



ARTIFICIAL INTELLIGENCE, DEEPFAKES, AND THE MISINFORMATION CHALLENGE JOURNALISTS' PERSPECTIVES FROM PAKISTAN

Hafiz Muhammad Muddser

hafizmuhammadmuddassir642@gmail.com

PhD Scholar, Department of Media and Communication
Studies, The Islamia University of Bahawalpur, Pakistan.

Awais H. Gillani

ahgillani@iub.edu.pk

Assistant Professor, Department of Media and
Communication Studies, The Islamia University of
Bahawalpur, Pakistan.

Abstract

The proliferation of deepfakes and AI-generated misinformation has introduced unprecedented ethical and operational challenges for contemporary journalism. This study investigates how Pakistani journalists perceive the impact of deepfakes and AI-assisted misinformation on news credibility, verification practices, and audience trust. Drawing on survey data from 387 journalists across print, broadcast, and digital media, the study employs quantitative analysis using Likert-scale measures to assess perceptions of risk, ethical concerns, and professional safeguards. Reliability of the survey instrument was confirmed via Cronbach's alpha ($\alpha = 0.87$), indicating strong internal consistency. Findings reveal that journalists are acutely aware of the threats posed by deepfakes and misinformation, reporting heightened challenges in verifying sources, maintaining accuracy, and preventing the dissemination of fabricated content. Moreover, respondents emphasize the necessity of ethical safeguards, including disclosure of AI-assisted content and rigorous editorial oversight. The study highlights the growing gap between technological innovation and existing verification practices, suggesting an urgent need for targeted training, institutional policies, and regulatory frameworks. By foregrounding journalists' perspectives, the research contributes to the global discourse on AI-driven misinformation and ethical journalism in developing media ecosystems.

Keywords: *Deepfakes, AI-Generated Misinformation, Journalism Ethics, News Credibility, Content Verification, Editorial Oversight, Pakistani Newsrooms, AI In Media*

Corresponding Author: Hafiz Muhammad Muddser (PhD Scholar, Department of Media and Communication Studies, The Islamia University of Bahawalpur, Pakistan).

Email: hafizmuhammadmuddassir642@gmail.com

1. Introduction

The rapid advancement of Artificial Intelligence (AI) has significantly altered the contemporary information environment, particularly through the emergence of synthetic media and deepfake technologies. AI-driven tools capable of generating realistic text, images, audio, and video have intensified the global challenge of misinformation, posing serious risks to journalism, democratic discourse, and public trust. For news organizations, the proliferation of AI-generated misinformation represents not only a technological challenge but also a fundamental threat to the credibility and authority of professional journalism.

Deepfake technologies, which use machine learning algorithms to manipulate or fabricate audio-visual content, have gained increasing attention due to their potential misuse in political communication, social manipulation, and reputational harm. Unlike traditional forms of misinformation, deepfakes are often difficult to detect with the naked eye, making verification more complex and time-consuming for journalists. As a result, newsrooms are confronted with unprecedented challenges in distinguishing authentic content from AI-generated fabrications while maintaining the speed and accuracy expected in the digital news cycle.

Existing research highlights that AI has become both a driver and an amplifier of misinformation. Algorithmic content generation and automated dissemination systems can accelerate the spread of false information across digital platforms, often outpacing journalistic fact-checking mechanisms (Vosoughi et al., 2018; Tandoc et al., 2018). While scholars have examined the political and societal implications of deepfakes, empirical research focusing on journalists' lived experiences with AI-driven misinformation remains limited, particularly in developing media systems.

In Pakistan, the misinformation challenge is intensified by high social media penetration, political polarization, and limited institutional capacity for digital verification. Journalists operate in an environment where viral misinformation can rapidly influence public opinion, yet technological resources for detecting deepfakes and synthetic media are unevenly distributed across news organizations. Despite these challenges, systematic empirical evidence capturing Pakistani journalists' perceptions of AI-driven misinformation risks is scarce.

This study addresses this gap by examining journalists' perceptions of AI-generated misinformation and deepfake threats in Pakistan. Using survey data collected from journalists working across print, electronic, and digital media, the study explores how media professionals assess the severity of AI-driven misinformation, the challenges it poses to verification practices, and its implications for journalistic authority. By centering journalists' perspectives, this research contributes to broader debates on information

disorder and highlights the need for technological and institutional responses to AI-enabled misinformation.

1.1. Research Questions

- RQ1: How do Pakistani journalists perceive the threat of AI-generated deepfakes to journalism?
- RQ2: To what extent do journalists believe AI accelerates the spread of misinformation and fake news?
- RQ3: What challenges do journalists face in verifying AI-generated or manipulated content?
- RQ4: How do journalists assess newsroom preparedness to counter AI-driven misinformation?

2. Literature Review

2.1. Artificial Intelligence and the Rise of Synthetic Media

The emergence of Artificial Intelligence has accelerated the production of synthetic media, fundamentally reshaping the contemporary information ecosystem. AI-driven systems—particularly those based on deep learning and generative adversarial networks (GANs)—enable the creation of highly realistic text, images, audio, and video content that is often indistinguishable from authentic media. While these technologies have legitimate applications in journalism and media production, scholars argue that their misuse has significantly intensified the global misinformation crisis (Floridi et al., 2018; Diakopoulos, 2019).

Synthetic media challenges traditional epistemological assumptions in journalism by undermining visual and textual evidence as reliable indicators of truth. Historically, photographic and video materials have served as powerful forms of journalistic verification. However, AI-generated content erodes this evidentiary value, forcing journalists to rethink verification norms and credibility frameworks (Gillespie, 2014; Carlson, 2015).

2.2. Deepfakes as a New Form of Misinformation

Deepfakes represent one of the most disruptive manifestations of AI-generated misinformation. Defined as synthetic audio-visual content created through machine learning techniques to falsely depict individuals saying or doing things they never did, deepfakes pose unique threats to journalism and democratic discourse (Chesney & Citron, 2019). Unlike conventional misinformation, deepfakes exploit human cognitive biases and visual trust, making them particularly persuasive and difficult to debunk.

Research indicates that deepfakes can be weaponized for political manipulation, character assassination, and social destabilization (Vaccari & Chadwick, 2020). Their potential impact extends beyond misinformation, threatening the legitimacy of authentic

media by fostering a “liar’s dividend,” where genuine evidence can be dismissed as fake (Chesney & Citron, 2019). This phenomenon significantly complicates journalistic verification and accountability.

2.3. AI, Misinformation, and Algorithmic Amplification

AI contributes to misinformation not only through content generation but also through algorithmic amplification. Automated recommendation systems and social media algorithms prioritize engagement-driven content, often facilitating the rapid spread of sensational or false information (Vosoughi et al., 2018; Napoli, 2019). Studies show that false news spreads faster and more widely than verified information, placing journalists in a reactive position where verification lags behind viral misinformation.

Tandoc et al. (2018) argue that the structural design of digital platforms amplifies misinformation by privileging speed and virality over accuracy. For journalists, this creates a paradox: while AI-driven tools can assist in news production, the same technologies undermine journalists’ ability to control information flows and uphold professional standards.

2.4. Verification Challenges and Journalistic Authority

The rise of AI-generated misinformation has significantly increased the complexity of verification practices in journalism. Traditional fact-checking methods are often insufficient to detect sophisticated deepfakes, requiring advanced technical tools and interdisciplinary expertise (Silverman, 2015; Graefe, 2016). However, access to such resources remains uneven, particularly in developing media systems.

Scholars emphasize that journalists’ authority as gatekeepers is increasingly contested in an environment where audiences encounter misinformation directly through digital platforms (Lewis et al., 2019). The growing reliance on AI detection tools introduces new ethical and professional dilemmas, including overdependence on automated verification systems and reduced human judgment.

2.5. Misinformation, Trust, and Audience Perceptions

AI-driven misinformation has profound implications for public trust in journalism. Studies suggest that repeated exposure to manipulated or false content erodes audience confidence in news media, even when misinformation is later corrected (Waisbord, 2018). Deepfakes exacerbate this erosion by creating uncertainty about the authenticity of all media content, not just fabricated materials.

Research also indicates that audiences often struggle to distinguish between verified news and AI-generated misinformation, placing greater responsibility on journalists to act as credibility anchors (Newman, 2023). Failure to effectively counter misinformation risks long-term damage to journalistic legitimacy and democratic engagement.

2.6. Misinformation and Journalism in Developing Media Systems

Most existing literature on AI-driven misinformation focuses on Western contexts, leaving a critical gap in research on developing media systems. In countries like Pakistan, journalists operate within environments characterized by political polarization, limited regulatory oversight, and resource constraints. These conditions amplify the risks posed by AI-generated misinformation while limiting institutional capacity for detection and response.

Emerging studies suggest that journalists in developing countries face disproportionate challenges in combating digital misinformation due to limited access to verification technologies and professional training (Beckett, 2019; Marconi, 2020). However, empirical evidence capturing journalists' perceptions of deepfakes and AI-driven misinformation in such contexts remains scarce.

2.7. Research Gap

Despite growing scholarly attention to AI, deepfakes, and misinformation, limited empirical research examines how journalists themselves perceive and experience these challenges, particularly in non-Western media systems. Existing studies largely focus on technological capabilities or audience effects, overlooking newsroom-level perspectives and professional concerns.

This study addresses this gap by empirically investigating Pakistani journalists' perceptions of AI-generated misinformation and deepfake threats. By foregrounding journalists' voices, the research contributes to a more nuanced understanding of how AI-driven misinformation reshapes journalistic practice, authority, and credibility in developing media environments.

3. Methodology

3.1. Research Design

This study adopted a quantitative cross-sectional survey design to examine journalists' perceptions of AI-generated deepfakes and misinformation and their implications for contemporary journalism. A quantitative approach was considered appropriate to capture measurable attitudes and trends across a large professional sample.

3.2. Population and Sample

The target population consisted of working journalists employed in print, electronic, and digital media organizations. A random sampling technique was used to collect data from journalists based in Islamabad, Lahore, and Karachi, representing Pakistan's major media hubs. A total of 386 valid responses were included in the final analysis, ensuring adequate statistical representation.

3.3. Instrumentation

Data were collected using a structured questionnaire developed based on existing literature on AI, deepfakes, misinformation, and media ethics. The instrument employed a five-point Likert scale ranging from *Strongly Disagree (1)* to *Strongly Agree (5)*. The

questionnaire measured key variables such as perceptions of deepfakes, fake news proliferation, ethical concerns, and the need for human editorial oversight.

3.4. Data Collection

The survey was administered using Google Forms, allowing efficient and anonymous data collection. Participation was voluntary, and respondents were informed about the academic purpose of the study.

A total of 386 valid responses were obtained using random sampling techniques, meeting the recommended sample size for quantitative social research. Data collection was conducted both online and in person to maximize response diversity and participation. Ethical considerations were strictly observed; participation was voluntary, anonymity was assured, and informed consent was obtained from all respondents.

The collected data were analyzed using Statistical Package for the Social Sciences (SPSS). Descriptive statistical techniques, including frequencies, percentages, means, and standard deviations, were employed to summarize respondents' perceptions. The findings were presented through tabular representations and bar charts to facilitate interpretation. This methodological approach enabled a systematic examination of how AI-driven misinformation and deepfakes are perceived within contemporary journalistic practice.

Collected data were analyzed using descriptive statistical techniques, including frequencies, percentages, and bar charts, to illustrate response patterns. The results were presented visually to enhance clarity and interpretability.

3.5. Ethical Considerations

Ethical principles were strictly followed. Respondents' identities were kept confidential, informed consent was obtained, and data were used solely for research purposes.

3.6. Data Analysis

The collected survey data were coded and analyzed using the Statistical Package for the Social Sciences (SPSS). Prior to analysis, the dataset was screened for missing values, inconsistencies, and outliers to ensure accuracy and reliability. All responses were found to be complete and suitable for statistical analysis.

Descriptive statistical techniques were primarily employed to analyze journalists' perceptions regarding AI-driven misinformation and deepfake technologies. Frequencies and percentages were used to summarize demographic characteristics and response distributions across Likert-scale items. Measures of central tendency, including means and standard deviations, were calculated to assess the overall level of agreement or disagreement with statements related to deepfakes, fake news production, ethical concerns, and the erosion of journalistic credibility.

To enhance interpretability, results were visually represented through bar charts illustrating respondents' attitudes toward the perceived seriousness of deepfakes, the role

of AI in facilitating misinformation, and the importance of human editorial oversight. These visualizations allowed for clearer comparison of response patterns across key variables.

The analysis focused on identifying dominant trends and perceptual patterns rather than causal relationships, in line with the study's exploratory objectives. The findings were interpreted in relation to existing literature on AI ethics, misinformation, and journalism studies to contextualize journalists' concerns within broader global debates. This analytical approach provided empirical insight into how emerging AI technologies are reshaping trust, responsibility, and ethical norms in contemporary journalism.

Descriptive statistics were employed to summarize journalists' perceptions of AI-driven misinformation and deepfake technologies. Frequencies and percentages were calculated for demographic variables and Likert-scale items, while means and standard deviations assessed the overall levels of agreement or disagreement with statements on deepfakes, fake news, ethical concerns, and journalistic credibility.

To evaluate the internal consistency of the survey instrument, a reliability analysis using Cronbach's alpha was conducted. The scale assessing perceptions of AI-generated misinformation and ethical risks demonstrated good reliability ($\alpha = 0.87$), indicating that the items consistently measured the underlying construct of concern regarding deepfakes and misinformation.

Results were further illustrated using bar charts to visually represent journalists' perceptions regarding the seriousness of deepfakes, AI's role in facilitating fake news, and the importance of human editorial oversight. This visualization facilitated the identification of response patterns across key variables and enhanced the clarity of the analysis.

The study focused on identifying dominant trends and perceptual patterns rather than establishing causal relationships, consistent with its exploratory design. Findings were interpreted in relation to existing literature on AI ethics, misinformation, and journalism studies, providing empirical insight into how emerging AI technologies are influencing trust, responsibility, and ethical standards in contemporary newsrooms

4.Results

This section presents the findings on journalists' perceptions of AI-driven misinformation, deepfakes, and their impact on journalistic credibility and ethics. A five-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree) was used. Visual bar charts are included as prompts for your graphs.

4.1. Perceived Threat of Deepfakes to Journalism

Respondents were asked whether deepfakes pose a serious challenge to journalistic integrity. A majority of journalists agreed that AI-generated deepfakes undermine public trust and complicate verification processes.

Figure 1. Journalists' Perceptions Deepfakes as a Serious Threat to Journalism

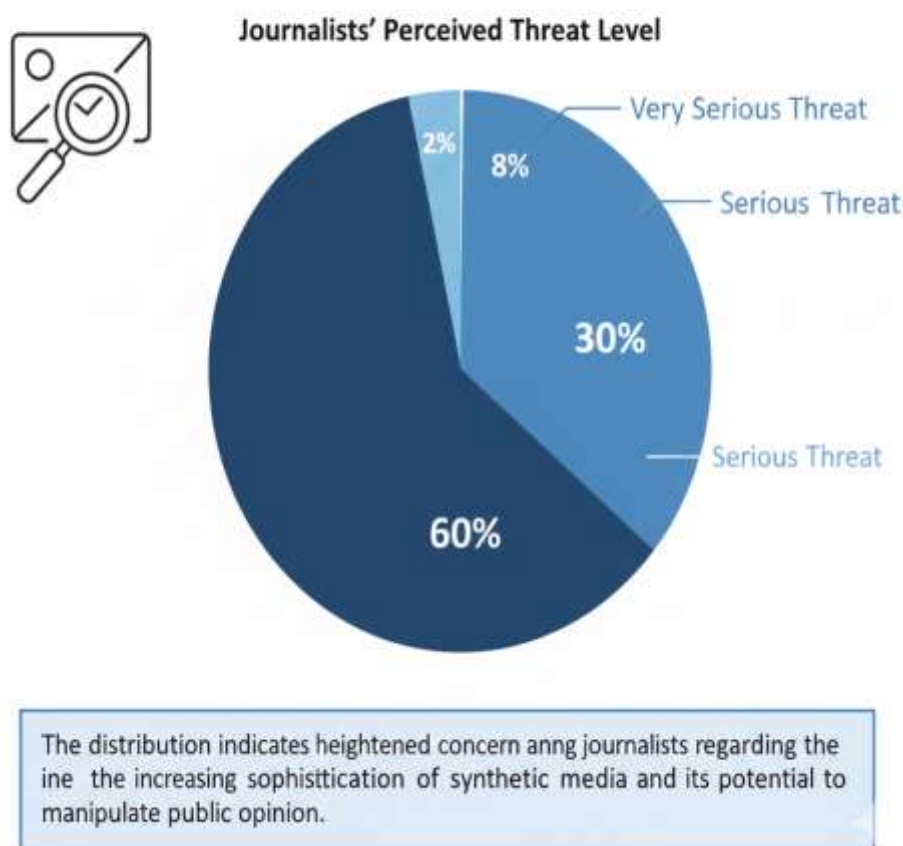


Figure 1. Journalists' perceptions of deepfakes as a serious threat to journalism

The distribution indicates heightened concern among journalists regarding the increasing sophistication of synthetic media and its potential to manipulate public opinion.

4.2. AI's Role in Facilitating Fake News

Journalists were also surveyed about AI's contribution to the creation and dissemination of fake news. Most respondents agreed that AI technologies accelerate the spread of misinformation, creating ethical and professional challenges in newsrooms.

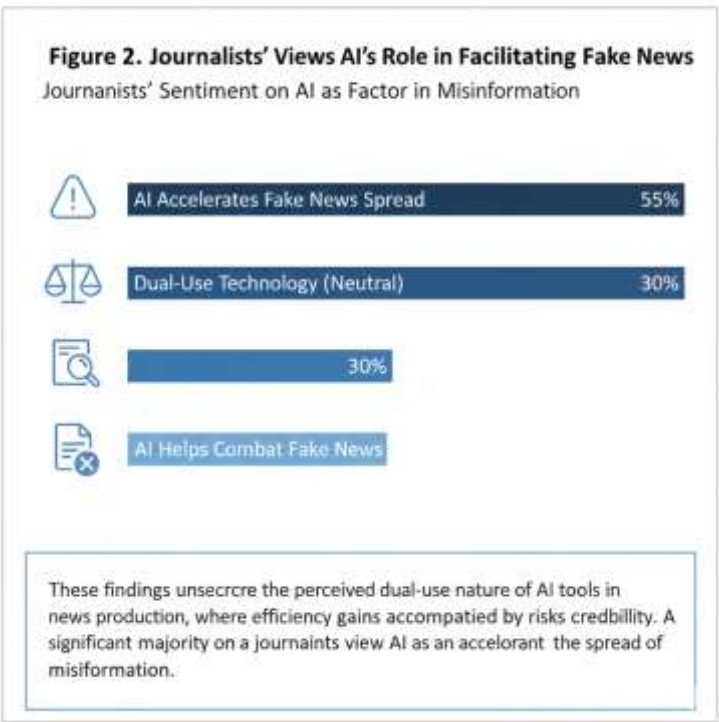


Figure 2. Journalists' views on AI's role in facilitating fake news

These findings underscore the perceived dual-use nature of AI tools in news production, where efficiency gains are accompanied by risks to credibility.

4.3. Ethical Concerns Related to Deepfakes and Misinformation

Respondents evaluated specific ethical concerns such as plagiarism, misattribution, and loss of originality in AI-assisted content. A substantial proportion of journalists expressed strong agreement that these issues compromise journalistic standards.

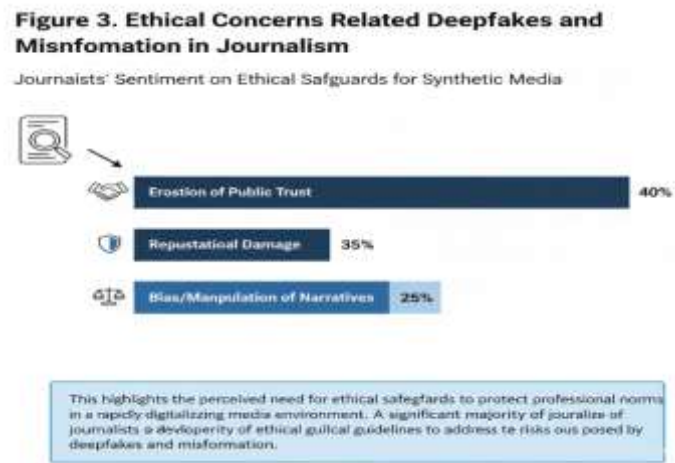


Figure 3. Ethical concerns related to deepfakes and misinformation in journalism

This highlights the perceived need for ethical safeguards to protect professional norms in a rapidly digitalizing media environment.

4.4. Disclosure and Editorial Oversight

Journalists were asked about the importance of disclosing AI-generated content and maintaining human editorial control. There was near-universal agreement that disclosure is essential and that editorial oversight cannot be replaced by automated systems.

Figure 4. Support for Disclosure of AI-Generated Content and Human Editorial Review

Journalists' Support for Transparency and Oversight

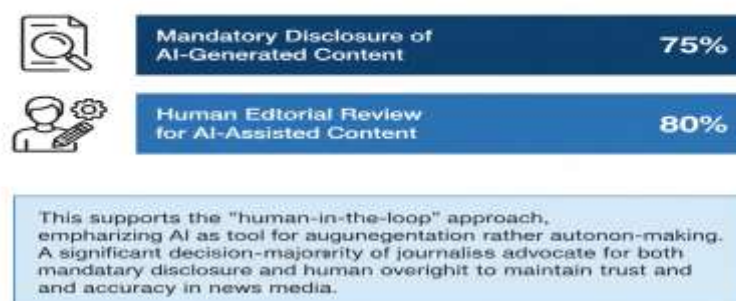


Figure 4. Support for disclosure of AI-generated content and human editorial review

This supports the "human-in-the-loop" approach, emphasizing AI as a tool for augmentation rather than autonomous decision-making.

4.5. Summary of Key Findings

- Journalists perceive deepfakes as a serious threat to credibility and public trust.
- AI is recognized as facilitating the production and spread of fake news.
- Ethical concerns such as plagiarism, misattribution, and loss of originality are prominent.
- Strong support exists for disclosure and human editorial oversight.

5. Discussion

The findings of this study provide significant insights into how journalists perceive the ethical and professional challenges posed by deepfakes and AI-driven misinformation. Overall, the results suggest that journalists are acutely aware of the risks associated with synthetic media and automated content generation, reflecting a global concern about credibility, trust, and verification in the digital age.

One of the most salient results is the perceived threat of deepfakes to journalistic integrity. The majority of respondents agreed that highly realistic AI-generated audio-

visual content undermines public trust and complicates verification processes. This aligns with international research indicating that deepfakes can manipulate political discourse, influence elections, and erode confidence in media institutions (Chesney & Citron, 2019; Vaccari & Chadwick, 2020). Journalists in Pakistan appear particularly sensitive to these risks, likely due to the country's highly polarized media environment and limited regulatory oversight of digital misinformation.

AI's role in facilitating fake news was also highlighted as a critical concern. Respondents acknowledged that automated tools accelerate the spread of misinformation, supporting findings from global studies showing that social media algorithms amplify AI-generated content, making it difficult for journalists to filter and fact-check rapidly (Vosoughi, Roy, & Aral, 2018; Tandoc et al., 2018). These perceptions reinforce the understanding that AI can serve as a double-edged sword in news production, offering efficiency gains while exacerbating ethical and credibility challenges.

Ethical concerns such as plagiarism, misattribution, and diminished originality emerged strongly in the data. Journalists perceive AI-driven content generation as a potential threat to professional norms, mirroring prior research suggesting that automated writing tools may inadvertently encourage content homogenization and shortcuts in verification (Carlson, 2015; Beckett, 2019). The findings underscore that ethical risks extend beyond overt misinformation to more subtle erosions of journalistic standards.

Finally, the strong support for disclosure and human editorial oversight reflects the enduring value of the human-in-the-loop model in ethical journalism. Journalists emphasized that AI should assist rather than replace human decision-making, reinforcing broader recommendations that transparency and editorial accountability are essential to maintain public trust (Guzman & Lewis, 2020; Lewis, Sanders, & Carmody, 2019).

Collectively, these results indicate that Pakistani journalists are technologically aware but ethically cautious. They recognize that AI-driven misinformation requires not only technical mitigation but also professional and institutional safeguards, highlighting the need for newsroom policies, training, and regulatory frameworks tailored to local contexts.

6. Conclusion

This study examined journalists' perceptions of AI-driven misinformation and deepfakes within Pakistani newsrooms, focusing on credibility, trust, and professional ethics. The results indicate that journalists view deepfakes and automated fake news as significant threats to content authenticity and public confidence in media institutions. AI-generated misinformation not only complicates verification processes but also challenges traditional editorial practices, potentially undermining journalistic standards and professional identity.

Ethical concerns, including plagiarism, misattribution, and reduced originality, were emphasized, suggesting that risks extend beyond overt misinformation to subtle compromises in professional norms. Despite the technological efficiency offered by AI, journalists strongly favored human editorial oversight, disclosure of AI-generated content, and adherence to ethical guidelines. These findings highlight that while Pakistani journalists recognize the transformative potential of AI, they remain cautious and prioritize ethical safeguards to protect credibility and trust.

6.1. Recommendations

Based on the study's findings, the following recommendations are proposed:

2. **Ethical Guidelines Development:** Media organizations should formulate clear ethical protocols for AI-assisted news production, emphasizing verification, accuracy, and attribution standards.
3. **Mandatory Disclosure Policies:** Newsrooms should disclose when content is AI-generated or AI-assisted, enhancing transparency and audience trust.
4. **Human Editorial Oversight:** Automated content must undergo rigorous human review to ensure ethical compliance and adherence to professional norms.
5. **Capacity Building and Training:** Journalism institutions and media organizations should offer structured training on AI literacy, detection of deepfakes, and strategies to mitigate misinformation.
6. **Regulatory and Policy Frameworks:** National media regulators should establish guidelines governing AI-generated news, addressing fake news proliferation, deepfakes, and content verification.
7. **Future Research Directions:** Further studies should explore the evolving impact of AI-driven misinformation longitudinally, including audience perceptions and cross-platform effects.
8. In sum, the study demonstrates that AI-driven misinformation poses both technological and ethical challenges, necessitating a combination of professional, institutional, and policy-oriented interventions to preserve journalistic integrity and public trust.

References

- Chesney, R., & Citron, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review*, 107(6), 1753–1819. <https://doi.org/10.2139/ssrn.3213954>
- Diakopoulos, N. (2019). *Automating the news: How algorithms are rewriting the media*. Harvard University Press.
- Farid, H. (2018). Photo and video forgery detection. In K. P. Murphy (Ed.), *Annual review of statistics and its application*, 5, 93–116. <https://doi.org/10.1146/annurev-statistics-031017-100318>

- Hansen, A., Allan, S., & Cottle, S. (2018). *Digital news and the politics of misinformation: Challenges for journalism*. Routledge.
- King, G., Pan, J., & Roberts, M. E. (2017). How the Chinese government fabricates social media posts for strategic distraction, not engaged argument. *American Political Science Review*, 111(3), 484–501. <https://doi.org/10.1017/S0003055417000144>
- Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., ... Zittrain, J. L. (2018). The science of fake news. *Science*, 359(6380), 1094–1096. <https://doi.org/10.1126/science.aao2998>
- Marwick, A., & Lewis, R. (2017). *Media manipulation and disinformation online*. Data & Society Research Institute.
- 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), 1–13. <https://doi.org/10.1177/2056305120903408>
- West, D. M. (2019). *The future of work: Robots, AI, and automation*. Brookings Institution Press.
- Zhou, P., Han, X., Morariu, V., & Davis, L. S. (2018). Learning rich features for image manipulation detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1053–1061). <https://doi.org/10.1109/CVPR.2018.00116>