *Heliyon*, vol. 11, Jan. 2025.
24. Shraddha Suratkar, Mukul S. "Exposing DeepFakes Using Convolutional Neural Networks and Transfer Learning Approaches".
25. R. Babaei, S. Cheng, R. Duan, and S. Zhao, "Generative Artificial Intelligence and the Evolving Challenge of Deepfake Detection: A Systematic Analysis," *Journal of Sensor and Actuator Networks*, vol. 14, 2025.
26. S. H. Al-Khazraji, H. H. Saleh, A. I. Khalil, and I. A. Mishkal, "Impact of Deepfake Technology on Social Media: Detection, Misinformation and Societal Implications," *The Eurasia Proceedings of Science, Technology, Engineering & Mathematics*, vol. 23, 2023.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, June 2026
Impact Factor: 8.2

27. X. H. Nguyen, T. S. Tran, V. T. Le, K. D. Nguyen, and D.-T. Truong, "Learning Spatio-temporal features to detect manipulated facial videos created by the Deepfake techniques," *Forensic Science International: Digital Investigation*, vol. 36, 2021.
28. Xiao Han, W. Wang, Macau University of Science and Technology, "Low Resolution Facial Manipulation Detection," Artifacts-Focus Super-Resolution framework.
29. Junet Setiawan, Rachmat Irvan, "Regulatory Strategies for the Prevention and Legal Enforcement Against the Misuse of Deepfake Technology in Elections," 2024.
30. "Rising Menace of Deepfakes with the Help of AI: Legal Implications in India," *Indian Journal of Integrated Research in Law*, vol. 4, no. 3, 2024.
31. "Regulating Deepfakes to Protect Indian Elections," *Partners Universal Innovative Research Publication (PUIRP)*, vol. 1, no. 2, Nov. 2023.
32. U. Aneja, A. Chamuah, and A. Reddy K, "The Impact of Artificial Intelligence on Elections," *International Journal for Multidisciplinary Research*, May 2020.
33. S. Li, "The Social Harms of AI-Generated Fake News: Addressing Deepfake and AI Political Manipulation," *Digital Society & Virtual Governance*, vol. 1, no. 1, Feb. 2025.
34. Markus Appel , Fabian Prietzel, "The detection of political deepfakes," Open Science Framework Analysis.
35. "The ethical challenges posed by deepfake technology during the 2024 Indian general elections," Executive Summary.