VNU Journal of Social Sciences and Humanities
Vol 2, No 2 (2026) 225-239
© The Author(s) 2026

p-ISSN: 3093-3889
e-ISSN: 3093-3900



Deepfakes and the Crisis of Credibility: AI, Disinformation, and the Public Discourse in India

Mohammad Reyaz [*], Kaifia Ancer Laskar [**]

Abstract: Technological advances have blurred the lines between what is considered to be real and imaginary, natural and synthetic. In the post-truth era, reality itself has seemingly become contested and developments in generative artificial intelligence (AI) have changed both the perceptions of people and their media consumption patterns. The "fake news" ecosystem has created an AI-supported synthetic reality that fuels propaganda and untruths. This article explores the proliferation of "deepfakes" and AI-generated synthetic media in India's digital space, taking five interrelated themes of political misrepresentation, political rumour, false attribution, manipulation of current affairs, and gendered digital violence. We foreground how AI disinformation impacts the everyday news cycle, politics, and geopolitics in the digital age, typically exploiting audience response and platform logics to amplify reach. The article argues that these AI-generated synthetic deepfakes pose as epistemic disruptions, paving the way for easier deception, and leading to the polarization and humiliation of individuals and communities. The article concludes by drawing attention to critical aspects of the regulatory, technological, and pedagogical response to the emerging crisis of credibility in digital media ecosystems.

Keywords: deepfakes; AI misinformation; post-truth; affective publics; propaganda; gender-based violence; digital media; India.

Received: 11th May, 2026; Revised: 13th July, 2026; Accepted: 17th September, 2026

DOI: https://doi.org/10.33100/jossh.2026.2.2.3

1. Introduction

The swift diffusion of artificial intelligence (AI) into routine media consumption has transformed the way information is collated, distributed, and accessed. One of its most disruptive manifestations is the rise of "deepfakes" and AI-generated content. These technologies enable the production of hyper-realistic, but fabricated, audio-visual content, which is often difficult to distinguish from real (i.e., non-AI generated) recordings. In India, one of the world's most populous, and both politically and digitally active societies, deepfakes have gone from novelty to a weaponised tool for misquoting politicians and celebrities, fabricating journalistic voices, sexualizing women, and building persuasive, but fake, stories about present-

[*] Department of Journalism and Mass Communication, Aliah University, India; email: reyaz@aliah.ac.in.
[**] Department of Journalism and Mass Communication, Aliah University, India.

day geopolitical crises. Its impacts are not limited to misinformation as they shape and influence public opinion, erode institutional trust, and, in some instances, have serious personal consequences.

The rise and wide use of social media platforms such as X (formerly Twitter), Facebook, and WhatsApp have changed how we communicate in the digital age. These platforms allow users to share information instantly across large networks, overcoming geographical and time limitations. However, along with these benefits, social media has also made it easier for fake news and misinformation to spread quickly (Aïmeur et al. 2023; Naeem et al. 2020; Shu et al. 2017). The lack of strict editorial oversight and rapid pace of information-sharing allow false content to reach many people in mere seconds, often before anyone can check its accuracy. Research from the Massachusetts Institute of Technology showed that false news spreads much faster and more widely than accurate news on social media (Vosoughi et al. 2018). This behavior shows how human tendencies help misinformation thrive online. The effects of misinformation are extensive and can have serious consequences for society. For example, misinformation significantly influenced the 2016 U.S. presidential election, where false stories circulated widely on social media (Allcott and Gentzkow 2017). Similarly, during the COVID-19 pandemic, the spread of false information about treatments and safety measures put public health at risk (World Health Organization 2020; Laskar and Reyaz 2021). These instances show how misinformation can sway public opinion, erode trust in institutions, and cause genuine harm. AI has only hastened this process, and made deepfakes more accessible.

This article uses as case studies recent credible examples in five interrelated realms – the misquoting of politicians and famous personalities, false attributions to media and journalists, manipulation of both current affairs and political rumours, and gendered violence – to identify and analyze a coherent and growing ecosystem of AI-enabled disinformation. Consequently, in this article, we study the phenomenon of disinformation using AI or synthetic media content that thrives on the attention economy. Building upon several case studies, the article argues that the disinformation ecosystem can further amplify an already polarized, post-truth society. The so called affective economy thriving in these intentional digital disruptions exploits pre-existing biases among social media users and shapes the emotions in ways that benefit the disinformation infrastructure.

2. Literature Review

Scholarship on misinformation and digital manipulation has grown in the past decade, but the rise of deepfakes represents a recent development whose relevance can only be partly predicted. The early work on misinformation conceptualizes "information disorder" as a spectrum composed of misinformation, disinformation, and malinformation (Wardle and Derakhshan 2017). The spread of deepfakes has further complicated the disinformation ecosystem by blurring the lines between two different kinds of narratives – fabricated and plausible.

The issue of post-truth politics further shapes the broader intellectual underpinning. McIntyre (2018) and Keyes (2004) contend that contemporary political communication is privileging belief over evidence. Deepfakes exacerbate this condition as they shake one of our most trusted forms of evidence: visual media. The epistemic authority of images and videos that has previously been viewed as relatively

trustworthy is therefore now open to suspicion. The works of Bennett and Segerberg (2012) and Papacharissi (2015) are critical in our understanding of why such contents spread persuasively. Their conceptualization of *connective action* and *affective publics* demonstrate how digital engagement is mediated not so much by institutional frameworks as by highly personal, emotionally loaded engagement. According to Laskar and Reyaz (2021: 93), "personal bias as well as unverified content sharing through the click of a button has not only led to the epidemic of fake content across the world, but in countries like India, it has also led to lynching and violence in various places".

Deepfakes and AI-generated content are especially well-suited in this space because they can be created to provide quick emotional reactions and to look genuine. Recent research has also begun to explore the deepfake phenomenon in the Indian context (Verma 2026). Studies indicate a sudden surge in the application of AI-generated media for political manipulations, financial fraud, and personal attacks (Raj and Yadav 2025; Jain 2025). These changes are also happening amid a rapid growth of digital penetration that offers new opportunities and new vulnerabilities. According to one study, ease of access to deepfake tools has decreased the minimum entry requirements, allowing people with relatively basic technical skills to produce convincing manipulated content (Sharma and Shaikh 2025). Indian legal scholarship has also pointed to several gaps in regulatory frameworks – for example, the need for technological solutions to preserve privacy and to develop AI-driven detection strategies (Kishwar et al. 2025). While extant laws already exist to deal with defamation and identity theft, they are rarely used against generated and propagated synthetic media, except in certain politically motivated cases (Jha and Jain 2021; Srivastava 2026). At the same time, emerging research supports the potential societal implications of deepfakes on elections, communal polarization, and digital crime (Ghosh and De Sarkar 2024; Roy et al. 2025). Indian studies have also focussed on the technological and linguistic aspects of deepfakes. The rise of datasets specifically in Indian languages catering to indigenous demographics emphasize the seriousness of the issue and the need for context-specific solutions (Paul and Dutta 2025; Kaur et al. 2024).

Several studies have foregrounded the use of generative AI in propagating deepfakes to cater to the attention economy (Kishwar et al. 2025; Verma 2026; Ohman 2022). Ohman (2022: 318), for example, notes that synthetic media technology opens floodgates to new forms of "cyber-bullying, sexual harassment and humiliation". Forms of humiliation-inducing synthetic content like pornographic or nude content to hate speech have appeared on digital media platforms (Ohman 2019; Young 2021). Intentional synthetic content intended to produce harm to an individual may also be disruptive to the family and life of the individual, but the harm targeted towards groups or a community related to political ecosystems is considered more serious (Ohman 2022). These synthetic deepfake contents are of an epistemic nature and disrupt the entire authenticity of the medium (Ohman 2022). Warzel (2018) argues that AI generated content can create an environment where users cannot make out the natural from the synthetic. This leads to deception that is more sophisticated, tougher to detect, and works with other technological forces "that are not only currently unknown but likely unpredictable". He calls this post-truth era, under the onslaught of synthetic content, the "information apocalypse" (Warzel 2018).

These studies thus highlight how deepfakes have made deep inroads in India as in other parts of the world. Further, because of easily accessible technologies, synthetic media can easily be produced in regional languages as well as English, affecting political circumstances and even electoral behavior. While generative AI has received much academic attention foregrounding it as a transformative tool to ease pedagogy or machine learning, its potential to manipulate political communications and trust in media sources, particularly in the context of India, has received less attention. This article hopes to fill this gap.

3. Theoretical Framework

3.1. Post-Truth and the Crisis of Evidence

AI and synthetic media intensify the impact of deepfakes by fundamentally destabilizing the evidentiary value of audio-visual media. Visual content was traditionally considered as evidence. But with the advent of AI-generated synthetic media, even highly realistic videos cannot be immediately regarded as reliable, accurate or valid. This loss of confidence feeds into what Wardle and Derakhshan (2017) identify as "information disorder," wherein misinformation, disinformation and malinformation live and support one another. Deepfakes serve as part of all three categories. Not only do they fake events and cause doubt in the face of credible evidence, they also generate what many academics have labelled an epistemic crisis (Okocha 2026; Naffi 2025, Fallis 2021). The contemporary media environment is a product of the "post-truth" era, in which objective facts are less important than emotional (or pre-formed) appeals and beliefs (McIntyre 2018; Keyes 2004). The information overload on the internet has created a space for a new kind of economy, often known as the "economy of attention" (Baeza-Yates and Fayyad 2022). Developments in society and public opinion depend heavily on digital technologies. Baeza-Yates and Fayyad have argued that competition for people's attention has increased exponentially as it proportionately boosts the growth of digital advertising revenue (2022: 125). Intense competition in the attention economy creates space for manipulation and disruption of digital information in favour of the select few who own the data and information. The data economy creates biases and privacy concerns, where those who possess the maximum personal data will dominate the society by the virtue of digital manipulation (Baeza-Yates and Fayyad 2022: 131).

In an age when social media has the power to create filter bubbles and echo chambers, digital media users are susceptible to receiving a one-dimensional image of the world (Cerro 2024). These filter bubbles and online echo chambers often shape public opinion by focusing on identity and polarization (Pariser 2011). Algorithms are used to expose users based on their previous social media history, operating differently for different users following the mechanism of selective exposure and personalization (Bozdag 2013; Iyengar and Hahn 2009). The selective exposure technology of filter bubbles is problematic as it can reinforce or enhance polarization (Flaxman et al. 2016). Similarly, echo chambers are defined as the phenomenon of selective self-exposure and selective avoidance of certain information and people (Tucker et al. 2018). Further, both echo chambers and filter bubbles "involve the selective exposure to information that reinforces existing beliefs, they operate in slightly different ways and have distinct implications for political polarization and information consumption" (Cerro 2024: 3). The *confirmation bias* of

individual "prosumers" – that is, the tendency for individuals to seek out and interpret only information that affirms pre-existing personal beliefs rather than contradictory opinion – contributes to the phenomenon (Nickerson 1998; Kahneman 2011). Users also face confirmation bias or disconfirmation bias, where they go through either selective exposure or selective avoidance (Festinger 1957), which helps to explain the emergence of polarization (Cerro 2024: 3). Confirmation bias plays a crucial role in misinformation as individuals are more likely to consume, like, and share content that aligns with their views (Nickerson 1998; Kahneman 2011).

3.2. Affective Publics and Emotional Circulation

The rapid proliferation of deepfakes is directly associated with the emergence of what Zizi Papacharissi (2015) refers to as "affective publics" – that is, social networks formed on the basis of shared emotions rather than rational thought processes. Social media algorithmic dynamics push emotive posts and amplify outrage, fear, or anger. Deepfakes work particularly well in such ecosystems as they couple emotional triggers with AI-generated visual realism. The "connective action theory" of Bennett and Segerberg (2012) better explains this phenomenon of digital participation than any traditional systematic media consumption framework: it is more personalised and driven by emotions. Consequently, media consumption and circulation dynamics are determined by emotional resonance rather than any factual basis; that is, members of the public share certain posts or information not because they are true, but because they appear to be real.

3.3. Propaganda Models and Platform Logics

The traditional model of propaganda in media narratives, as explained by Herman and Chomsky (1988), relied on structural filters like sourcing, advertising, and above all ownership. Although these filters are still relevant, AI-generated misinformation points to a new type of propaganda model where digital platforms and their algorithmic-based filter bubbles introduce new dynamics along with more decentralized production and consumption, and are networked and amplified by algorithms (Benkler et al. 2018). Deepfakes are perfect for such a model as they are cheap to produce and the same content can be replicated in diverse languages, and scaled across different platforms. Further, they can be easily adapted to varying ideological contexts. Algorithms are not neutral intermediaries as they play an active role in determining the visibility and attention garnered by certain posts or contents (Gillespie 2018). Such a model is more likely to prioritise deepfake content because it is emotionally triggering, sensational, and engaging. This makes deepfake and synthetic videos a powerful tool for propaganda.

3.4. The Public Sphere, Public Legitimacy, and Democratic Decay

The sudden proliferation of AI-generated synthetic media in a country like India, where digital literacy remains very low, has serious implications for the public sphere, understood as a space for rational debate and discussion (Habermas 1989). In deliberative democracy, the public sphere plays a critical role in the free flow of information, but the spread of synthetic media disinformation undermines the relevance of these spaces. This is because citizens will find it increasingly difficult to distinguish between real and fake content, a capacity further

eroded by advances in technology. This development has implications for what future deliberative democracy demands: when citizens are unable to distinguish between genuine and counterfeit content, the conditions are ripe for democratic decline. Couldry and Hepp (2017: 10) posit that "technologically based media of communication are now fundamental to the construction of everyday reality," thereby emphasising the role of mediatisation. In such a scenario, deepfakes add a dangerous layer by creating synthetic media content that is difficult to distinguish, particularly for ordinary citizens.

3.5. Methodology

We employ a qualitative case study method to explore the proliferation of deepfakes and AI-based synthetic media in the misinformation ecosystem of India. We use a curated series of instances that have been verified by fact-checking platforms, such as Alt News, Factly, Boom Live, India Today Fact Check, and Quint Webqoof, along with historically more reliable media outlets such as the *Times of India* and NDTV.com. For convenience and greater diversity, we have chosen several cases in each of the five categories of political misquotation, media misattribution, manipulation of current affairs, large-scale political rumour, and gendered targeting. We employ qualitative analysis to identify the trajectories of deepfakes in India and to highlight how they are created, consumed, and circulated. The analysis takes into account the following three facets: the nature of manipulation, the context of circulation, and its potential effect on audiences. Such an analysis will help us to understand the nature of the AI-driven misinformation ecosystem, and how it affects the everyday reality of prosumers in this highly mediatised society.

4. Case Studies

4.1. Misquoting Politicians and Celebrities: Manufacturing Consent and Confusion

False attribution of statements to politicians is one of the most widely employed applications of deepfake technology in India, as is the use of artificial material. These doctored videos, purported screenshots of social media posts, and increasingly now synthetic media, created using generative AI, are usually designed to warp stances on sensitive matters like foreign policy, national security, or communal relations. For example, a video shared on the internet falsely showed Indian politician Shashi Tharoor lauding Pakistan's diplomatic moves during a cricket dispute. Tharoor dismissed the video, claiming that the voice as well as the language did not match his, and that he had never made the statements (Fact Check Bureau 2026; TOI News Desk 2026). In another video, later credibly proven to be digitally manipulated (TOI News Desk, 2025), Minister for External Affairs (MEA) S. Jaishankar was seen as apologizing to the people of India for purposedly failing to teach Pakistan lessons. Similarly, a deepfake video of an MEA spokesperson allegedly "threatening" Iran went viral during the Iranian crisis (Alam 2026). Narendra Modi, the Prime Minister of India, has also been the victim of such manipulation. One such deepfake video widely circulated recently claimed that he had announced aid to Afghanistan during his address to the Israel's parliament, just days before the Israel-Iran conflict. Factcheckers noted that the video was manipulated and magnified by cross-border networks (Dutta 2026).

These cases exemplify the ways in which AI is being used not only to influence and manipulate domestic audiences but also to

influence geopolitical narratives. Such misquotes also function by leveraging the confidence provided by familiar public figures. When a well-known figure has the appearance of a leader giving a speech, or talking to media, audiences are less likely to suspect its authenticity.

Similarly, in 2024, in the run up to the Indian general elections, a short video, purported to be of Bollywood star Amitabh Bachchan being very critical of government policies, went viral. The clip appeared very realistic, leading many to believe it, until corroboration revealed that the audio had been AI-generated (Mahalanobish 2024). Bachchan's case is another example of how the voices of famous personalities, not just politicians, can be appropriated to give legitimacy to any narrative for propaganda purposes. There have been similar viral videos and audio clips of other celebrities, including actor Akshay Kumar and former cricketer Gautam Gambhir (Deshpande 2025; TOI Legal Desk 2026).

These examples show how deepfakes prey on familiarity and authority. The post-truth state is increasingly defined by a dissociation between credibility and verifiability in the digital age (McIntyre 2018). In this scenario, the visual appearance of politicians or popular personalities is enough to lend credence to content even where it is entirely fictitious.

4.2. Political Rumours and Synthetic Narratives

AI-generated media is increasingly used to manufacture wide-reaching political gossip and general narratives. For example, AI-generated and misleading videos were used in political campaigning recently, such as in the Assam Bharatiya Janata Party-linked clip that communalised voter awareness messaging (Mahalanobish 2025). In another instance, political actors circulated misleading graphics about the Prime Minister's new official residence (Majumder 2026). Meanwhile, deepfakes have also helped in both scams and endorsements as political phenomena at the intersection of celebrity culture. Fictional videos of cricketer Virat Kohli and prominent businessman Anant Ambani supporting false promises was another manifestation of how trust in public figures can be monetised (CloudSEK 2025). These stories are effective because they are untraceable and consistent with prosumer biases. AI adds credibility to them, making them much more persuasive than traditional forms of misinformation.

4.3. False Attribution to Media: Erosion of Credibility

The misuse of synthetic media to impersonate journalists is an increasingly prominent aspect of the misinformation landscape in India. Deepfakes and manipulated social media posts have also been used against journalists and media institutions, leading to eroding credibility and the spread of propaganda masquerading as news. An article on fact-checking website BOOM Live, titled "Indian journalists sound alarm over their AI clones on YouTube," noted that "AI deepfakes of Hindi news journalists are propagating disinformation and warping perception" (Kumar 2025).

Indian journalist Ravish Kumar is one of the better-known cases, with his identity frequently simulated in fake videos circulating online. But several other journalists have also been victims of such manipulation. For example, in 2025, videos attributed to journalists Palki Sharma and Shiv Aroor went viral on social media and purportedly captured them "criticising India's diplomacy and praising Pakistan's role in the West Asia conflict" (Varma

2025; Dhonge 2025). None of these journalists had made these statements in any social media posts or in any news bulletin, as the videos seem to suggest, and were, in fact, AI-generated according to factcheckers. The viral videos purportedly made by these journalists also falsely claimed that "key Bihar NDA [National Democratic Alliance] leaders are under investigation for allegedly funding Pakistan's political party, Pakistan Tehreek-e-Insaf (PTI)" (Dhonge 2025). One aspect of these cases that is especially troubling is that they contribute to a generalized loss of trust. When reputable and well known journalists apparently express sensational, outlandish, or improbable views, even briefly, it can breed confusion and suspicion among audiences and result in declining confidence in legitimate journalism. Other Indian journalists have also suffered similar AI-generated mischaracterizations.

There are also fake social media posts intended to mimic reputable news organizations – especially around opinion polls or breaking news – that circulate, but are later debunked (Anand 2025; Mehrotra 2024). These point to a larger trend of using the logos of credible media outlets or videos with the fabricated voices of reputable individuals to lend credence to fake narratives.

4.4. Contemporary Power Plays: War, Clickbait, and the Growth of Open-Source Intelligence Propaganda

AI-generated synthetic media and deepfakes have begun to play crucial roles in how current affairs are explained, consumed, and understood on social media. During the recent Iranian crisis of 2026, for example, social media platforms were flooded with dramatic visuals featuring diverse claims – for example, burnt U.S. fighter planes, destroyed Israeli homes, and the injured body of Benjamin Netanyahu (Mahalanobish 2026; Chowdhury 2026; Dwivedi 2026a). Indeed, allegations about Benjamin Netanyahu's death – with purported bunker footage, his damaged house, and the appearance of an extra finger in a subsequent video – kept the social media busy for days till they were finally debunked (Chowdhury 2026; Dwivedi 2026a). Such visual content originated from all quarters, each pushing their perspective on triumph and destruction. However, large numbers of these videos were subsequently found to be either old or AI-generated. This is not a new phenomenon. Earlier forms of social media as well as many media channels have used old or even computer-game footage misleadingly, but with the proliferation of AI-generated synthetic media, it has become much easier to stage and platform such content. Synthetic videos thrive on clickbait algorithms by using recycled data, sensationalized captions, and emotionally triggering scripts to create maximum noise online, sometimes promoted via attention-grabbing captions such "breaking," "leaked footage," "special investigation," or "exclusive."

Another growing practice observed on social media is the rise of a set of digital content creators and social media influencers who thrive on the "attention economy" by creating content that keeps the audience closely engaged for extended periods. This narrative economy works not through claims to credibility but by producing visually convincing and emotionally gripping content. Some pretend to be Open Source Intelligence (OSINT) experts every time there is a new conflict. OSINT uses publicly available information, such metadata, satellite locations, images, and geolocations, to clarify ambiguous information. However, many self-styled

warfare experts misuse this label for traction and engagement, and spread misleading or false information to satisfy the curiosity of audiences who want simple explanations for complex events. Consequently, these audiences accept misinformation that aligns with their confirmation bias (Dwivedi 2026b).

When mainstream media seems slow, or bound to a corner, audiences increasingly look to OSINT-style accounts for real-time updates. But many of these accounts are produced by analysts who are far from neutral, and who amplify propaganda. In some instances, coordinated networks have been identified. For example, an investigation by India Today Fact Check found the social media account @InsiderWB was circulating misleading and doctored videos aimed at Indian policymakers and institutions (Dwivedi 2026b). The hybrid nature of the content makes it more difficult to detect. Real footage, AI-generated visuals, and selectively edited clips are often presented together, making verification difficult even for trained factcheckers.

Recent reports highlight how pro-state content creators, media, and non-state actors use AI-generated images and videos combined with real material on-site to create narratives so convincing that it is difficult to untangle them. For example, AI-generated propaganda warfare has become a new tool between Indians and Pakistanis during each episode of tension (Shah 2025; Ellis-Petersen 2025; Awasthi 2025).

At one level are obviously manufactured deepfakes that are overtly meant to be shocking or provocative, but at another, semi-credible actors use OSINT language and aesthetics to justify dubious conclusions. The result, together, is what one might call an "information fog:" the distinction between fact, interpretation, and fabrication becomes less and less clear (Badke 2021).

In this competitive climate, clickbait is not only a stylistic strategy but also a tactical tool that helps maintain visibility in a cluttered digital space, amplifies algorithmically to drive up the volume of similar content, and helps fuel further production. The result is a cycle that denies the truth, and that is not only consumed, but is also rewarded for being the source and structure of an untruth. Such a situation is indicative of the circulation effect of affective publics (Papacharissi 2015), where emotionally charged (and aesthetically appealing) media is circulated faster than verified information. In this environment, even loosely framed "analysis" by self-styled experts can gain legitimacy if it is couched in technical language or based on visual evidence.

4.5. Gendered Violence: Deepfake as a Means of Control & Manipulation

Perhaps the most troubling use of AI-generated media concerns gender-based violence. Women have been most disproportionately affected by sexually explicit deepfakes and morphing, including public figures and journalists, as well as young students. A 2023 deepfake video of the South Indian actor Rashmika Mandanna had for the first time brought national attention to such manipulation (Metro Vartha 2023). Meanwhile, journalist Rana Ayyub has faced serious deepfake attacks since 2018, with frequent deepfake images of her own personal photos taken from Instagram, and others have been the victims of deepfake sexual content designed to intimidate them (George 2024; Posetti et al. 2023). Several prominent female journalists and activists, as well as female members of

the general public, have faced similar deepfake attacks online. A news report from Kolkata reveals that deepfake images are a popular tool of blackmail and coercion, include threats that attackers will release deepfake pornographic images of the targeted women (Niyogi 2025). The longstanding issue of gender-based violence has now turned into digital misogyny (Raj and Yadav 2025). This violence is not only reputational; it is deeply psychological and has driven the victim to suicide in many instances (Niyogi 2025; Chitre 2025).

5. Discussion and Conclusion

Deepfakes, AI manipulations, and synthetic media can be lethal tools of disinformation and propaganda. The emerging pattern points to their use as a deception tool, which erodes trust in visual media. In this post-truth era, the proliferation of synthetic media that often aligns with the confirmation bias of prosumers is creating filter bubbles and echo chambers. Even though many of those videos are later fact-checked and found to be fabricated, they have already served their purpose: to make large numbers of consumers believe them. Even though current technology offers limited scope to detect out-of-sync audio, it seems only a matter of time before near perfect synthetic videos can be easily generated. Their circulation of emotional appeals that trump credibility and facts shapes the dynamics of affective publics.

While traditional propaganda models were often centralized, top-down, and had set targets, digital propaganda aided by synthetic media and AI reveal a more diffused, decentralized, and networked model of proliferation within bee-hives like networked filter bubbles. They do not necessarily need direct institutional support but can manage through what Benkler et al. (2018) term "networked propaganda." Partisan social handles, influencers, anonymous users, and digital content creators, pretending to be self-styled experts on OSINT, economics, or politics, act as important nodes in this process of creating, rehashing, and disseminating narratives and unverified claims for traction, the pursuit of ideological agendas or direct monetary gains.

Such deepfakes and AI-generated content are not merely shaping engagement and the attention economy, but in certain cases they may also lead to real-world harm. While manipulated quotations or attributions damage the credibility of celebrities, politicians, or journalists, many young women have also found themselves in vulnerable positions due to AI-manipulated images or videos. Although misogyny is a longstanding social problem, new technologies have further augmented its scope, leading to many instances of digital gender violence.

Further, synthetic media has aggravated the growing trust deficit in media by contributing to a climate in which audio-visuals are often viewed with suspicion and seen as no longer reliable. This leads to the fragmentation of the public sphere and produces diverse groups who remain within and believe in their own social realities, that is, their own filter bubbles. In relation to current affairs and global conflicts, such fragmentation further blurs the line between what is real and plausible, and what is fake. The belief that *anything can be fake* will only increase cynicism and the lack of trust in media, and may lead to further polarization in an already polarized Indian society.

Deepfakes and AI-generated synthetic media are thus not just technological whims or interesting novelties. They have turned into a potent weapon in the disinformation

ecosystem. By misquoting politicians, constructing inaccurate stories during conflicts, spreading false rumours, and sexualising women, AI-generated content can have devastating effects due its accessibility and scalability. While fact-checking groups and legal systems have taken steps to counter these developments, they are, at best, responding to a growing and evolving danger and have yet to be effective.

Finally, the crisis of deepfakes is not only technological but its cumulative effect also threatens the very premise of our belief system, in which images, video, and audio evidence have largely been seen as sacrosanct and credible. Use of synthetic media as deepfakes challenges that very premise – is what we see on TV or the internet real or fabricated? Who decides which is real and which is fake? Unless you have seen or met Benjamin Netanyahu, how do you know whether the AI-generated images of his death are true? The case studies analyzed in this article are symptomatic of serious problems in the information ecosystem of India, in which the production and dissemination of AI-generated deepfakes is rapidly increasing and for which there seems to be no immediate remedy. Tackling these problems will require more coordinated global interventions – involving technological filters, legal and political readiness, and regulatory frameworks – and, above all, a reconceptualization of media practice.

Declaration of AI Use

No AI tool has been specially used in the writing of this article or editing. The authors used AI tools such as DeepSeek and ChatGPT to locate earlier instances of AI abuse and cite them as examples.

References

Aïmeur, E., S. Amri, and G. Brassard. 2023. "Fake News, Disinformation and Misinformation in Social Media: A Review." *Social Network Analysis and Mining* 13(1):30. https://doi.org/10.1007/s13278-023-01028-5.

Alam, M. M. 2026. "Did MEA Spokesperson Threaten Iran?" *India Today Fact Check*, March 12. Accessed on March 25, 2026. https://www.indiatoday.in/fact-check/story/fact-check-did-mea-spokie-threaten-iran-at-a-presser-no-this-video-is-a-deepfake-2881080-2026-03-12.

Allcott, H. and M. Gentzkow. 2017. "Social Media and Fake News in the 2016 Election." *Journal of Economic Perspectives* 31(2):211–236. https://doi.org/10.1257/jep.31.2.211.

Anand, A. 2025. "Did India Today Predict BJP's Win in Delhi Elections in Its Pre-Poll Survey? No!" *The Quint*, February 4. Accessed on February 20, 2026. https://www.thequint.com/news/webqoof/india-today-pre-poll-survey-delhi-elections-bjp-48-seats-viral-graphic-fact-check.

Awasthi, S. 2025. "Pakistan's Information Warfare: Strategic Implications and India's Response." *Observer Research Foundation*, October 22. Accessed on March 10, 2026. https://www.orfonline.org/research/pakistan-s-information-warfare-strategic-implications-and-india-s-response.

Badke, W. 2021. *Research Strategies: Finding Your Way Through the Information Fog*. Bloomington, IN: iUniverse.

Baeza-Yates, R. and U. M. Fayyad. 2022. "The Attention Economy and the Impact of Artificial Intelligence." In *Perspectives on Digital Humanism*, edited by H. Werthner, E. Prem, E. A. Lee, and C. Ghezzi. Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-030-86144-5_18.

Benkler, Y., R. Faris, and H. Roberts. 2018. *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. Oxford, UK: Oxford University Press.

https://doi.org/10.1093/oso/9780190923624.001.0001.

Bennett, W. L. and A. Segerberg. 2012. "The Logic of Connective Action." *Information, Communication & Society* 15(5):739–768. https://doi.org/10.1080/1369118X.2012.670661.

Bozdag, E. 2013. "Bias in Algorithmic Filtering and Personalization." *Ethics and Information Technology* 15(3):209–227.

Chitre, M. 2025. "Faridabad Man Blackmailed With AI-Generated Pics Of 3 Sisters. He Dies By Suicide." *NDTV*, October 27. Accessed on February 14, 2026. https://www.ndtv.com/india-news/faridabad-man-blackmailed-with-ai-generated-pics-of-3-sisters-he-dies-by-suicide-9522971.

Chowdhury, A. 2026. "Is The Israeli PM Alive? Why Three Videos Couldn't Settle The Debate." *BOOM Live*, March 20. Accessed on April 5, 2026. https://www.boomlive.in/explainers/netanyahu-deepfake-video-analysis-israel-iran-conflict-misinformation-30884.

Chueca Del Cerro, C. 2024. "The Power of Social Networks and Social Media's Filter Bubble in Shaping Polarisation: An Agent-Based Model." *Applied Network Science* 9:69. https://doi.org/10.1007/s41109-024-00679-3.

CloudSEK. 2025. "Deepfake Scams Using Virat Kohli and Anant Ambani." Accessed on March 22, 2026. https://www.cloudsek.com/blog/deepfake-controversy-scammers-use-deepfakes-of-virat-kohli-anant-ambani-to-fraud.

Couldry, N. and A. Hepp. 2017. *The Mediated Construction of Reality*. Cambridge, UK: Polity Press.

Deshpande, S. 2025. "'Truly Alarming': Bombay High Court Orders Removal of Deepfake Content Infringing Akshay Kumar's Personality Rights." *The Times of India*, October 16. Accessed on January 10, 2026. https://timesofindia.indiatimes.com/city/mumbai/truly-alarming-bombay-high-court-orders-removal-of-deepfake-content-infringing-akshay-kumars-personality-rights/articleshow/124611724.cms.

Dhonge, N. K. 2025. "These Viral Videos Allegedly Showing Indian Journalists Reporting That NDA Leaders Funded Pakistan's Political Party PTI Are Fake." *Factly*, November 24. Accessed on February 28, 2026. https://factly.in/these-viral-videos-allegedly-showing-indian-journalists-reporting-that-nda-leaders-funded-pakistans-political-party-pti-are-fake/.

Dutta, R. 2026. "Fake Video of Modi Announcing Aid Debunked." *India Today Fact Check*, February 26. Accessed on March 18, 2026. https://www.indiatoday.in/fact-check/story/fact-check-pak-accounts-share-fake-video-of-modi-announcing-aid-for-afg-in-israel-2874823-2026-02-26.

Dwivedi, J. 2026a. "Fact Check: Israeli Reporter Confirms Netanyahu's Death? Viral Clip Shows Something Else." *India Today Fact Check*, March 17. Accessed on April 2, 2026. https://www.indiatoday.in/fact-check/story/fact-check-israeli-reporter-confirms-netanyahus-death-viral-clip-shows-something-else-2883325-2026-03-17.

Dwivedi, J. 2026b. "OSINT in Name, Propaganda's the Game: Meet the 'War Experts' Spreading Disinformation." *India Today Fact Check*, March 11. Accessed on April 2, 2026. https://www.indiatoday.in/fact-check/story/osint-propaganda-meet-war-experts-spreading-disinformation-2880563-2026-03-11.

Ellis-Petersen, H. 2025. "How Social Media Lies Fuelled a Rush to War Between India and Pakistan." *The Guardian*, May 28. Accessed on January 5, 2026. https://www.theguardian.com/media/2025/may/28/how-social-media-lies-fuelled-a-rush-to-war-between-india-and-pakistan.

Fact Check Bureau. 2026. "Shashi Tharoor Deepfake Video Debunked." *India Today Fact Check*, March 26. Accessed on April 10, 2026. https://www.indiatoday.in/fact-check/story/shashi-tharoor-deepfake-video-false-praise-pak-diplomacy-2887497-2026-03-26.

Fallis, D. 2021. "The Epistemic Threat of Deepfakes." *Philosophy & Technology* 34(4):623–643. https://doi.org/10.1007/s13347-020-00419-2.

Festinger, L. 1957. *A Theory of Cognitive Dissonance*. Evanston, IL: Row Peterson.

Fisher, S. A., J. W. Howard, and B. Kira. 2024. "Moderating Synthetic Content: The Challenge of Generative AI." *Philosophy & Technology* 37:133. https://doi.org/10.1007/s13347-024-00818-9.

Flaxman, S., S. Goel, and J. M. Rao. 2016. "Filter Bubbles, Echo Chambers, and Online News Consumption." *Public Opinion Quarterly* 80(S1):298–320.

George, A. 2024. "Defamation in the Time of Deepfakes." *Columbia Journal of Gender and Law* 45(1):122–172. https://doi.org/10.52214/cjgl.v45i1.13186.

Ghosh, D. S. and S. De Sarkar. 2024. "Combatting Deepfakes in India: Legal and Regulatory Challenges." *Indian Journal of Law and Justice* 15(1):346-376. Accessed on March 5, 2026. https://www.researchgate.net/publication/381302175.

Gillespie, T. 2018. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. New Haven, CT: Yale University Press.

Habermas, J. 1989. *The Structural Transformation of the Public Sphere*. Cambridge, MA: MIT Press.

Herman, E. S. and N. Chomsky. 1988. *Manufacturing Consent: The Political Economy of the Mass Media*. New York: Pantheon Books.

Iyengar, S. and K. Hahn. 2009. "Red Media, Blue Media: Evidence of Ideological Selectivity in Media Use." *Journal of Communication* 59(1):19–39.

Jain, A. 2025. "Deepfakes and Misinformation: Legal Remedies and Legislative Gaps." *Indian Journal of Law* 3(2):23–28. https://doi.org/10.36676/ijl.v3.i2.86.

Jha, P. and S. Jain. 2021. "Detecting and Regulating Deepfakes in India: A Legal and Technological Conundrum." *SSRN Electronic Journal*. http://dx.doi.org/10.2139/ssrn.4411227.

Kahneman, D. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Kaur, S., M. Buhari, N. Khandelwal, P. Tyagi, and K. Sharma. 2024. "Hindi Audio-Video-Deepfake (HAV-DF): A Hindi Language-Based Audio-Video Deepfake Dataset." *arXiv*, November 23. Accessed on January 20, 2026. https://arxiv.org/abs/2411.15457.

Keyes, R. 2004. *The Post-Truth Era: Dishonesty and Deception in Contemporary Life*. New York: St. Martin's Press.

Kishwar, S. D., A. Tripathi, S. Khatoon, D. Poddar, and B. Khurana. 2025. "Regulating Deep Fakes and Synthetic Media: Privacy, Policy and Global Regulatory Challenges." *Journal of Data Protection & Privacy* 8(1):78-96. https://doi.org/10.69554/HBGG8150.

Kumar, R. 2025. "Indian Journalists Sound Alarm Over Their AI Clones On YouTube." *BOOM Live*, November 7. Accessed on February 8, 2026. https://www.boomlive.in/decode/indian-journalists-sound-alarm-over-their-ai-clones-on-youtube-29932.

Laskar, K. A. and M. Reyaz. 2021. "Mapping the Fake News Infodemic Amidst the COVID-19 Pandemic: A Study of Indian Fact-Checking Websites." *Journal of Arab & Muslim Media Research* 14(1):93-116. https://doi.org/10.1386/jammr_00026_1.

Mahalanobish, A. 2024. "Not Amitabh Bachchan: Viral Video Uses AI-Generated Voice." *Alt News*, November 26. Accessed on January 15, 2026. https://www.altnews.in/not-amitabh-bachchan-voice-making-scathing-remarks-on-state-of-the-nation-in-viral-video-is-not-of-the-actor/.

Mahalanobish, A. 2025. "SIR Gives BJP Fresh Excuse to Demonize Muslims." *Alt News*, December 10. Accessed on February 5, 2026. https://www.altnews.in/sir-gives-bjp-fresh-excuse-to-demonize-muslims-and-create-polarization-bengal-delhi-units-lead-the-way/.

Mahalanobish, A. 2026. "Documenting the War or Fabricating It? AI-Generated Visuals on Social Media Blur the Line as Iran Conflict Escalates." *Alt News*, March 9. Accessed on April 12, 2026. https://www.altnews.in/documenting-the-war-or-fabricating-it-ai-generated-visuals-on-social-media-blur-the-line/.

Majumder, S. 2026. "Year in Review: Deepfakes and AI Supercharged Disinformation

Propaganda." *Alt News*, January 1. Accessed on March 1, 2026. https://www.altnews.in/year-in-review-deepfakes-and-ai-supercharged-disinformation-propaganda-in-2025-like-never-before/.

McIntyre, L. 2018. *Post-Truth*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/11483.001.0001.

Mehrotra, K. 2024. "Fact-Check: Fake Exit Poll Graphic From TNM Predicting NDA Win in AP Goes Viral!" *The Quint*, May 31. Accessed on January 25, 2026. https://www.thequint.com/news/webqoof/fake-graphic-of-andhra-pradesh-exit-polls-with-tnm-logo-viral-fact-check.

Metro Vaartha. 2023. "Govt Issues Advisory to Social Media Companies on Rashmika Mandanna Deep Fake Video." *Metro Vaartha*, November 7. Accessed on January 18, 2026. https://english.metrovaartha.com/news/national/govt-issues-advisory-to-social-media-companies-on-rashmika-mandanna-deep-fake-video.

Naeem, S. B., R. Bhatti, and A. Khan. 2020. "An Exploration of How Fake News is Taking Over Social Media and Putting Public Health at Risk." *Health Information & Libraries Journal* 38(2):143–149. https://doi.org/10.1111/hir.12320.

Naffi, N. 2025. "Deepfakes and the Crisis of Knowing." *UNESCO.org*, October 1. Accessed on February 22, 2026. https://www.unesco.org/en/articles/deepfakes-and-crisis-knowing.

Nickerson, R. S. 1998. "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises." *Review of General Psychology* 2(2):175–220. https://doi.org/10.1037/1089-2680.2.2.175.

Niyogi, S. 2025. "Trolls to Deepfake: Navigating Gender-Based Online Violence." *The Times of India*, November 26. Accessed on March 8, 2026. https://timesofindia.indiatimes.com/city/kolkata/trolls-to-deepfake-navigating-gender-based-online-violence/articleshow/125572131.cms.

Öhman, C. 2019. "Introducing the Pervert's Dilemma: A Contribution to the Critique of Deepfake Pornography." *Ethics and Information Technology* 22(2):133–140. https://doi.org/10.1007/s10676-019-09522-1.

Öhman, C. 2022. "The Identification Game: Deepfakes and the Epistemic Limits of Identity." *Synthese* 200:319. https://doi.org/10.1007/s11229-022-03798-5.

Okocha, M. U. 2026. "Journalism in the Shadows: AI Deepfakes, Credibility, and the Fragility of Media Trust." *Journalism Practice*:1–23. https://doi.org/10.1080/17512786.2026.2719552.

Papacharissi, Z. 2015. *Affective Publics: Sentiment, Technology, and Politics*. Oxford, UK: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199999736.001.0001.

Pariser, E. 2011. *The Filter Bubble: What the Internet is Hiding from You*. London: Penguin Group.

Paul, S. and B. Dutta. 2025. "Unveiling Trends in Deepfake Research: A Global Bibliometric Perspective." *DESIDOC Journal of Library & Information Technology* 45(3):224–237. https://doi.org/10.14429/djlit.20858.

Pennycook, G. and D. G. Rand. 2019. "Lazy, Not Biased: Susceptibility to Partisan Fake News." *Cognition* 188:39–50. https://doi.org/10.1016/j.cognition.2018.06.011.

Posetti, J., K. Bontcheva, H. Zaffar, N. Shabbir, D. Maynard, and M. Pandya. 2023. *Rana Ayyub: Targeted Online Violence at the Intersection of Misogyny and Islamophobia, Big Data Case Study*. International Center for Journalists (ICFJ). Accessed on January 30, 2026. https://www.icfj.org/sites/default/files/2023-02/Rana%20Ayyub_Case%20Study_ICFJ.pdf#page=2.11.

Raj, D. and O. K. Yadav. 2025. "Unmasking Deception: Investigating the Rise and Impact of Deepfake Crimes in India." *Asian Journal of Advances in Research* 8(1):256–265. https://doi.org/10.56557/ajoair/2025/v8i1524.

Roy, E., K. A. Laskar, and M. Reyaz. 2025. "Retroactive Gatekeepers: Fact-Checkers, Electoral Disinformation, and Audience Engagement in India." *Emerging Media* 3(4):582-600. https://doi.org/10.1177/27523543251395190.

Shah, A. A. 2025. "AI, Deepfakes, and the Fog of War - Disinformation in the 2025 India-Pakistan Conflict." *International Federation of Journalists (IFJ)*, June 25. Accessed on February 12, 2026. https://www.ifj.org/media-centre/news/detail/category/ai/article/ifjblog-ai-deepfakes-and-the-fog-of-war-disinformation-in-the-2025-india-pakistan-conflict.

Sharma, M. D. and A. N. Shaikh. 2025. "A Case Study on Deepfake Awareness, Mumbai, India." *Indian Journal of Computer Science and Technology* 4(1):121-126. https://doi.org/10.59256/indjcst.20250401019.

Shu, K., A. Sliva, S. Wang, J. Tang, and H. Liu. 2017. "Fake News Detection on Social Media." *ACM SIGKDD Explorations Newsletter* 19(1):22–36. https://doi.org/10.1145/3137597.3137600.

Srivastava, R. 2026. "Deepfakes and Digital Deception: A Study of India's Legal Preparedness." *International Journal for Multidisciplinary Research (IJFMR)* 8(2):1-10. Accessed on April 8, 2026. https://www.ijfmr.com/papers/2026/2/71322.pdf.

Tandoc, E. C., Z. W. Lim, and R. Ling. 2018. "Defining 'Fake News'." *Digital Journalism* 6(2):137–153. https://doi.org/10.1080/21670811.2017.1360143.

TOI Legal Desk. 2026. "Delhi High Court Protects Cricketer Gautam Gambhir's Personality Rights Against AI Deepfakes." *The Times of India*, March 29. Accessed on April 20, 2026. https://timesofindia.indiatimes.com/legal/news/delhi-high-court-protects-cricketer-gautam-gambhirs-personality-rights-against-ai-deepfakes/articleshow/129877893.cms.

TOI News Desk. 2025. "Fake Video of Jaishankar Apologising Exposed." *The Times of India*, May 10. Accessed on January 28, 2026. https://timesofindia.indiatimes.com/india/video-of-eam-jaishankar-apologising-fake-pib-exposes-false-propaganda/articleshow/121056141.cms.

TOI News Desk. 2026. "'Neither My Language Nor My Voice': Shashi Tharoor Flags Fake AI Video Claiming He Praised Pakistan's T20 World Cup Boycott." *The Times of India*, February 12. Accessed on March 15, 2026. https://timesofindia.indiatimes.com/india/neither-my-language-nor-my-voice-shashi-tharoor-flags-fake-ai-video-claiming-he-praised-pakistans-t20-world-cup-boycott/articleshow/128262272.cms.

Tucker, J., A. Guess, P. Barbera, et al. 2018. "Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature." *Social Science Research Network (SSRN)*:1–95.

Varma, A. 2025. "Deepfake Videos Falsely Attribute Statements to Indian Journalists." *The Quint*, November 25. Accessed on February 25, 2026. https://www.thequint.com/news/webqoof/ravish-kumar-palki-sharma-shiv-aroor-talking-about-bihar-nda-funding-pakistan-pti-deepfakes-fact-check.

Verma, A. 2026. "Deepfakes and the Crisis of Digital Authenticity: Ethical Challenges in the Age of Synthetic Media." *Journal of Information, Communication and Ethics in Society* 24(1):59–76. https://doi.org/10.1108/JICES-04-2025-0083.

Vosoughi, S., D. Roy, and S. Aral. 2018. "The Spread of True and False News Online." *Science* 359(6380):1146–1151. https://doi.org/10.1126/science.aap9559.

Wardle, C. and H. Derakhshan. 2017. *Information Disorder: Toward an Interdisciplinary Framework*. Council of Europe Report. Accessed on March 12, 2026. https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework/168076277c.

Warzel, C. 2018. "Believable: The Terrifying Future of Fake News." *BuzzFeed News*, February 12. Accessed on January 11, 2026. https://www.buzzfeednews.com/article/charliewarzel/the-terrifying-future-of-fake-news.

Young, G. 2021. *Fictional Immorality and Immoral Fiction*. Lanham, MD: Lexington Books.