



Truth in the Age of Illusion: Fake News, Deepfakes and the Crisis of Credibility in Digital Journalism

Dr. Vikas Dogra Professor, Department of Journalism and Mass Communication Himachal Pradesh University, Shimla, India Chetan Tripathi Research Fellow, Department of Journalism and Mass Communication Himachal Pradesh University, Shimla, India	Citation: Dogra, V., & Tripathi, C. (2026). Truth in the Age of Illusion: Fake News, Deepfakes and the Crisis of Credibility in Digital Journalism. In Innovative Research Thoughts in Social Sciences (Vol. 2, Number 2, pp. 36–52). DOI: https://doi.org/10.5281/zenodo.20637379
--	---

Abstract

The rapid expansion of digital media ecosystems has fundamentally transformed the production, dissemination, and consumption of news, while simultaneously intensifying concerns over credibility in journalism. The proliferation of fake news and the emergence of advanced deepfake technologies have blurred the boundaries between truth and fabrication, posing serious challenges to journalistic integrity and audience trust in the Indian media landscape between 2019 and 2024.

The primary objectives of this study are: (1) To conceptualize fake news and deepfakes within the framework of digital journalism; and (2) To analyze the impact of fake news and deepfakes on journalistic credibility and public trust, with particular reference to India's digital media environment. Adopting a systematic review and critical synthesis methodology, this paper draws on peer-reviewed journal articles, books, policy documents, and credible media reports published between 2019 and 2024. Conceptual and thematic analysis is employed to examine key constructs are redibility, authenticity, and trust situating them within communication theories such as agenda setting, framing, and post truth discourse.

The study concludes that the rise of fake news and deepfakes represents a profound epistemological challenge to journalism, necessitating renewed emphasis on ethical journalism practices, robust verification mechanisms, platform accountability, and systemic media literacy to sustain trust in India's evolving digital news environment.

Keywords: Fake News, Deepfakes, Digital Journalism, Credibility Crisis, Public Trust, India, Media Literacy, Post-Truth

1. Introduction

The contemporary media landscape in India stands at a critical juncture. With over 820 million internet users and a social media penetration rate that has grown exponentially since 2019, India has emerged as one of the world's most dynamic and consequential digital news markets (TRAI, 2023). The proliferation of smartphones, affordable data plans driven by the Jio revolution, and the ubiquity of messaging platforms such as WhatsApp have democratized information access on an unprecedented scale. Yet this democratization carries with it a profound paradox: the same channels that empower citizens to participate in public discourse have also become conduits for misinformation, manipulation, and outright fabrication.

'Fake news' a contested but analytically useful term denoting deliberately fabricated or misleading information presented in the guise of legitimate journalism, has emerged as one of the defining crises of twenty first-century democracy. In India, this crisis is not merely academic. Between 2019 and 2024, misinformation campaigns have been documented in various and diverse contexts ranging from the 2019 and 2024 general elections to the COVID-19 pandemic, from health emergencies to communal violence, each of them demonstrating the potential of fake news or false information to generate real world harm (Sharma et al., 2021; Internet Freedom Foundation, 2020).

Compounding this challenge is the rapid evolution of deepfake technology artificial intelligence generated synthetic media capable of

rendering realistic but entirely fabricated audio, video, and images. The accessibility of deepfake generation tools has lowered the barrier for sophisticated disinformation campaigns. In India, high-profile deepfake incidents involving political figures and public personalities have emerged with alarming number since 2022, raising urgent questions about authentication, consent, and the viability of visual evidence as a journalistic tool (Dhawan, 2023; MIT Technology Review, 2023).

The Principle of credibility in journalism that is considered as the normative foundation of the profession is now facing a systemic erosion. Credibility, understood as the audience's perception of a news source's trustworthiness and expertise (Hovland & Weiss, 1951; Flanagin & Metzger, 2007), is not merely a professional virtue but a sociological and communicative function essential to the democratic role of the press. When audiences can no longer distinguish authentic journalism from algorithmically generated disinformation or AI synthesized video, the epistemic foundations of informed citizenship are undermined.

This paper seeks to engage critically with these developments through a theoretically grounded and methodologically rigorous review of existing scholarship. Specifically, it pursues two primary objectives: first, to conceptualize fake news and deepfakes within the framework of digital journalism theory and second, to analyze the impact of these phenomena on journalistic credibility and public trust, with sustained attention to the Indian context between 2019 and 2024.

2. Theoretical Framework

The analytical architecture of this paper rests upon an integrated theoretical framework that draws on three established paradigms within communication studies: agenda setting theory, framing theory, and post-truth discourse. Together, these frameworks provide the conceptual vocabulary necessary to understand how fake news and deepfakes operate as communicative phenomena and why their consequences for journalistic credibility are so profound.

2.1 Agenda -Setting Theory and the Digital Disruption

Originally formulated by McCombs and Shaw (1972) through their seminal Chapel Hill study, agenda-setting theory posits that mass media do not tell audiences what to think, but are strikingly effective at telling them what to think about. The agenda-setting function rests on the media's capacity to assign salience to particular issues, thereby shaping public perception of what matters. In the digital era, this function has been fundamentally destabilized.

The rise of social media platforms as primary news distribution channels has fragmented the agenda-setting process, distributing it across a multiplicity of actors including but not limited to bots, trolls, partisan platforms, and algorithmic recommendation engines none of whom are regulated by the ethical guidelines or norms governing professional journalism. Research conducted during the 2019 Indian general elections

demonstrated that WhatsApp forwards played a significant role in determining which issues gained traction within localized communities, often amplifying fabricated stories and bypassing mainstream media gatekeeping entirely (Banaji & Bhat, 2019).

Deepfakes further complicate the agenda-setting framework. A manufactured video of a political leader making a fabricated statement can, if strategically timed and distributed, set the media agenda around a false premise forcing journalists into a reactive posture of denial and debunking rather than original reporting. This represents a qualitative shift in the power to set agendas from credentialed journalists to malicious actors equipped with AI generation tools.

Extending McCombs' second-level agenda-setting, which focuses not merely on what is covered but on the attributes through which issues are framed, one can observe how fake news systematically constructs attribute agendas that distort public understanding of persons, events, and communities. False information that is repeated often on different platforms starts to feel true because people recognize it more easily. This is called the illusory truth effect (Pennycook et al., 2018).

2.2 Framing Theory and the Construction of Misinformation

Framing theory, developed through the contributions of Goffman (1974), Gitlin (1980), and subsequently systematized by Entman (1993), holds that media coverage does not merely report reality but actively constructs it by

selecting, emphasizing, and excluding particular aspects of events. Frames organize experience and guide interpretation; they tell audiences not simply what happened, but what it means.

Fake news operates fundamentally as a framing device. By constructing false narratives through strategic frames like fear, threat, and moral outrage, misinformation exploits known cognitive heuristics to bypass critical evaluation. Research on the emotional architecture of misinformation demonstrates that fabricated stories are more likely to be shared when they invoke strong affective responses, particularly anger and anxiety (Vosoughi, Roy, & Aral, 2018). This observation has profound implications for the Indian context, where communal, regional, and political sensitivities provide fertile ground for frame manipulation.

Deepfakes represent a particularly insidious evolution of frame manipulation because they operate at the level of perceptual reality rather than narrative construction. A deepfake video does not merely frame an event but it fabricates the empirical evidence upon which frames are constructed. While doing this, it attacks the pre interpretive layer of journalistic communication based on the the raw footage, the recorded statement, the visual document that has historically been treated as authentic source.

The Indian fact-checking ecosystem, including organizations such as BOOM, AltNews, and The Quint's WebQoof, has documented numerous cases in which selectively edited or entirely fabricated videos were framed to inflame communal tensions, particularly in states like Uttar Pradesh, Assam, and Delhi (Sharma et al., 2021). These cases reveal the

intersection of framing and fabrication as a strategy of political and social manipulation.

2.3 Post-Truth Discourse and the Epistemological Crisis

The concept of post-truth, named Oxford Dictionaries' Word of the Year in 2016, describes a condition in which objective facts are less influential in shaping public opinion than appeals to emotion and personal belief. Keyes (2004) anticipated this trajectory, and subsequent scholars including McIntyre (2018) and D'Ancona (2017) have theorized it as a systemic feature of late modern political communication.

Post-truth discourse is not simply about lying; it represents a deeper epistemological shift in which the very notion of shared, verifiable reality becomes contested. When citizens routinely encounter professionally produced false information, when verified facts can be dismissed as 'fake news,' and when deepfakes render visual evidence unreliable, the conditions for post-truth discourse are not merely present but they are actively cultivated.

For journalism, post-truth presents an existential challenge. The profession's normative authority rests on its claim to provide verifiable, objective accounts of reality. In a post-truth environment, this claim is perpetually under attack through the strategic deployment of doubt. India's media environment between 2019 and 2024 has exhibited characteristic features of post-truth culture through the systematic labeling of inconvenient journalism as 'presstitute' , the proliferation of partisan news channels that blur the line between reporting and advocacy, and the

use of AI-generated content to manufacture consensus (Jeffrey & Doron, 2021; Raghavan, 2022).

Together, these three theoretical frameworks provide the analytical tools to examine how fake news and deepfakes function not merely as individual acts of deception but as structural forces reshaping the communicative ecology in which journalism operates. The integration of agenda-setting, framing, and post-truth analysis constitutes the theoretical backbone of this paper's inquiry.

3. Methodology

This paper adopts a qualitative, interpretive research design rooted in the tradition of systematic literature review and critical synthesis. Given the conceptual objectives of the study, to theorize the phenomena of fake news and deepfakes within digital journalism and to analyze their impact on credibility and public trust. An empirical data-collection approach would be neither appropriate nor sufficient. The study instead constructs its analysis through a rigorous engagement with existing scholarly, policy, and investigative literature.

3.1 Research Design and Philosophical Underpinnings

The study operates within an interpretivist epistemological paradigm, which holds that social phenomena including journalism, trust, and misinformation are not objectively given facts awaiting measurement but are socially constructed realities whose meaning

must be interpreted through contextually sensitive analysis (Creswell, 2014; Denzin & Lincoln, 2011). Consistent with this ontological position, the paper does not seek to generate statistically generalizable findings but to produce theoretically grounded, analytically nuanced understanding of complex communicative phenomena.

The review methodology employed is the systematic-narrative hybrid, which combines the rigour of systematic searching with the analytical flexibility of narrative synthesis. This approach, advocated by Snyder (2019) for conceptual and theoretical review papers, permits the identification of patterns, tensions, and gaps across a diverse body of literature while retaining the analytical depth appropriate for developing theory.

3.2 Inclusion and Exclusion Criteria

Literature was identified and selected according to explicit criteria across four dimensions. In terms of temporal scope, only literature published between January 2019 and December 2024 was included as primary material, ensuring alignment with the study's defined timeframe. Pre-2019 foundational theoretical works such as McCombs and Shaw (1972), Entman (1993), and Hovland and Weiss (1951) were retained as theoretical anchors where necessary.

Regarding geographical relevance, literature with direct applicability to the Indian context was prioritized, while global scholarship

was included where it provided theoretical frameworks or comparative insights applicable to India's media environment. Concerning source type, peer-reviewed journal articles, academic monographs, edited book chapters, government and regulatory policy documents (including those from the Ministry of Information and Broadcasting, the Press Council of India, and TRAI), reports from credible civil society organizations (Internet Freedom Foundation, Article 19, Reporters Without Borders), and documented investigations by recognized fact-checking organizations (BOOM, AltNews, The Quint's WebQoof) were included. Thematic relevance required that only literature directly addressing fake news, deepfakes, digital journalism, journalistic credibility, media literacy, or platform governance was included.

3.3 Search Strategy and Database Selection

Systematic searches were conducted across the following academic databases: Scopus, Web of Science, Google Scholar, JSTOR, and the Directory of Open Access Journals (DOAJ). Additional manual searches were conducted in the archives of leading Indian communication journals including Media Watch, the Journal of Content, Community and Communication, and Communicator: Journal of the Indian Institute of Mass Communication. Gray literature was accessed through the websites of regulatory bodies, civil society organizations, and recognized journalistic institutions.

The primary search strings employed included combinations of the following terms:

'fake news AND India,' 'deepfake AND journalism,' 'digital journalism AND credibility,' 'misinformation AND Indian media,' 'post-truth AND journalism,' 'media trust AND India,' 'deepfake AND political communication,' and 'fact-checking AND India.' Searches were conducted in English, as this represents the dominant language of peer-reviewed scholarship on these topics. Regional language scholarship was accessed through translated reports and citations within English language studies.

3.4 Analytical Approach

The analytical process proceeded through two interconnected stages. In the first stage, thematic synthesis, the literature was inductively coded to identify recurring conceptual themes. Key constructs including credibility, authenticity, trust, verification, media literacy, and platform accountability were mapped and cross-referenced across sources. This stage generated a thematic architecture that structures the presentation of findings in subsequent sections.

In the second stage, conceptual analysis, the identified themes were subjected to critical theoretical interrogation using the three frameworks outlined in Section 2. This stage moved beyond description to interpretation, examining not merely what the literature reveals but what it means for understanding the relationship between misinformation technologies and journalistic credibility. Following Fink (2020) and Torracco (2016), the paper's contribution lies not in the aggregation of findings but in their reinterpretation through a

theoretically coherent and contextually grounded analytical lens. The integrated analytical framework developed in Section 8 represents the original scholarly contribution of this synthesis.

4. Conceptualizing Fake News in the Digital Journalism Framework

4.1 Definitional Landscape and Terminological Debates

The term 'fake news' has become one of the most contested concepts in contemporary communication scholarship. Its definitional ambiguity is not incidental but is itself politically significant: the same label is deployed by disinformation producers to delegitimize authentic journalism and by media scholars to describe deliberate fabrication. Wardle and Derakhshan (2017), in their foundational typology commissioned by the Council of Europe, propose the more precise conceptual framework of 'information disorder,' which distinguishes between misinformation (false information shared without intent to harm), disinformation (false information shared with deliberate intent to deceive), and malinformation (true information shared with intent to cause harm).

This taxonomy is particularly useful for analyzing the Indian context, where all three categories operate simultaneously and often in overlapping configurations. The COVID-19 infodemic in India provides illustrative cases of all three: fabricated cures (disinformation), mistakenly shared incorrect advice (misinformation), and the strategic leaking of

private medical data to stigmatize communities (malinformation) (Sharma et al., 2021; Goel, 2020).

For the purposes of this paper, 'fake news' is employed as a broad analytical category encompassing content that falsely presents itself as professional journalism while lacking adherence to verification, sourcing, editorial accountability, or transparency. This usage, consistent with Tandoc, Lim, and Ling (2018), acknowledges the term's limitations while recognizing its utility as a shorthand for a cluster of related phenomena with shared characteristics and consequences for journalistic credibility.

4.2 Typologies and Modalities of Fake News in India (2019–2024)

The literature on fake news in India between 2019 and 2024 reveals several distinct modalities of production and circulation, each with characteristic features and impacts on journalistic credibility.

Fabricated news articles represent the most straightforward form, involving entirely invented stories presented in the visual and linguistic conventions of legitimate journalism. These range from sophisticated operations involving multiple fake news websites to crude WhatsApp forwards with no editorial infrastructure. The 'Pratibha Devi Singh' case in 2019 in which a fake article attributed to a mainstream news outlet was widely circulated claiming that the President had rescinded poll results illustrates the capacity of well-designed fabrications to penetrate public discourse (BOOM, 2019).

Manipulated authentic content constitutes a significant and particularly insidious category, involving the selective editing, decontextualization, or caption manipulation of genuine news footage and photographs. AltNews has documented hundreds of such cases between 2019 and 2024, many involving photographs from other countries or historical events relabeled to suggest contemporary Indian events. During the COVID-19 pandemic, images of migrant labor crises from other nations were routinely circulated as representations of the Indian situation, distorting public understanding of the crisis's scale and character (Sharma et al., 2021).

Satire and parody, while not inherently problematic, generate misinformation when satirical content is stripped of its contextual markers and recirculated as factual reporting. Several Indian satirical websites have seen their content misappropriated in this manner, leading to institutional confusion about the boundaries of journalistic genre.

State aligned partisan content, occupying the contested terrain between political communication and journalistic fraud, has been extensively documented in the Indian context. Studies examining WhatsApp political networks during the 2019 elections identified organized networks for the production and distribution of partisan fabrications, often originating from IT cells affiliated with major political parties (Banaji & Bhat, 2019; Lokniti-CSDS, 2019).

4.3 WhatsApp and the Indian Misinformation Ecosystem

No analysis of fake news in India is complete without sustained attention to WhatsApp, which functions not merely as a messaging platform but as a primary news medium for hundreds of millions of Indians, particularly in semi-urban and rural contexts where English-language print journalism has limited penetration. With approximately 487 million users in India as of 2023, WhatsApp constitutes the world's largest single-country user base for the platform (Statista, 2023).

WhatsApp's architecture presents particular challenges for fake news mitigation. Its end-to-end encryption, while essential for user privacy, renders the platform largely opaque to external researchers and regulatory scrutiny. The platform's group-based structure, in which content is forwarded through chains of interconnected groups, creates a viral distribution mechanism ideally suited to the rapid propagation of misinformation. Research conducted by the Internet and Mobile Association of India (IAMAI, 2020) found that over 62% of WhatsApp users in India encountered at least some potentially false news content weekly, while studies by the Reuters Institute (Newman et al., 2022) consistently identified WhatsApp as the platform on which Indian users most frequently encountered misinformation.

The linkage between WhatsApp misinformation and real-world violence in India represents one of the most disturbing dimensions of the fake news crisis. Between 2017 and 2020, over 30 deaths were directly attributed to mob violence triggered by viral WhatsApp messages falsely accusing individuals of being child kidnappers (Aldwaik, 2020; Human Rights

Watch, 2019). These cases, which span states from Maharashtra and Telangana to West Bengal and Jharkhand, underscore the life-or-death stakes of the credibility crisis in digital journalism.

5. Deepfakes: Technology, Proliferation, and Journalistic Consequences

5.1 Technical Genesis and Evolution

Deepfakes represent the application of deep learning technologies specifically generative adversarial networks (GANs) and, more recently, diffusion models to the synthesis of realistic but fabricated audiovisual content. The term itself derives from the pseudonymous Reddit user 'deepfakes' who, in 2017, began publishing AI-generated non-consensual intimate imagery of celebrities, thereby inaugurating public awareness of the technology's malicious potential. Since then, the technology has evolved with remarkable speed: tools that required significant computational expertise and resources in 2018 are now accessible through consumer-grade applications requiring no specialized knowledge (Chesney & Citron, 2019).

The technical mechanism underlying deepfakes involves the training of neural networks on large datasets of images or video of a target individual, enabling the synthesis of new content in which the target's likeness is mapped onto another body or voice pattern. Early deepfakes were identifiable through characteristic artifacts unnatural blinking, skin texture inconsistencies, lighting anomalies but advances in generation technology between 2020 and 2024

have substantially narrowed the gap between synthetic and authentic media (Tolosana et al., 2020).

Audio deepfakes, which clone voice patterns from limited sample data, have developed in parallel with video synthesis. By 2022, commercially available voice cloning platforms could generate convincing audio replicas from as little as three seconds of authentic speech, enabling the fabrication of statements attributed to identifiable individuals with high acoustic fidelity (Nautyal & Sharma, 2023).

5.2 Deepfakes in the Indian Context (2019–2024)

India's encounter with deepfake technology has been rapid and consequential. Several high-profile cases between 2019 and 2024 illustrate the range of applications and their implications for political communication, journalistic practice, and public trust.

The 2019 general elections witnessed what researchers identified as some of the earliest documented uses of deepfake video in an Indian electoral context. A video purportedly showing a prominent political figure making inflammatory statements was circulated through political WhatsApp networks in multiple states. While the video's fabricated nature was eventually established by technical analysts and fact-checkers, its initial circulation during a sensitive electoral period demonstrated the potential for

deepfakes to operate as election interference tools (Sharma et al., 2021).

The period 2022 to 2024 saw a marked acceleration in deepfake incidents targeting Indian public figures. A manipulated video of Prime Minister Narendra Modi was circulated in late 2023, depicting him engaged in an activity inconsistent with his public conduct. Multiple Bollywood actors, including Rashmika Mandanna and Kajol, became subjects of non-consensual deepfake intimate imagery that circulated widely on social media platforms, prompting public statements from central government ministers and accelerated legislative consideration (Press Trust of India, 2023; Dhawan, 2023).

The government's response to deepfake proliferation culminated in the IT (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules of 2023, which imposed explicit obligations on social media platforms to prevent the dissemination of deepfake content. The Ministry of Electronics and Information Technology issued advisory notices to major platforms including Meta, Google, and X (formerly Twitter) emphasizing their legal liability for the spread of AI-generated misinformation (MeitY, 2023). However, enforcement has remained inconsistent, and the regulatory framework has been critiqued by digital rights organizations for its potential to enable governmental censorship rather than genuine content accountability (Internet Freedom Foundation, 2023).

5.3 Deepfakes and the Epistemology of Visual Evidence

The deepfake crisis introduces a dimension that transcends the operational challenges posed by conventional misinformation: it attacks the epistemological status of visual and auditory evidence itself. Journalism's evidentiary tradition has long accorded particular epistemic authority to audiovisual documentation the photograph, the recorded statement, the video testimony. This authority is not absolute; documentary manipulation has been practiced since the earliest photographic era. However, the scale, accessibility, and sophistication of modern deepfake technology transforms what was historically a specialized capability into a democratized threat.

Chesney and Citron (2019) identify this dynamic as the 'liar's dividend': even in the absence of actual deepfakes, the knowledge of their existence enables bad-faith actors to dismiss authentic video evidence as potentially fabricated. In the Indian political context, this dynamic has been observed with increasing frequency since 2022, as politicians and their communications teams have preemptively raised the possibility of deepfake manipulation in response to authentic video documentation of inconvenient statements or conduct (Raghavan, 2022).

For journalism, the implications are severe. Verification processes that once relied on cross-referencing visual documentation with

independent sources must now incorporate technical deepfake detection as a standard element of editorial practice. Indian news organizations have been slow to develop this capacity; a survey of editorial guidelines at major Indian news networks conducted by the Network of Women in Media, India (NWMI, 2022) found that fewer than 12% had formal protocols for deepfake detection and verification as of 2022.

6. The Crisis of Credibility in Indian Digital Journalism

6.1 Conceptualizing Credibility in the Digital Environment

Credibility, as a construct within communication research, has been theorized across multiple dimensions since Aristotle's identification of ethos that is the character of the speaker as a primary determinant of persuasive authority. In the modern communication research tradition, the foundational framework established by Hovland and Weiss (1951) distinguished two primary components: expertise (the perceived knowledge and competence of the source) and trustworthiness (the perceived honesty and motivational integrity of the source). Subsequent scholars have extended this framework to incorporate dimensions of dynamism, attractiveness, and perceived similarity (McCroskey & Teven, 1999).

In the context of digital journalism, Flanagin and Metzger (2007) argued for a reconceptualization of credibility that accounts for the shift from a broadcast mode in which

audiences evaluate the credibility of identified institutional sources to a networked model in which information circulates without clear authorial or institutional attribution. In networked environments, credibility assessment becomes more cognitively demanding, less reliable, and more susceptible to manipulation by actors who can mimic the visual and discursive conventions of legitimate journalism.

For Indian audiences navigating a media ecosystem characterized by hyperpartisanship, commercial pressures on editorial independence, and high volumes of circulatory misinformation, credibility assessment has become both more essential and more difficult. Surveys conducted by the Reuters Institute for the Study of Journalism (Newman et al., 2022, 2023, 2024) consistently identified India among the countries with the lowest levels of overall news trust globally, with only 38% of Indian respondents in 2023 reporting that they trust most news most of the time—a figure that has declined from 44% in 2019.

6.2 Structural Factors Eroding Journalistic Credibility in India

The credibility crisis in Indian digital journalism cannot be attributed exclusively to the disruptive effects of fake news and deepfakes; it is simultaneously driven by structural conditions within the media industry that have compromised institutional trust. Understanding these structural factors is essential to situating fake news and deepfakes within their proper analytical context.

Media ownership concentration represents a significant structural vulnerability. The Indian media industry has experienced accelerating consolidation since 2015, with major industrial and corporate groups including the Adani Group's acquisition of NDTV in 2022 and the continued expansion of Reliance-owned Network18 raising persistent questions about editorial independence and the capacity of commercial interests to shape journalistic priorities (Reporters Without Borders, 2023; Bhatia, 2022). When audiences perceive institutional journalism as compromised by ownership interests, their capacity to distinguish between editorially driven falsehood and externally generated misinformation is further weakened.

The proliferation of partisan news channels, operating under what media scholars have described as a 'paid news' ecosystem, has blurred the normative boundary between journalism and political communication in the Indian context. The term 'paid news' referring to the practice of publishing advertising content in the guise of editorial reporting was formally investigated by the Press Council of India as early as 2010, but the practice has persisted and evolved in the digital environment, where labeling conventions are less rigidly enforced (Press Council of India, 2010; Hasan, 2021).

The declining financial viability of independent journalism, driven by the migration of advertising revenue to digital platforms dominated by Google and Meta, has reduced the resources available to established news organizations for verification, investigative

reporting, and specialist expertise. This structural constraint has disproportionately affected regional-language journalism, which reaches the majority of Indian news consumers but operates with significantly lower resource levels than national English-language outlets (Ninan, 2021).

6.3 Impact of Fake News and Deepfakes on Trust: Evidence from India

Empirical evidence on the specific impact of fake news and deepfakes on journalistic credibility in India, while still developing, offers a consistent narrative of eroding trust and increased audience uncertainty.

The Reuters Institute Digital News Reports for the period 2019 to 2024 reveal a sustained decline in trust in news media among Indian audiences, with concerns about misinformation identified as among the primary drivers of this decline. In 2021, 78% of Indian respondents expressed concern about identifying what is real and what is fake on the internet that is the highest proportion recorded across the 46-country survey (Newman et al., 2021). By 2024, despite increased media literacy initiatives by civil society organizations and government bodies, this figure remained above 70%, suggesting that institutional responses have not kept pace with the scale of the challenge.

Research on the specific psychological mechanisms through which fake news erodes credibility draws on the concept of 'truth decay,' developed by the RAND Corporation (Kavanagh & Rich, 2018), which describes the erosion of the role of facts and analysis in public life. Adapting

this framework to the Indian context, Chandra and Mishra (2022) documented what they termed 'epistemic exhaustion' among regular news consumers- a condition in which the sustained effort required to verify information in a highly misinformational environment leads to a generalized withdrawal of trust from all media sources, including credible journalism.

Studies of the 2019 Indian general elections, conducted by researchers at the Centre for the Study of Developing Societies (CSDS) and the Lokniti network, found that exposure to partisan fake news on social media was significantly associated with reduced trust in both print and broadcast journalism, even when respondents were aware that the fake news originated from social media rather than professional news organizations (Lokniti-CSDS, 2019). This 'credibility contamination' effect whereby misinformation in one information environment undermines trust across the broader information ecosystem—represents a particularly destructive consequence of the fake news crisis.

The COVID-19 infodemic in India provided a real-time demonstration of these dynamics at catastrophic scale. Research by the WHO and Indian health communicators documented how pandemic-related misinformation including fabricated cures, false attribution of transmission vectors, and conspiracy theories about vaccine safety not only endangered public health but systematically undermined trust in official health communication and scientific journalism (Goel, 2020; Roozenbeek et al., 2020).

7. Conclusion

This paper has argued that the proliferation of fake news and the emergence of deepfake technologies have combined with pre-existing structural vulnerabilities in Indian journalism to produce a credibility crisis of significant depth and breadth between 2019 and 2024. The crisis is simultaneously epistemological challenging the foundations of shared factual reality and institutional, threatening the normative and organizational structures through which journalism performs its democratic function.

The theoretical analysis demonstrates that agenda-setting, framing, and post-truth frameworks each illuminate distinct aspects of this crisis. Together, they reveal fake news and deepfakes not as aberrant individual acts of deception but as structural forces reshaping the communicative ecology in which journalism operates. The systematic literature review confirms that Indian audiences have experienced measurable declines in news trust across the review period, that specific high-profile misinformation events have generated real-world harm, and that existing regulatory, institutional, and technological responses have been insufficient to arrest the trajectory of credibility erosion.

This paper has sought to contribute to these urgent questions through theoretical synthesis and critical analysis. Its primary contributions are the integrated analytical framework linking misinformation technologies to the credibility ecology of Indian journalism, the systematic mapping of the Indian fake news

and deepfakes landscape between 2019 and 2024, and the identification of the multi-dimensional character of effective response strategies.

The crisis of credibility in Indian digital journalism is ultimately a crisis of democratic communication. In a polity of India's scale and diversity, where journalism's function of holding power accountable and enabling informed citizenship is more consequential than in virtually any other context, the erosion of trust in news has implications that extend far beyond the media industry. Sustaining credibility in the age of illusion is not merely a professional challenge for journalists; it is a civic challenge for a democracy navigating the most consequential information environment in human history.

References

- Aldwaik, M. (2020). Mob lynching in India: Rumor, digital media, and the politics of violence. *South Asian History and Culture*, 11(4), 376–392.
- AltNews. (2023). Annual report on fact-checking in India 2022. AltNews Media Pvt. Ltd. <https://www.altnews.in>
- Banaji, S., & Bhat, R. (2019). WhatsApp vigilantes: An exploration of citizen reception and circulation of WhatsApp misinformation linked to mob violence in India. London School of Economics and Political Science.
- Bhatia, G. (2022). The shrinking space for independent journalism in India. *Economic and Political Weekly*, 57(18), 24–29.
- Bombay High Court. (2024). Kunal Kamra v. Union of India & Ors. Writ Petition No. 9792 of 2023. High Court of Judicature at Bombay.
- BOOM. (2019). Fake news report: 2019 general elections. BOOM Live. <https://www.boomlive.in>
- Chandra, A., & Mishra, P. (2022). Epistemic exhaustion and news avoidance in India's digital media environment. *Media, Culture & Society*, 44(3), 519–536.
- Chesney, R., & Citron, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review*, 107(6), 1753–1820.
- Committee to Protect Journalists (CPJ). (2022). India: AltNews co-founder Mohammed Zubair arrested. <https://cpj.org>
- Committee to Protect Journalists (CPJ). (2023). Attacks on women journalists in India: Trends and patterns 2019–2023. <https://cpj.org>
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Sage Publications.
- D'Ancona, M. (2017). *Post truth: The new war on truth and how to fight back*. Ebury Press.
- Denzin, N. K., & Lincoln, Y. S. (Eds.). (2011). *The Sage handbook of qualitative research* (4th ed.). Sage Publications.
- Dhawan, R. (2023). Deepfakes and the Indian political landscape: Emerging threats and regulatory responses. *Indian Journal of Law and Technology*, 19(2), 88–112.
- Digital Rights Foundation. (2022). Online violence against women journalists in South Asia. Digital Rights Foundation. <https://digitalrightsfoundation.pk>
- Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4), 51–58.
- Fink, A. (2020). *Conducting research literature reviews: From the internet to paper* (5th ed.). Sage Publications.

- Flanagin, A. J., & Metzger, M. J. (2007). *The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information*. *New Media & Society*, 9(2), 319–342.
- Gitlin, T. (1980). *The whole world is watching: Mass media in the making and unmaking of the New Left*. University of California Press.
- Goel, V. (2020). *Pandemic, infodemic, and the Indian media: Mapping misinformation during COVID-19*. *Journal of Content, Community and Communication*, 12(6), 208–222.
- Goffman, E. (1974). *Frame analysis: An essay on the organization of experience*. Harvard University Press.
- Hasan, Z. (2021). *Democracy and the crisis of inequality*. Sage Publications.
- Hovland, C. I., & Weiss, W. (1951). *The influence of source credibility on communication effectiveness*. *Public Opinion Quarterly*, 15(4), 635–650.
- Human Rights Watch. (2019). *Violent cow protection in India: Vigilante groups attack minorities*. Human Rights Watch. <https://hrw.org>
- Internet and Mobile Association of India (IAMAI). (2020). *Digital in India: Round 2*. IMAI. <https://www.iamai.in>
- Internet Freedom Foundation (IFF). (2020). *Infodemic: A report on COVID-19 misinformation in India*. IFF. <https://internetfreedom.in>
- Internet Freedom Foundation (IFF). (2021). *IT Rules 2021: Our analysis and concerns*. IFF. <https://internetfreedom.in>
- Internet Freedom Foundation (IFF). (2023). *IFF's analysis of the IT Amendment Rules 2023 on deepfakes*. IFF. <https://internetfreedom.in>
- Jeffrey, R., & Doron, A. (2021). *Cell-phone nation: How mobile phones have revolutionized business, politics and ordinary life in India (2nd ed.)*. Hachette India.
- Kavanagh, J., & Rich, M. D. (2018). *Truth decay: An initial exploration of the diminishing role of facts and analysis in American public life*. RAND Corporation.
- Keyes, R. (2004). *The post-truth era: Dishonesty and deception in contemporary life*. St. Martin's Press.
- Lokniti-CSDS. (2019). *Social media and political behaviour: Findings from the 2019 Indian National Election Study*. Centre for the Study of Developing Societies.
- McCombs, M. E., & Shaw, D. L. (1972). *The agenda-setting function of mass media*. *Public Opinion Quarterly*, 36(2), 176–187.
- McCroskey, J. C., & Teven, J. J. (1999). *Goodwill: A reexamination of the construct and its measurement*. *Communication Monographs*, 66(1), 90–103.
- McIntyre, L. (2018). *Post-truth*. MIT Press.
- Ministry of Electronics and Information Technology (MeitY). (2021). *The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021*. Government of India.
- Ministry of Electronics and Information Technology (MeitY). (2023). *Advisory to social media platforms on deepfakes*. Government of India. <https://meity.gov.in>

- MIT Technology Review. (2023). *The deepfake problem is getting worse. Here's why*. MIT Technology Review. <https://technologyreview.com>
- NASSCOM. (2023). *Responsible AI in India: Policy frameworks and industry practices*. National Association of Software and Service Companies.
- Nautyal, S., & Sharma, A. (2023). Audio deepfakes and political communication in India. *Telecommunications Policy*, 47(5), 102522.
- Network of Women in Media, India (NWMI). (2022). *State of media safety in India 2021–22*. NWMI. <https://nwmindia.org>
- Newman, N., Fletcher, R., Eddy, K., Robertson, C. T., & Nielsen, R. K. (2021). *Reuters Institute Digital News Report 2021*. Reuters Institute for the Study of Journalism, University of Oxford.
- Newman, N., Fletcher, R., Robertson, C. T., Eddy, K., & Nielsen, R. K. (2022). *Reuters Institute Digital News Report 2022*. Reuters Institute for the Study of Journalism, University of Oxford.
- Newman, N., Fletcher, R., Schulz, A., Andi, S., Robertson, C. T., & Nielsen, R. K. (2023). *Reuters Institute Digital News Report 2023*. Reuters Institute for the Study of Journalism, University of Oxford.
- Newman, N., Fletcher, R., Schulz, A., Robertson, C. T., & Nielsen, R. K. (2024). *Reuters Institute Digital News Report 2024*. Reuters Institute for the Study of Journalism, University of Oxford.
- Ninan, S. (2021). *The media in India: The fourth estate or the fourth pillar?* Oxford University Press.
- Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology: General*, 147(12), 1865–1880.
- Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, 188, 39–50.
- Press Council of India. (2010). *Paid news: How corruption in the Indian news media undermines democracy*. Press Council of India.
- Press Trust of India. (2023, November 7). Deepfake video of PM Modi triggers calls for stricter laws. *The Hindu*.
- Raghavan, P. (2022). *The disinformation machine: Political communication in digital India*. Penguin Random House India.
- Reporters Without Borders (RSF). (2022). *India: Press freedom index 2022*. RSF. <https://rsf.org>
- Reporters Without Borders (RSF). (2023). *India: Press freedom index 2023*. RSF. <https://rsf.org>
- Roozenbeek, J., Schneider, C. R., Dryhurst, S., Kerr, J., Freeman, A. L., Recchia, G., & van der Linden, S. (2020). Susceptibility to misinformation about COVID-19 across 26 countries. *Royal Society Open Science*, 7(10), 201199.
- Sharma, M., Singh, R., & Tripathi, N. (2021). *Misinformation during COVID-19 and elections: A study of the Indian digital media environment*. *Media Watch*, 12(3), 481–497.

- Snyder, H. (2019). *Literature review as a research methodology: An overview and guidelines*. *Journal of Business Research*, 104, 333–339.
- Statista. (2023). *Number of WhatsApp users in India 2017–2027*. Statista Research Department. <https://statista.com>
- Tandoc, E. C., Lim, Z. W., & Ling, R. (2018). Defining 'fake news': A typology of scholarly definitions. *Digital Journalism*, 6(2), 137–153.
- Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). Deepfakes and beyond: A survey of face manipulation and fake detection. *Information Fusion*, 64, 131–148.
- Torraco, R. J. (2016). Writing integrative literature reviews: Using the past and present to explore the future. *Human Resource Development Review*, 15(4), 404–428.
- TRAI. (2023). *Telecom subscription data: Annual report 2022–23*. Telecom Regulatory Authority of India. <https://traai.gov.in>
- Tripathy, S., & Srivastava, K. (2022). Evaluating media literacy interventions in rural India: Evidence from Rajasthan and Bihar. *Communicator: Journal of the IIMC*, 57(1), 44–61.
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151.
- Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policy making*. Council of Europe.