Journal
of the
Oriental Institute
M.S. University of Baroda
ISSN: 0030-5324
UGC CARE Group 1

Fake News, Deep Fakes, and Misinformation in Digital Media: Issues of Media Credibility

Dr. P. Shantha Bai

Assistant Professor,
Department of Mass Communication,
Telangana University, Nizamabad - 503322

Abstract:

The rapid and widespread convergence of misinformation, disinformation, and synthetic media is causing the digital media ecosystem to struggle with an "Information Disorder" that is undermining the fundamental purpose of informed public conversation. Platform economics, which put user engagement ahead of factual accuracy and take use of cognitive biases in people like in-group preference and emotional contagion, is what drives this phenomenon. It speeds up the viral spread of false narratives. By undermining the validity of visual and aural evidence, the development of generative artificial intelligence (AI), especially Deepfakes, poses a threat to political stability and judicial certainty and creates an epistemological crisis. The most severe manifestation of this issue is the sharp fall and political division of public confidence in news outlets, which enables dishonest individuals to take advantage of widespread suspicion by employing tactics like the "Liar's Dividend." A hybrid, multi-vector strategy that includes strong legal requirements for AI disclosure, significant funding for professional, scalable fact-checking projects, and thorough media literacy training is needed to mitigate these systemic dangers.

Keywords:

Disinformation, Deepfakes, Media Credibility, Information Disorder, Echo Chambers, Algorithmic Bias, Synthetic Media, Liar's Dividend, Epistemology, Computational Social Science.

Introduction: The Crisis of Digital Authenticity

- **Context and Problem Statement**

Journalism scholars refer to this setting as the "Information Disorder" because of the significant changes in the public realm brought about by the movement in information consumption toward decentralized, digital platforms. This condition encompasses sophisticated purposeful and inadvertent types of false content that are widely shared, going beyond simple factual errors. The modern media environment, characterized by previously unheard-of speed and accessibility, puts traditional journalistic institutions and other

established information gatekeepers to the test and poses serious threats to public health programs, democratic processes, and efforts to mitigate climate change worldwide. The systemic erosion of faith in authoritative organizations is a crucial and extremely harmful consequence of this widespread information pollution. For journalism, which is primarily responsible for supplying the common factual basis required for informed civic life and community cohesiveness, this fall is especially severe. The phrase "fake news," which is frequently employed in an ambiguous manner in public discourse, is unable to adequately convey the complexity of the issue. To shift conversations away from partisan rhetoric and toward practical policy solutions, it is imperative that concepts like "Information Disorder," as proposed by prominent journalism scholars, be rigorously adopted. Findings showing that incorrect information routinely travels more quickly on social networks than genuine information further highlight how urgent this problem is. As a result, truth is fundamentally at a structural disadvantage, and reactionary defences like verification and correction are inevitably involved in an expensive, trailing battle with content created with the goal of spreading as widely as possible. Instead of depending just on post-hoc rectification, this structural reality necessitates the development of proactive intervention models that emphasize preventive and platform accountability.

The Taxonomy of Problematic Information: Defining the Information Disorder

A clear, widely recognized nomenclature that separates harmful content according to the intention of the creator or sharer is necessary for the efficient management of the information disorder. The tripartite framework of misinformation, disinformation, and mal-information (MDM) is used in this paper.

- **Distinguishing Misinformation, Disinformation, and Mal-information (MDM)**

Misinformation (Unintentional Error)

Misinformation is characterized as incorrect or misleading information that is spread by someone who is not intentionally harmful but is merely misrepresenting the facts. Twelve Human error or a simple lack of verification are frequently at blame for this. A well-known politician might, for instance, post a news article with a false headline without realizing that the publisher had since fixed the mistake, which led to others spreading the original, inaccurate post. Correction, openness, and education are frequently the main focuses of the proper management plan because the sharer is not purposefully misleading the audience.

ISSN: 0030-5324
UGC CARE Group 1

Disinformation (Intentional Deception)

Disinformation, on the other hand, is intentionally misleading information that is either incorrect or erroneous. It is created, advertised, and promoted specifically with the intention of misleading people, making money, or hurting the general public. Disinformation is usually linked to well-planned operations and bad individuals looking to persuade others or stir up problems. The objectives are intentional and essentially detrimental.

Mal-information (Truth-Based Harm)

Because term refers to information that is founded on truth—it may be factually accurate—but is disseminated specifically with the intention of disparaging, harassing, or attacking a concept, person, group, or nation, mal-information inhabits a complicated space. The importance of misinformation stems from the fact that information integrity cannot be adequately measured by factual correctness alone. Digital media policy needs to go from evaluating content alone (Is this true?) to including context and intent analysis. Regulatory actions can be precisely targeted at malevolent Grade 2 and 3 actors thanks to the clear academic distinction between misinformation (error), disinformation (falsehood with malicious intent), and mal-information (truth with malicious intent).

- **Categorizing Information Disorder Syndromes (Typology and Grading)**

Different content typologies, which go beyond the MDM framework, explain the ways in which harmful material appears. These include: Fabricated Content; Manipulated Content; Impostor Content Misleading Content; and False Connection. Additionally, in order to manage the disorder, the players engaged must be graded according to their part and intent, which guides the proper response. Based on the actor's degree of complicity, the Information Disorder Syndrome grading system makes distinctions:

- **Grade 1 (The Sharer):** Most people in this group share content, frequently without knowing its accuracy or any repercussions (e.g., sharing a meme or rumour like "Garlic cures COVID-19"). To promote critical awareness, management uses psycho-social counselling and community involvement.
- **Grade 2/3 (The Originator/Capitalizer):** These people either create misinformation or have the ability to profit from its dissemination for their own gain or to purposefully hurt others (e.g., by participating in anti-vaxer campaigns or making unfounded charges). This level necessitates strict regulatory action and enforcement since malevolent intent is involved.

ISSN: 0030-5324
UGC CARE Group 1

The difference between these grades demonstrates a necessary policy intervention split: while intentional disinformation and mal-information require platform interventions and legal sanction because of the deliberate intent to cause harm, unintentional misinformation requires literacy and correction.

The Dual Nature of "Fake News"

Despite frequent criticism for its impreciseness, the phrase "fake news" endures in popular discourse with two separate dimensions:

1. The term "fake news" describes stuff that is purposefully presented to appear like authentic journalistic news, even if it may be false or disinformation.
2. "Fake News as a Label" refers to the use of the term itself as a rhetorical device to disparage or undermine the credibility of the real news media. By characterizing accurate reporting as politically driven lies, this labelling technique aims to undermine the press's fundamental, fact-gathering function, which is extremely damaging to media credibility.

Digital Architecture and Mechanisms of Information Spread

False content is spreading more widely than ever before, and this is a direct result of the way digital platforms are structured and how they interact with the psychological traits of people.

- **Algorithmic Amplification and the Engagement Economy**

A basic feature of digital media platforms is the structural tension between their business model and the general welfare. The main goal of social media algorithms is to increase user engagement, which is in direct opposition to the goal of disseminating factual information. By methodically giving preference to news that evokes strong emotional reactions, like fear and fury, over content that may be correct but elicits less visceral reaction, this engagement-focused business model inherently fosters divisiveness and disinformation. As a result of this design, it has been shown that fake news spreads on social networks far more quickly than authentic content. The polarized network structures that promote deceptive information are created and maintained by the purposeful exploitation of recognized human cognitive deficiencies in the goal of engagement profit. Platforms are therefore more than just impartial information channels; the main source of the Information Disorder is their well-designed architecture.

- **Psychological Susceptibility: Cognitive Biases and Emotional Drivers**

In the digital world, human cognition is intrinsically susceptible to manipulation. When people are presented with new information, they typically concentrate on comprehending it and choosing their next move instead of taking the time and effort to assess its accuracy. Because of this cognitive efficiency, those who spread misleading information have an advantage. One of the main factors influencing the propagation of false information is its emotional content. Knowing that messages that evoke strong emotions are the most widely shared, miscreants deliberately craft their messages to maximize emotional impact, avoiding critical evaluation. Furthermore, material that supports preexisting opinions and originates from sources that are seen as belonging to one's "in-group," taking advantage of confirmation bias and social identity, is more likely to be trusted. These effects are exacerbated by the Illusory Truth Effect, which is the tendency for people to trust information they have heard often, even if it clearly contradicts what they already know.

- **Network Effects: Polarization and the Formation of Homogeneous Echo Chambers**

The World Wide Web's abundance of user-generated material makes it easier for individuals to naturally come together around common interests, worldviews, and storylines. This tendency is replicated and made worse by algorithm-driven personalized feeds, which produce echo chambers—self-reinforcing informational settings. Selective exposure restricts exposure to outside perspectives inside these homogeneous groups, which results in group polarization—a process where belief sets are pushed to more extreme extremes and shared ideas are reinforced. This homogeneity and polarization within echo chambers are the primary factors utilized to forecast the final size and dissemination of rumour cascades, according to computational social science analysis. Its optimality for sharing within these insulated, already polarized social structures is directly related to the structural reality that highly emotive, novel, and deceptive material acquires a velocity advantage.

Deepfakes: The Technical Apex of Disinformation

The technological edge of deception is represented by deepfakes, which combine high-fidelity artificial intelligence with the capacity to produce incredibly lifelike but completely fake audio and video content.

- **Generative Adversarial Networks (GANs) and Synthetic Media Creation**

The main tool used to create deepfakes is Generative Adversarial Networks (GANs). Two neural networks make up this architecture: a discriminator that assesses the veracity of

the generated data and a generator that produces artificial data. Even skilled human observers find it more and more difficult to differentiate the produced deepfakes from real media as a result of this adversarial process, which continuously improves the quality of the created content. Political stability, social trust, and information integrity are all seriously threatened by the widespread use of these hyper realistic technologies, especially in underdeveloped areas where there is a lack of media literacy and technological infrastructure for fact-checking.

- **Challenges in Deepfake Detection and Verification**

Synthetic media detection is part of an ongoing technical arms race with creative technologies. Detection systems sometimes depend on the discriminator network's capacity to spot minute irregularities that the human eye frequently misses, like irregularities in pixel patterns, illumination, abnormal eye movements, or mismatched facial features. Self-supervised learning, which lessens the need for labelled data, temporal analysis, which tracks motion and timing irregularities, and hybrid models, which combine GANs with other machine learning approaches, are examples of emerging detection techniques. However, there are significant drawbacks to detection technologies, especially for experienced journalists who require prompt, conclusive results. These technologies usually yield results that are unclear or deceptive, which might cause more confusion than clarity. An over-reliance on these technologies can result in serious editorial mistakes, such as false positives that cause legitimate content to be unnecessarily rejected and false negatives that let sophisticated deepfakes go unnoticed. The difficulty is exacerbated by the labelling issue: labelling any content as even partially "artificial" can have a negative impact on how the public perceives its authenticity. For example, a real photo that has only been colour-corrected by AI could be mistakenly believed to be completely fake. A significant epistemological collapse is shown by the combination of the deepfake danger and technology detection flaws: if visual evidence can be flawlessly replicated and technological detectors are untrustworthy, society loses its common empirical anchor of truth.

- **Ethical and Legal Implications: Identity, Privacy, and Consent**

Synthetic media has far-reaching ethical and legal implications. Unauthorized use of personal information and likenesses to produce synthetic media raises serious ethical issues because it can result in identity theft, financial fraud, and serious harm to one's reputation. Reputable synthetic media companies try to enforce moral guidelines, such as forbidding malevolent creation and asking users to attest their assent, however dishonest individuals who violate rights can simply get over these restrictions. This illustrates the inadequacy of creators'

self-regulation and ethical standards, highlighting the need for legal requirements that put accountability not only on the creative tools but also on malevolent usage or intent to hurt. Deepfakes upend the conventional framework employed in the legal system to authenticate audio and video evidence in court, which usually depends on established chain-of-custody protocols and human perception. The fundamental presumption that juries can accurately assess the veracity of audiovisual evidence is called into question by the growing complexity of deepfakes. Because of this, legal experts have suggested extending the judicial gatekeeping power, which would require judges to assess the reliability of evidence using criteria akin to those used for complex technical evidence before it is shown to juries.

The Systemic Erosion of Media Credibility

A systemic crisis of trust has been triggered by the widespread inflow of fabricated and manipulated content, radically changing the public-media relationship.

- **Public Trust Trends and Partisan Polarization**

It is now very difficult for consumers and journalists to accurately assess the truth due to the introduction of highly realistic, phony information. Even when further stories are correct, the audience starts to seriously distrust media integrity after seeing even one example of distorted media presented as authentic. The audience's inclination to reject both false and authentic communications is a characteristic of their reaction to this ambiguity. The trend of eroding confidence is highly political, according to public opinion analysis. Partisan divisions have grown considerably as a result of a sharp drop in trust between particular political factions. For example, over a five-year period, the number of Republicans who said they had at least some faith in national news outlets was halved. According to research, readers' trust in the mainstream news is severely eroded as a result of increased exposure to online partisan media, rather than necessarily experiencing a significant change in political views. This successfully polarizes public trust and guarantees that various societal segments function using essentially disparate, frequently incompatible sets of knowledge.

- **The Weaponization of Doubt: The "Liar's Dividend"**

The establishment of widespread epistemological uncertainty, where the public defaults to skepticism regardless of factual correctness, is the most detrimental effect of the information disorder. The "Liar's Dividend" mechanism has made it possible to politically take advantage of this widespread mistrust. Politicians profit from deliberately and deceptively claiming that genuine scandal reports are "fake news" or "deepfakes," a tactic known as the Liar's Dividend. According to studies, these kinds of assertions effectively boost politician support across

political subgroups, regardless of whether they evoke informational confusion or inspire opposing rallying of core followers. It has been shown that this approach benefits the scandal-plagued politician more than other options like apologizing or keeping quiet. Political accountability is seriously threatened by this phenomenon because it gives dishonest actors a highly developed and efficient way to proactively reject indisputable textual or visual evidence.

- **Societal Impact on Democracy and Judicial Process**

This lack of confidence affects more than just political discourse. Hyper-realistic media content has the power to sway public opinion, jeopardize political stability, and spark upheaval in underdeveloped nations with weak technology infrastructure and fact-checking capabilities. Information integrity is compromised by deepfakes' capacity to create fake reality, which is especially important in democracies that are already in danger. There is also a fundamental redefining of the legal system. Enhanced burden requirements for video and audio evidence in high-stakes cases are among the practical strategies courts are creating to address the authenticity issues presented by synthetic media. A suggested change to the rules governing evidence, for instance, would require the proponent to show that the "probative value outweighs its prejudicial effect" of the evidence if a challenging party is successful in proving that a piece of digital information is most likely a deepfake. Additionally, there is a drive to mandate that the party contesting the evidence provide a threshold demonstration that the evidence may be artificial intelligence (AI) generated, such as expert testimony or a study of AI detection. While recognizing that the sophisticated nature of deepfake technology upends the traditional dependence on juror perception, this strict criterion is required to safeguard the legal system against baseless objections.

Multi-Vector Countermeasures and Policy Solutions

A coordinated, multi-vector approach combining legislative action, education, and technology intervention is needed to address the challenge of digital authenticity.

- **Fact-Checking: Efficacy, Scaling, and Epistemological Challenges**

In order to combat misleading information on social media and in the media, fact-checking has become a crucial aspect of journalism. According to research, fact-checks successfully mitigate one of the main negative effects of misinformation by reducing inaccurate beliefs over time. Subsequent research suggests that people are often eager to alter incorrect views when supplied with appropriate knowledge, which contradicts criticisms of the "backfire effect"—the idea that corrections reinforce faulty beliefs. According to experts, fact-checkers

shouldn't refrain from making changes because of broad worries about potential negative consequences.

But there is a sustainability conundrum facing the field. Major social media platforms like Meta and X have reduced their financial support and participation, placing a great deal of financial hardship on independent fact-checking organizations that depend on independent and rigorous investigative work. Independent organizations are forced to seek public action as a result of this divestment, exposing the systemic failure of depending on platforms' competing financial interests to maintain basic democratic information integrity. A hybrid model that combines professional journalists with automated scaling capabilities is considered necessary for speed and volume because non-professional approaches like crowdsourced or AI-driven fact-checking lack the necessary investigative rigor and critical thinking when used in isolation. Fact-checkers must also openly acknowledge subjective aspects of their work, such as presentation and selection criteria, and refrain from reproducing biases, especially when using AI-powered verification systems, in order to bolster their validity and respond to epistemological critiques.

- **Media Literacy and Critical Digital Citizenship**

It is commonly acknowledged that media literacy is an essential weapon in the fight against misinformation. The public is better able to evaluate information, recognize reliable sources, and make informed judgments when they are critically and digitally literate. It is a lifelong process that entails gaining knowledge and awareness of the ever-changing digital environment in addition to practical skills. Public engagement in formal media literacy training is still low, despite its significance; for example, 58% of Europeans questioned indicated interest in receiving training, but just 9% had done so. Policymakers must make sure that educational initiatives are planned to work in tandem with regulation, avoiding the trap of "burdening the citizen" with the sole responsibility of deciphering an often illegible and complex online world, even though 72% of Americans agree that media literacy skills are important for identifying misinformation.

- **Regulatory and Judicial Frameworks for Synthetic Media**

Legislative frameworks are being implemented by governments to combat the negative effects of synthetic media, especially in high-stakes domains like elections and private digital forgeries. The DEEPFAKES Accountability Act is one example of federal legislation in the US that attempts to safeguard national security and give victims legal remedies. According to the Protecting Consumers from Deceptive AI Act, the National Institute of Standards and

Technology should draft rules requiring disclosure for information produced or significantly altered by generative AI. Legislation at the state level frequently targets direct harms, such as making it illegal to create false movies meant to hurt a candidate or sway an election or making deepfakes featuring children having sex. Judicial rules are changing in the courtroom to control the veracity of digital evidence. The idea of requiring a challenger to demonstrate that the evidence is an AI-generated deepfake before demanding further evidence of the proponent's credibility is being discussed by courts. 5. This prevents baseless accusations from being made in every case using digital audio-visual evidence while acknowledging the complex nature of AI manipulation. This coordination is used in the most successful overall countermeasure strategy: Hybrid fact-checking offers quick, reliable correction (content management); literacy increases public resilience; and regulation targets deliberate harm and platform architecture.

Conclusion:

The confluence of deliberate malice (disinformation, mal-information, and deepfakes) and inadvertent transmission exacerbated by platform economic systems has resulted in the multilayered systemic issue known as the Information Disorder. False material secures its velocity advantage by taking use of well-known cognitive flaws in humans, such as the predilection for in-group bias and emotional validation over objective factual examination. With the emergence of deepfakes, the public's trust in audiovisual evidence has been undermined, and sophisticated political accountability avoidance through the "Liar's Dividend" has been made possible. Importantly, the removal of funding from the digital platforms that host the content poses a serious threat to the necessary scale and sustainability of proactive countermeasures like professional fact-checking, which have been shown to produce long-lasting reductions in false beliefs. This underscores the necessity for structural policy intervention.

References:

1. Allcott H., Gentzkow M. (2017). Social media and fake news in the 2016 election. Journal of Economic Perspectives, 31(2), 211–236.
2. Borji, A. (2023). Qualitative failures of image generation models and their application in detecting deepfakes. arXiv, 1.
3. Broda, E., and Strömbäck, J. (2024) “Misinformation, Disinformation, and Fake News: Lessons from an Interdisciplinary, Systematic Literature Review”, Annals of the International Communication Association, 48(2), pp. 139-166. DOI: 10.1080/23808985.2024.2323736

4. Chesney, B., & Citron, D. (2019). Deep fakes: a looming challenge for privacy, democracy, and national security. California Law Review, 107, 1753.
5. Danry, V., Leong, J., Pataranutaporn, P., Tandon, P., Liu, Y., Shilkrot, R., Punpongsanon, P., Weissman, T., Maes, P., & Sra, M. (2022). AI-Generated Characters: Putting Deepfakes to Good Use. CHI Conference on Human Factors in Computing Systems Extended Abstracts.URL: https://dl.acm.org/doi/10.1145/3491101.3503736
6. Gregory, S. (2021). Deepfakes, misinformation and disinformation and authenticity infrastructure responses: Impacts on frontline witnessing, distant witnessing, and civic journalism. Journalism, 23, 708-729. URL: https://journals.sagepub.com/doi/10.1177/14648849211060644
7. Gamage, D., Ghasiya, P., Bonagiri, V., Whiting, M. E., & Sasahara, K. (2022). Are deepfakes concerning? analysing conversations of deepfakes on reddit and exploring societal implications. Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22). Association for Computing Machinery, 1, 1–19
8. Sanghvi, B., Shelar, H., Pandey, M., & Sisodia, J. (2021). Detection of Machine Generated Multimedia Elements Using Deep Learning. 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), 1238-1243. URL: https://ieeexplore.ieee.org/document/9418008
9. Sundar, S. (2020). Rise of Machine Agency: A Framework for Studying the Psychology of Human-AI Interaction (HAII). Journal of Computer-Mediated Communication, 25, 74-88. URL: https://academic.oup.com/jcmc/article/25/1/74/5700811