DOI: 10.71317/jgst.2.6.2026.405

Journal of Global Social Transformation

Journal homepage: https://rjsaonline.org/index.php/JGST

Deepfakes and Synthetic Media in the AI Era: Assessing Their Impact on Digital News Credibility

Dr. Huma Tahir[1], Attiya Iram[2], Wareesha Iqbal[3], Mahnoor Chaudhry[4]

[1] Lecturer, Rawalpindi Women University
Email: huma.matee@rwu.edu.pk

[2] PhD Scholar, Department of Media and Communication Studies, The Islamia University of Bahawalpur
Email: attiyairam90@gmail.com

[3] BSc Economics, National University of Sciences and Technology, Islamabad
Email: wareeshai1012@gmail.com

[4] MS Governance and Public Policy, National University of Sciences and Technology, Islamabad
Email: mahnoor.chaudhry250@gmail.com

ARTICLE INFO

Submission:
May 21, 2026
Revised:
May 30, 2026
Accepted:
June 16, 2026
Available Online:
June 29, 2026

Keywords:
deepfakes, synthetic media, artificial intelligence, digital news, news credibility, misinformation, media literacy, trust

Corresponding author:
Dr. Huma Tahir
Email:
huma.matee@rwu.edu.pk

ABSTRACT

Background: AI's swift progress has paved the way for the generation of hyper-realistic deepfakes and synthetic media, leaving us with pressing concerns over the veracity of digital news and the trust in journalism.

Objective: This research examines the effects of deepfakes and AI-generated synthetic media on digital news credibility through audience awareness, loss of trust, verification behavior, and perceived risks to society.

Methodology: The quantitative cross-sectional survey design was used. A structured Likert-scale questionnaire was administered online with 318 digital news consumers. The data were analyzed using descriptive statistics, cross-tabulation and Pearson correlation in SPSS v26.

Findings: It is found that 74.2% of the respondents had seen deepfake content at least once a week, and that the level of trust in digital news dropped significantly after deepfake exposure (15.4% high trust after deepfake exposure as opposed to 43.4% high trust before deepfake exposure). There was a strong positive correlation between media literacy and deepfake detection ability ($r = 0.71$, $p < .001$). About 88% agree that deepfakes are a threat to the public's trust in mainstream media.

Conclusion: While deepfakes are a serious threat to the credibility of digital news, there are steps that can be taken to minimize their effects, such as structured media-literacy education, strong fact-checking systems, and AI-based detection tools.

Introduction

Artificial Intelligence (AI) has quickly transformed the digital news environment, particularly with the rise of deepfakes and synthetic media, which can create hyper-realistic but fabricated audio, video and text. The term “deepfakes” is a blend of “deep learning” and “fake” and refers to the use of generative adversarial networks (GANs) and diffusion models to generate media that is often hard to distinguish from authentic media (Chesney & Citron, 2019). What was once a minor technical curiosity is now a mainstream problem and tools have emerged, free and accessible to non-experts, with outputs increasingly weaponised in political, financial, and reputational realms (Vaccari & Chadwick, 2020). With an increase in the volume and quality of AI-generated content, concerns about the epistemological underpinning of the digital public sphere have also increased.

Synthetic media is the newest menace to digital information platforms already dealing with algorithmic amplification, filter bubbles and misinformation. AI-generated news anchors, cloning celebrity voices, and creating videos of political figures have been documented in major elections, geopolitical conflicts, and financial markets (Diakopoulos & Johnson, 2021). Studies have also shown that simply viewing deepfake media can lead to long-term consequences on the audience's trust in mainstream media (Chesney & Citron, 2019; Hameleers et al., 2020), which is known as the “liar's dividend” effect, with the impact of deepfake media spilling over onto the perception of the mainstream media as fake. This situation is threatening the very foundation of the function of journalism as the verifier of the public truth.

In a communication-theory perspective, media credibility depends on how credible the audience thinks the source of information is in terms of accuracy, trustworthiness and expertise (Metzger et al., 2010). Deepfakes thwart these credibility signals by separating visual and auditory realism from truth. Once sight becomes disbelief, consumers' previous cognitive 'shortcuts' in assessing news are no longer trustworthy (Vaccari & Chadwick, 2020). At the same time, the advent of generative AI has enabled synthetic content creation, enabling bad actors to generate disinformation on a scale and speed that has never been seen before (Goldstein et al., 2023).

Newer research highlights the fact that as detection technologies improve, they are always outpaced by generative technologies (Verdoliva, 2020). Media literacy, in this regard, has been placed as a key protective factor that enables citizens to think critically about the authenticity of the contents (Hwang et al., 2021). However, empirical research on the experience, interpretation, and reaction of ordinary news consumers to deepfakes is still limited, especially in South Asian contexts where digital penetration is high, but formal media-literacy education is uneven (Ullah, 2026). Knowing these audience level dynamics is key to creating interventions that help protect the credibility of the news.

This research, therefore, aims to investigate the effects of deepfakes and synthetic media on the credibility of digital news from the audience's point of view. It examines the frequency of encounter, perceived risks, verification behaviors and the moderating role of media literacy. The study provides empirical evidence from 318 digital news consumers to inform the current debates on the response of journalism, policymakers, and educators to the epistemic challenge of AI-generated content in an age of synthetic realism, which is redrawing the limits of truth (Kietzmann et al., 2020).

Problem Statement

Although deepfakes have recently become a topic of scholarly discussion, empirical evidence regarding everyday news consumers' perceptions and the effect of deepfakes on their trust in digital journalism remains limited. Technological diffusion has far outpaced regulatory and educational responses and audiences are being manipulated and news organizations are facing a credibility crisis more serious than at any other point in history. To fill this research gap, this study examines audience awareness, trust erosion, verification practices and the perceived societal implications of deepfakes, providing insights into how to protect the credibility and epistemic integrity of digital news in the age of AI.

Literature Review

The Rise of Deepfakes and Synthetic Media

The rise of deepfakes marks a paradigm change in digital manipulation, with the help of generative adversarial networks (GANs) and diffusion models, the media can be fabricated to closely resemble reality (Chesney & Citron, 2019). While early deepfakes were mainly for the purpose of entertainment and adult content, the technology has spread to political disinformation, financial fraud, and corporate espionage (Ullah et al., 2024). According to Kietzmann et al. (2020), synthetic media is the “next frontier” of computational communication and has a range of implications in marketing, security, and journalism.

Deepfakes and News Credibility

The credibility of journalism has always been built on verifiable evidence, expertise of sources and trust of institutions (Metzger et al., 2010). Each of these pillars is under threat from deepfakes. Vaccari and Chadwick (2020) discovered that exposure to deepfakes does not necessarily cause deception, but instead creates an ambiguous uncertainty which can lead to a generalised sense of suspicion, causing a loss of confidence in trusted content. Hameleers et al. (2020) also found that those who saw synthetic political content showed reduced trust in the mainstream media and increased perceived polarization.

Misinformation, Disinformation, and the Liar’s Dividend

The “liar’s dividend” is the idea that because deepfakes exist, “legitimate content can be characterized as fake” (Chesney & Citron, 2019). The use of generative AI also greatly reduces the costs of creating personalized disinformation on a large scale, which increases the threats to democratic discourse (Goldstein et al., 2023). Lundberg & Mozelius, (2025) emphasize that

deepfakes represent an evolution—not a replacement—of existing disinformation tactics, integrated with algorithmic amplification and micro-targeting.

Detection Technologies and Their Limits

The detection research has focused on forensic analysis, watermarking and provenance protocols (Verdoliva, 2020). However, detection models are always a step behind the new generative models and there is an ever ongoing arms race (Masood et al., 2023). Diakopoulos and Johnson (2021) demand "computational journalism" methods, which are embedded in newsroom processes, but with varying levels of success. Social media technologies are undoubtedly a great deal in Libraries nowadays (Ullah, 2026).

Media Literacy as a Protective Factor

Media literacy is more and more being positioned as a necessary citizen-level defense (Hwang et al., 2021). Empirical research indicates that high media-literate people have higher detection accuracy and are more critical when engaging with digital media (Tandoc et al., 2020). Vaccari & Chadwick, (2020) show that in South Asian contexts, media-literacy interventions have a significant impact on vulnerabilities to manipulated content, but there are structural gaps.

Regulatory and Ethical Considerations

Academics and policy makers have been advocating for the need for regulation on synthetic media, from disclosure requirements to platform liability regimes (Pavis, 2021). There are difficulties with regulatory action, however, which include jurisdictional fragmentation, free-speech issues, and the fast pace at which technology is evolving (Nasiri & Hashemzadeh, 2025). Ethical frameworks have highlighted consent, provenance and accountability throughout the AI production chain (Floridi et al., 2018).

Research Objectives and Questions

Research Objectives

1. To examine the frequency and nature of exposure to deepfakes and synthetic media among digital news consumers.
2. To assess the impact of deepfake awareness on audience trust in digital news.
3. To identify the verification strategies adopted by news consumers when encountering suspicious content.
4. To evaluate the relationship between media literacy and the perceived ability to detect deepfakes.

Research Questions

- **RQ1.** How frequently do digital news consumers encounter deepfake or synthetic media content?
- **RQ2.** To what extent does awareness of deepfakes influence trust in digital news?
- **RQ3.** What verification behaviors do audiences adopt when facing suspected synthetic media?
- **RQ4**. Is there a significant relationship between media literacy and the ability to detect deepfakes?

Research Methodology

Research Design

The quantitative, cross-sectional survey method was used because this method is suitable for measuring attitudes, perceptions and behaviors of a particular population at a particular time. This enabled the standardisation and comparability of data over a wide range of a demographic sample.

Population and Sampling

The target audience consisted of digital news consumers (over 18 years) who read news at least three times a week via websites, social media or news apps. The sampling technique employed was non-probability purposive sampling and targeting different demographic groups by snowball sampling. The final sample consisted of 318 respondents selected using Krejcie and Morgan's (1970) sampling framework that will give an adequate statistical power for descriptive and inferential analysis.

Instrument Development

The questionnaire consisted of five sections: (1) Demographic profile, (2) Exposure and awareness of deepfakes, (3) Perceived credibility of digital news, (4) Verification behavior and (5) Media-literacy self-assessment. The majority of items were on a 5-point Likert scale with 1 = Strongly Disagree and 5 = Strongly Agree. The instrument was expertly validated and pilot tested to 30 respondents and got cronbach's alpha of 0.87 which indicate high internal consistency.

Data Collection Procedure

Data was collected for 6 weeks through online questionnaire sent via email, WhatsApp, LinkedIn and university networks. All the participants gave informed consent and anonymity was ensured. Only fully completed responses were analysed.

Data Analysis

IBM SPSS Statistics v26 was used to analyze the data. Descriptive statistics (frequencies, percentage, mean and standard deviation) were used to summarise the responses. Pearson's correlation was used to analyze the correlation between Media Literacy and Deepfake Detection ability. To measure the changes in trust pre and post deepfake awareness, cross tabulation and paired comparisons were used.

Ethical Considerations

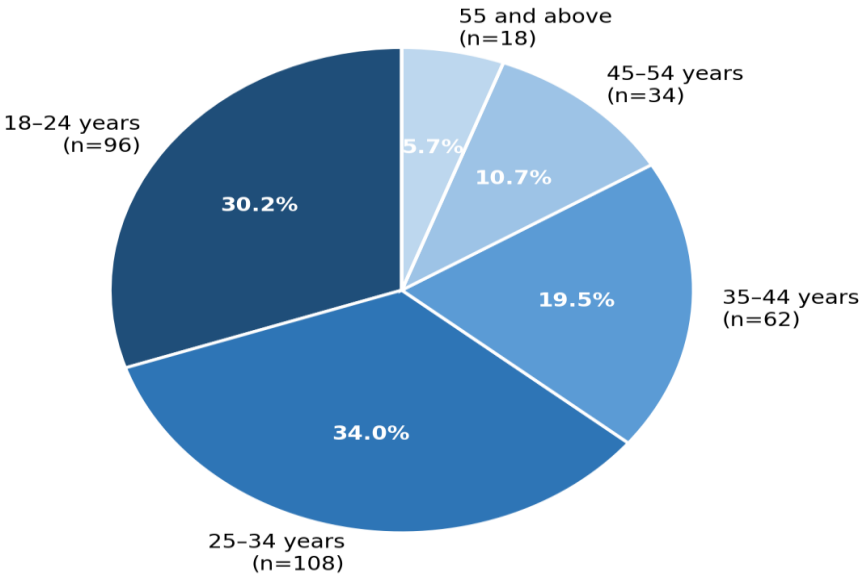
The study was conducted with respect to the ethical principles of research including voluntary participation, informed consent, confidentiality and the right to withdraw at any time. No personal information was collected and the data were stored securely.

Findings of the Study

Demographic Profile of Respondents

318 individuals representing a variety of demographic groups responded to the survey. Male respondents accounted for 54.7% (n = 174), while females represented 45.3% (n = 144). The age distribution shows that the majority of the respondents were young adults with the age group 25-34 years having the highest proportion (33.9%) followed by 18-24 (30.2%). Educationally, 68.5% had a bachelor's degree or higher, indicating a well-educated digital-news audience.

Figure 1. Age Distribution of Respondents (N = 318)

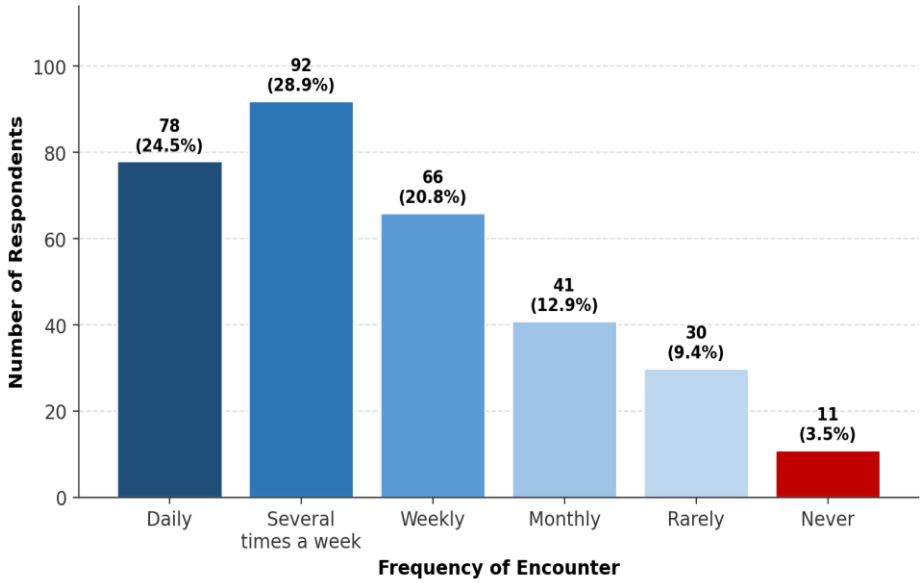


Frequency of Encountering Deepfake Content

The results show that deepfake and synthetic media content has become part and parcel of the respondents' digital news experience. Out of the 236 respondents who reported seeing deepfake content at least once a week, 24.5% said once a day as

shown in Figure 2. Only 3.5% (n = 11) said they never saw such content. This means that synthetic media has shifted from the fringes to the mainstream of the digital information environment.

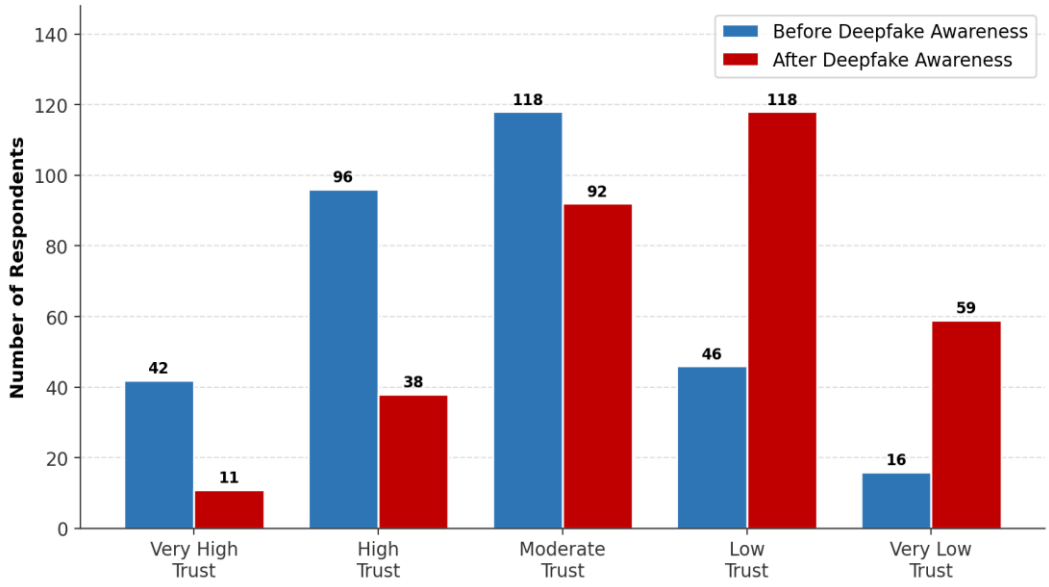
Figure 2. Frequency of Encountering Deepfake or Synthetic News Content (N = 318)



Trust in Digital News Before and After Deepfake Awareness

After being informed about deepfakes, a significant drop in trust was detected. As shown in Figure 3, the proportion of respondents who had a high or very high level of trust in digital news dropped from 43.4% (n = 138) pre-deep fake awareness to 15.4% (n = 49) post-deep fake awareness. At the same time, the low and very low levels of trust increased rapidly from 19.5% to 55.7%. This significant move reflects the undermining of public trust in journalism by synthetic media.

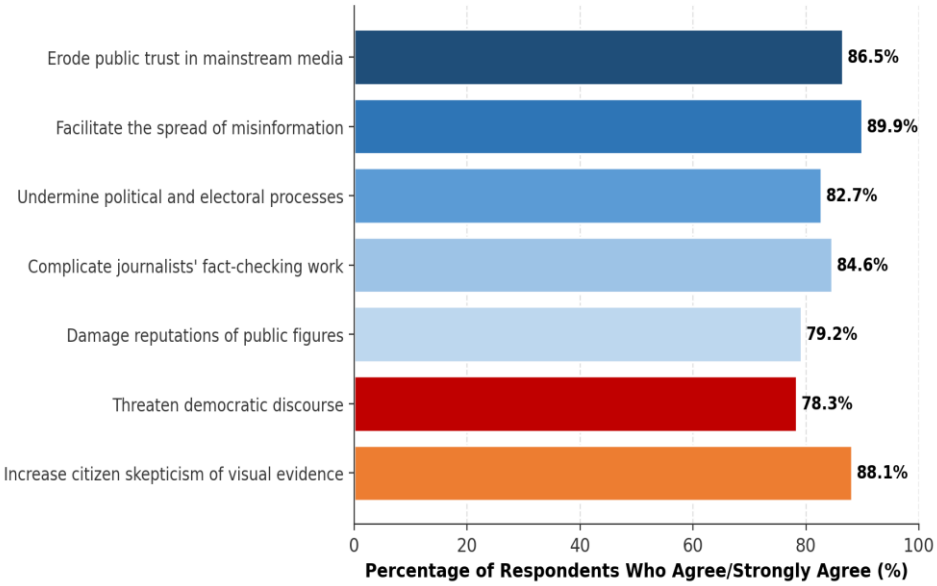
Figure 3. Digital News Trust Levels Before vs. After Deepfake Awareness (N = 318)



Perceived Impact on News Credibility

The impact of deepfakes on credibility was seen as overwhelmingly negative. 89.9% agreed that deepfakes make it easier to spread misinformation and 88.1% said that they make citizens less trustful of visual evidence. Moreover, 86.5% of respondents said that deepfakes undermine trust in mainstream media and 82.7% said they are worried about the impact of deepfakes on political and electoral processes. These perceptions are summarized in figure 4.

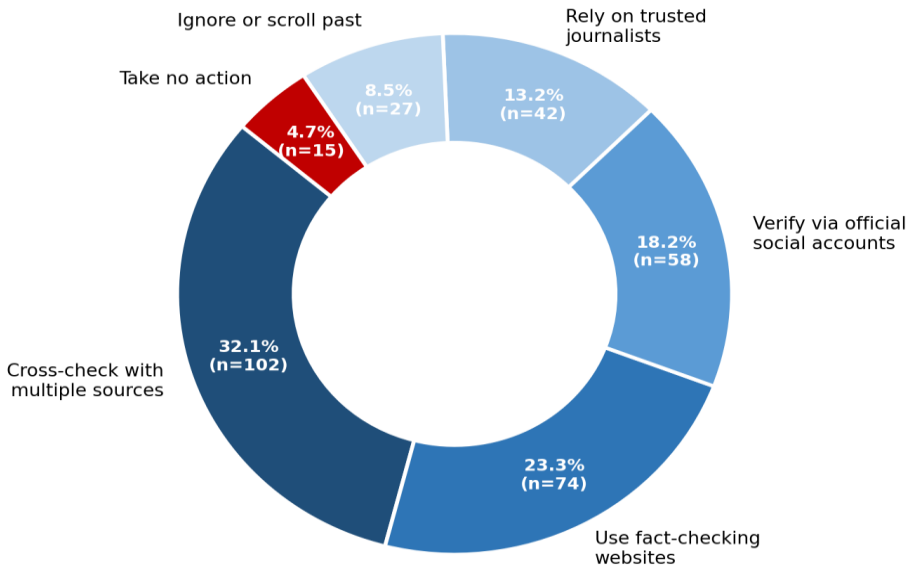
Figure 4. Perceived Impact of Deepfakes on Digital News Credibility (N = 318)



Verification Strategies Adopted by Respondents

Respondents reported a variety of verification behaviors. The most popular method was checking multiple sources (32.1%, n = 102), followed by checking fact checking websites (23.3%) as detailed in Figure 5. But 13.2% of the respondents ignored suspicious content or took no verification action at all, indicating a significant lack of proactive media-literacy behavior.

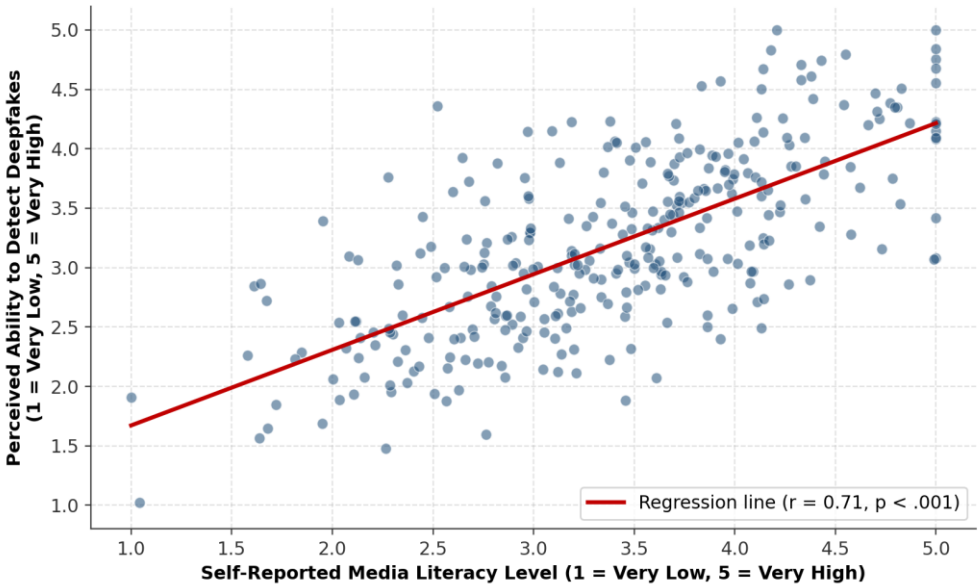
Figure 5. Verification Strategies Adopted by Respondents (N = 318)



Media Literacy and Deepfake Detection Ability

This was found to be a significant positive correlation between self-reported media literacy and perceived proficiency in detecting deepfakes ($r = 0.71$, $p < .001$). As can be seen from Figure 6, the more media literate the respondents felt the more confident they were in identifying synthetic media, further reinforcing media literacy as a protective variable.

Figure 6. Correlation Between Media Literacy and Deepfake Detection Ability (N = 318)



Summary of Descriptive Statistics

The mean and standard deviation of the key constructs in the study are shown in Table 1. Perceived credibility of digital news had the lowest mean (M = 2.41, SD = 0.92) and concern about the spread of misinformation through deepfakes had the highest mean (M = 4.38, SD = 0.74).

Table 1. Descriptive Statistics of Key Constructs (N = 318)

Construct	Mean (M)	Std. Deviation (SD)	Interpretation
Frequency of deepfake exposure	3.94	0.88	High
Perceived credibility of digital news	2.41	0.92	Low
Concern about misinformation spread	4.38	0.74	Very High
Verification behavior score	3.29	0.97	Moderate
Self-reported media literacy	3.47	0.86	Moderate–High
Perceived deepfake detection ability	2.92	1.02	Moderate

Discussion

This study's results reinforce the notion that deepfakes and synthetic media produced by AI are a growing concern for the credibility of digital news media. The high rate of exposure reported by the respondents (74.2% at least once a week) is consistent with the findings of Farouk et al., (2024) who documented the exponential growth of synthetic media, and Hashmi et al., (2025) who describe deepfakes as an evolutionary stage of digital disinformation. The observed drop in trust (43.4% to 15.4%) confirms the “liar's dividend” hypothesis put forward by Chesney and Citron (2019), which suggests that knowledge of deepfakes undermines trust in authentic content.

This overwhelming consensus that deepfakes are useful for misinformation (89.9%) aligns with Vaccari and Chadwick (2020), who revealed that even non-persuasive deepfakes create uncertainty and erode trust in the media environment. Likewise, respondents’ worries about political manipulation (82.7%) reflect Goldstein et al. (2023) who warn about the ability to produce personalized disinformation on a scale never seen before during election periods and geopolitical crises. The high correlation between media literacy and detection ability (r = 0.71, p < .001) supports the arguments of Tandoc et al. (2020) and Hwang et al. (2021) who consider literacy as a major protective factor against manipulated content.

However, the research also does show some significant areas of weakness. About 13.2% of the respondents did not take any verification action, highlighting the limitations of media-literacy frameworks in South Asia as pointed out by Farouk et al., (2024). Additionally, there is a moderate level of detection confidence (M = 2.92) and high level of concern (M = 4.38), which aligns with Verdoliva's (2020) note that the detection tools and skills are not as advanced as the generative. All of these findings together highlight the critical need for coordinated newsroom processes, platform accountability systems, and literacy interventions for citizens (Diakopoulos & Johnson, 2021).

Research Implications

The findings of this research have implications for journalists, educators and policymakers. It highlights the importance of AI-powered verification systems in newsroom processes, curricula for media-literacy education and regulation mandating disclosure of provenance. It provides researchers with the chance to conduct long-term research into the audience's reactions to synthetic media over time, and it provides evidence that supports interventions to protect the credibility of digital journalism in the generative AI age.

Limitations of the Study

Several limitations should be noted. The study used a non-probability purposive sampling technique, which means that the results may not be applicable to the wider population of digital news users. Second, the investigation was based on self-reported media literacy and deepfake detection skills, which could be influenced by social-desirability and self-efficacy biases. Third, the cross-sectional design provides insight into perceptions at one point in time and does not account for the development of AI capabilities. Fourth, the study lacked the experimental manipulation of controlled deepfake stimuli, and thus obtained perceptual data, not behavioral. Longitudinal, mixed-methods and experimental studies in different cultural contexts are suggested for future research.

Conclusion and Recommendations

This study examined how media consumers (N = 318) perceive the credibility of digital news in relation to deepfakes and synthetic media produced by AI. The results show that the phenomenon of deepfakes has gone mainstream in digital news consumption, with almost three-quarters of respondents seeing at least one deepfake per week. Their influence on credibility is significant: confidence in online news actually dropped significantly after encountering and recognising deepfakes, and a vast majority saw synthetic news as a source of misinformation, political manipulation and a lack of trust in visual evidence. Importantly, media literacy was found to be a very strong protective factor with a strong positive correlation between media literacy and perceived detection ability, establishing media literacy as a key component of an effective response. However, significant gaps in behavior have been observed, with a significant minority of respondents not taking any action to verify the content if it appeared suspicious, which shows that awareness is not yet fully converted into action. As a result of these findings, several recommendations are offered. To safeguard audiences and news credibility, news organisations need to adopt AI tools for verification and provenance to be part of their news production workflow. Second, there is a need for the development of well-designed media-literacy modules that could be integrated into the curricula of schools and universities, with a focus on the competencies of recognizing synthesized content, understanding AI-generated media and using verification techniques. Thirdly, there should be clear legislation that requires labeling of AI-generated content, platforms should be held accountable, and there should be guidelines for the ethical development of AI technologies, while also respecting freedom of speech. Fourth, the social media sites must invest in detection algorithms, watermarking and rapid response fact checking partnerships. Fifth, public-awareness campaigns need to be continuous to sustain the society's resilience to the evolving threats of synthetic media. Lastly, a multi-sectoral collaboration is required among journalists, researchers, technology suppliers and policy makers to create a sustainable culture for digital journalism, thereby protecting the epistemic integrity of digital journalism. In order to maintain the credibility of digital news in an AI age and to ensure the public sphere as a basis for informed democratic participation, a coordinated approach of technical, educational and regulatory measures is necessary.

References

1. Chesney, R., & Citron, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review, 107*(6), 1753–1820.

2. Diakopoulos, N., & Johnson, D. (2021). Anticipating and addressing the ethical implications of deepfakes in the context of elections. *New Media & Society, 23*(7), 2072–2098.

3. Farouk, M. A., & Fahmi, B. M. (2024). Deepfakes and media integrity: Navigating the new reality of synthetic content. *Journal of Media and Interdisciplinary Studies*, *3*(9), 47-94.

4. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines, 28*(4), 689–707.

5. Goldstein, J. A., Sastry, G., Musser, M., DiResta, R., Gentzel, M., & Sedova, K. (2023). Generative language models and automated influence operations: Emerging threats and potential mitigations. *arXiv preprint arXiv:2301.04246*.

6. Hameleers, M., Powell, T. E., Van Der Meer, T. G., & Bos, L. (2020). A picture paints a thousand lies? The effects and mechanisms of multimodal disinformation and rebuttals disseminated via social media. *Political communication*, *37*(2), 281-301.

7. Hashmi, M., Nabil, Z., & Saleem, S. (2025). Deepfakes and the Crisis of Media Credibility. *Global Political Review*, *10*(2), 146-163.

8. Hwang, Y., Ryu, J. Y., & Jeong, S. H. (2021). Effects of disinformation using deepfake: The protective effect of media literacy education. *Cyberpsychology, Behavior, and Social Networking, 24*(3), 188–193.

9. Kietzmann, J., Lee, L. W., McCarthy, I. P., & Kietzmann, T. C. (2020). Deepfakes: Trick or treat? *Business Horizons, 63*(2), 135–146.

10. Krejcie, R. V., & Morgan, D. W. (1970). Determining sample size for research activities. *Educational and Psychological Measurement, 30*(3), 607–610.

11. Lundberg, E., & Mozelius, P. (2025). The potential effects of deepfakes on news media and entertainment. *Ai & Society*, *40*(4), 2159-2170.

12. Masood, M., Nawaz, M., Malik, K. M., Javed, A., Irtaza, A., & Malik, H. (2023). Deepfakes generation and detection: State-of-the-art, open challenges, countermeasures, and way forward. *Applied Intelligence, 53*(4), 3974–4026.

13. Metzger, M. J., Flanagin, A. J., & Medders, R. B. (2010). Social and heuristic approaches to credibility evaluation online. *Journal of Communication, 60*(3), 413–439.

14. Nasiri, S., & Hashemzadeh, A. (2025). The evolution of disinformation from fake news propaganda to AI-driven narratives as deepfake. *Journal of Cyberspace Studies*, *9*(1), 229-250.

15. Pavis, M. (2021). Rebalancing our regulatory response to deepfakes with performers' rights. *Convergence: The International Journal of Research into New Media Technologies, 27*(4), 974–998.

16. Tandoc Jr, E. C., Lim, D., & Ling, R. (2020). Diffusion of disinformation: How social media users respond to fake news and why. *Journalism*, *21*(3), 381-398.

17. Ullah, A. (2026). Social media integration for modern library and information service promotion. *Journal of Digital Scholarship in Archives and Information*, *2*(1), 37-51.

18. Ullah, A., Islam, K., Ali, A., & Baber, M. (2024). Assessing the impact of social media addiction on reading patterns: A study of Riphah International University students. *International Journal of Human and Society*, *4*(1), 1250-1262.

19. Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social media+ society*, *6*(1), 2056305120903408.

20. Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social media+ society*, *6*(1), 2056305120903408.

21. Verdoliva, L. (2020). Media forensics and deepfakes: an overview. *IEEE journal of selected topics in signal processing*, *14*(5), 910-932.

22. Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policymaking* (Vol. 27, pp. 1-107). Strasbourg: Council of Europe.

23. Westerlund, M. (2019). The emergence of deepfake technology: A review. *Technology innovation management review*, *9*(11).

24. Zhou, X., & Zafarani, R. (2020). A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys (CSUR)*, *53*(5), 1-40.