

Trust Under Threat: The Influence of Fake News and AI-Manipulated Content on Mass Media Credibility

Jahangir Ashraf¹, Laiba Idrees² and Sara Arif³

<https://doi.org/10.62345/jads.2026.15.3.16>

Abstract

The current study aimed to find the influence of fake news and AI-manipulated content on mass media credibility, particularly through social media usage as a moderator, to fill the gap in the literature, as such studies are limited in the cultural context of Pakistan, where awareness regarding digital media use is limited and media illiteracy is very high. The study adopted a cross-sectional correlational study design and purposive sampling technique to collect data from the Pakistan cohort with at least intermediate qualification and ages 18 to 39. The study adopted structured questionnaires to measure variables with reliability ranges from 0.73 to 0.90. The Pearson correlational analysis reported a significant positive relationship between fake news and AI-manipulated content, whereas fake news has a significant negative relationship with media credibility. AI-manipulated content has a significant positive relationship with social media usage and a negative relationship with media credibility. Furthermore, social media usage possesses a significant negative relationship with media credibility. Furthermore, Hayes Macro Process 4.2 reported that social media usage significantly and positively moderated the negative influence of fake news and manipulated content on mass media credibility. The study provided comprehensive implications for youth of Pakistan.

Keywords: Fake News, AI-Manipulated Content, Mass Media Credibility.

Introduction

The world is getting technologized day by day, and the expansion of digital media has transformed the daily lives of people across the globe, as these modern digital media platforms like Facebook, Instagram, TikTok, Twitter/X, and YouTube provide access to education, make business convenient, and serve as a source of information, etc. (Elizebath, 2025; Irfan et al., 2025; Khan et al., 2024; Shahid et al., 2024). Social media has rapidly replaced traditional media as a primary source of news, as more individuals rely on digital platforms than on television (Zhang, 2024). The shift is evident not only in developed but also in developing countries, as access to social media remains convenient (Ali et al., 2019). However, the widespread use of social media has also contributed to the spread of misinformation, including fake news and AI-manipulated content, such as pictures, videos, and written posts (Aïmeur et al., 2023).

Fake news is deliberately misleading information, often designed to elicit emotional reactions and attract widespread attention (Munusamy et al., 2024; Baissa et al., 2025). Whereas AI-manipulated content is produced using advanced artificial intelligence techniques that generate

¹MPhil (Mass Communication), Alumni, Government College University Faisalabad, Pakistan.

Email: jahangir.ashraf@gmail.com, ORCID: <https://orcid.org/0009-0009-3010-1578>

²Lecturer, Department of Psychology, Riphah International University, Faisalabad Campus, Pakistan.

Email: laibaidrees64@gmail.com ORCID <https://orcid.org/0000-0003-3123-9631>

³MPhil (HRM), Alumni, Pakistan Institute of Quality Control, Lahore, Pakistan. Email: arifs1614@gmail.com



highly realistic but fabricated visual and audio content, making detection difficult (Acim et al., 2025; Singh et al., 2025).

The situation is further worsened by the liar dividend effect, in which individuals reject even genuine information by labelling it as false, ultimately reducing trust in authentic content (Goldberg, 2025). Furthermore, algorithm-driven content distribution on social media creates echo chambers where people mostly see ideas that align with their beliefs, making it easier for false information to spread (Ferreira, 2025). Consequently, public trust in mass media has significantly declined, and its credibility is therefore negatively affected across different age groups and regions worldwide (Denniss & Lindberg, 2025; Kioussis, 2001).

The fake news and AI-manipulated content in daily life examples include major political events, rumours of violence, discussions about celebrities, and the death of influential people; this content sheds light on serious risks related to misinformation (Alsaedi et al., 2017; Stubbs et al., 2024; Kulsum et al., 2025).

There is substantial literature on the negative influence of fake news and AI-manipulated content; however, the combined effect on mass media credibility remains under-studied, particularly in a country like Pakistan, where digital literacy is limited, and people widely believe in rumours, as the majority use social media platforms. Therefore, the aim of the study is to fill the gap in the cultural context of Pakistan by investigating social media usage as a moderator that strengthens the impact of fake news and AI-manipulated content on mass media credibility.

Hypothesis

H1: Social media usage moderates the relationship between AI-Manipulated content and mass media credibility, such that higher usage strengthens the negative effect.

Methodology

The study adopted a cross-sectional correlational design to examine the relationships among fake news, AI-manipulated content, and mass media credibility. Furthermore, studies also tend to find a moderating role for social media between predictors of fake news and AI-manipulated content and the outcome of mass media credibility.

The study used a purposive sampling technique to collect data from a sample of young adults aged 18 to 39 who have been actively engaged on social media platforms for the last year. These individuals must possess at least an intermediate qualification and must be Pakistani citizens. The study sample comprised 108 participants, which is considered adequate for moderation analysis using the PROCESS macro version 4.2, as moderation models can produce reliable estimates with moderate sample sizes.

The study followed APA 7 ethical considerations; i.e., before data collection, participants were provided with a consent form outlining voluntary participation, the right to withdraw at any time, and the confidentiality of their data. Upon their written signature, they were given the demographic questionnaire and the structured study questionnaire. The data were collected from individuals from Islamabad, Lahore, and Faisalabad. The demographic form included age, gender, education, socio-economic status, and the number of social media platforms used per day. All the structured questionnaires used a 5-point Likert scale ranging from strongly disagree (1) to strongly agree (5). The structured questionnaire on fake news obtained a reliability of 0.73, AI-manipulated content of 0.78, mass media credibility of 0.82, and social media usage of 0.90. Furthermore, content validity was established through expert review and research scholars in the communication and journalism fields, and factor loadings exceeded 0.62, indicating that the questionnaire precisely assesses the variables it intended to measure.

Example Items

Fake News: I frequently watch misleading news online.

AI-Manipulated content: It is hard for me to differentiate between real news and news with AI-manipulated content

Mass Media Credibility: I believe or trust information I get from mass media.

Social Media Usage: I depend on social media for receiving the latest new updates.

Results

Table 1: Demographic Characteristics

Characteristics	Frequency	Percentage	M (SD)
Age			23.41 (2.27)
Gender			
Men	60	55.55	
Women	48	44.55	
Education			
Intermediate	27	25	
Bachelor	48	44.44	
Master	21	19.44	
PhD	12	11.12	
Socioeconomic Status			
Lower Class	34	31.48	
Middle Class	52	48.15	
Upper Class	22	20.37	
Daily Social Media Usage			
1 to 4 hours	49	45.37	
5 to 8 hours	37	34.26	
8 to 12 hours	13	12.04	
More than 12 hours	9	8.33	

Note: M = Mean; SD = Standard Deviation

The table shows that participants had a mean age of 23.41 (SD = 2.27), with 60 men (55.55%) and 48 women (44.55%), most having a bachelor's degree (n = 48, 44.44%), followed by intermediate (n = 27, 25%), master's (n = 21, 19.44%), and PhD (n = 12, 11.12%), while socioeconomic status included lower class (n = 34, 31.48%), middle class (n = 52, 48.15%), and upper class (n = 22, 20.37%), and daily social media usage was 1–4 hours (n = 49, 45.37%), 5–8 hours (n = 37, 34.26%), 8–12 hours (n = 13, 12.04%), and more than 12 hours (n = 9, 8.33%)

Table 2: Correlation Matrix

Variable	1	2	3	4
1.Fake News	-	0.51**	0.43**	-0.59**
2.AI-Manipulated Content		-	0.44**	-0.55**
3.Social Media Usage			-	-0.46**
4.Media Credibility				-

Note: ** p < 0.01

Table 2 depicts a significant positive relationship between fake news and AI-manipulated content, whereas fake news has a significant negative relationship with media credibility. AI-manipulated content has a significant positive relationship with social media usage, and a negative relationship with media credibility. Furthermore, social media usage possesses a significant negative relationship with media credibility.

Table 3: Moderation Analysis

Predictor	β	t-value	Sig
Constant	4.01	9.21	.001
Fake News	-0.39	-7.01	.001
AI-manipulated content	-0.41	-4.79	.001
Social Media Usage	-0.24	-3.94	.001
Interaction Effect (Moderation)			
Fake News X Social Media Usage	-0.31	-2.52	.01
AI-Manipulated content X Social Media Usage	-0.37	-2.47	.02

Note: * $p < .05$, ** $p < .01$, *** $p < .001$

Table 3 indicates that fake news ($\beta = -.39$, $t = -7.01$, $p < .001$), AI-manipulated content ($\beta = -.41$, $t = -4.79$, $p < .001$), and social media usage ($\beta = -.24$, $t = -3.94$, $p < .001$) significantly negatively predicted media credibility. Moreover, the interaction effects of fake news $\times$ social media usage ($\beta = -.31$, $t = -2.52$, $p = .01$) and AI-manipulated content $\times$ social media usage ($\beta = -.37$, $t = -2.47$, $p = .02$) were significant, indicating moderation effects.

Discussion

The study aimed to find the influence of fake news and AI-manipulated content on mass media credibility, particularly through a moderator, social media usage, in the cultural context of Pakistan, to fill the gap in the indigenous literature and come up with proper implications for the youth of Pakistan.

The result of the study proved the assumption of the study, as social media usage moderated the influence of fake news and AI-manipulated content on mass media credibility. The result is aligned with various international studies conducted in China, Japan, the United States of America, Spain, France, and Egypt, which reported that those who frequently use social media tend to have lower mass media credibility and trust, as social media platforms like Facebook, TikTok, Instagram, etc., consistently allow the posting of AI-generated videos and content. The rumors of deaths, fights, marriage, divorce, engagement, etc., among celebrities are frequently posted on social media for the sake of views, and people who lack proper media literacy tend to believe mass media like social media, which declines mass media credibility (Boediman, 2025; Folorunsho & Boamah, 2025; Hussein & Özad, 2025; Mammadzada, 2021; Peña-Alonso et al., 2025; Sanchez-Acedo et al., 2024; Tsetsura et al., 2025; Wang, 2025).

The reason for such results in the cultural context of Pakistan may be a lack of awareness among people regarding online news authenticity. Furthermore, there is widespread social media use, so fake news and AI-made content quickly spread among users. People usually trust information shared by friends and peers on social media without verifying it. Consequently, mass media credibility decreases when social media use is very high.

Conclusion

The study aimed to find the moderating effect of social media usage between predictors fake news and AI-manipulated content on outcome mass media credibility to fill the gap in the cultural context of Pakistan. The study adopted a cross-sectional correlational study design and a purposive sampling technique to collect data from at least intermediate qualified individuals in Pakistan, aged 18 to 39. The study found that social media usage moderated the impact of fake news and AI-manipulated content on mass media credibility. The study thus filled the gap in the cultural context of Pakistan. Educational institutes must play a crucial role in enhancing digital media awareness among students. Furthermore, mental health professionals must spread awareness of safe digital media use through workshops, webinars, seminars, etc. It is

recommended for future studies to add more demographics like the names of apps, employment status, and parental education. Moreover, future studies need to adopt stratified sampling techniques for generalizability purposes, and scales in the Urdu version so that these could be easily understood among the less educated cohort of Pakistan as well.

References

- Acim, B., Boukhelif, M., Ouhnni, H., Kharmoum, N., & Ziti, S. (2025). A Decade of Deepfake Research in the Generative AI Era, 2014–2024: A Bibliometric Analysis. *Publications*, 13(4), 50.
- Aïmeur, E., Amri, S., & Brassard, G. (2023). Fake news, disinformation and misinformation in social media: a review. *Social Network Analysis and Mining*, 13(1), 30.
- Ali, M., Iraqi, K. M., Khan, L., & Salam, F. (2019). The Effect of Traditional Media Communication and Social Media Communication in Generating Consumer Based Brand Equity in Context of Pakistan. *South Asian Journal of Management Sciences*, 13(2).
- Alsaedi, N., Burnap, P., & Rana, O. (2017). Can we predict a riot? Disruptive event detection using Twitter. *ACM Transactions on Internet Technology (TOIT)*, 17(2), 1-26.
- Baissa, B., Fuoli, M., & Grieve, J. (2025). The news values of fake news. *Discourse & Communication*, 19(3), 418-440.
- Boediman, E. P. (2025). Exploring the impact of deepfake technology on public trust and media manipulation: A scoping review. *Jurnal Komunikasi*, 19(2), 131-152.
- Denniss, E., & Lindberg, R. (2025). Social media and the spread of misinformation: infectious and a threat to public health. *Health promotion international*, 40(2), daaf023.
- Elizabeth, L. (2025). The Rise of Social Media Platforms. In *The Practitioner's & Professional's Guide to Intimate Partner Violence Advocacy in the Digital Age* (pp. 21-34). Cham: Springer Nature Switzerland.
- Ferreira, L. (2025). Misinformation on Social Media: Between Usage Practices, Digital Literacy and Social Influence. *Estudos em Comunicação*, 1(41), 179-198.
- Folorunsho, F., & Boamah, B. F. (2025). Deepfake technology and its impact: ethical considerations, societal disruptions, and security threats in ai-generated media. *International journal of information technology and management information systems*, 16(1), 1060-1080.
- Goldberg, H. (2025). States Legislating against Digital Deception: A Comparative Study of Laws to Mitigate Deepfake Risks in American Political Advertisements. *Notre Dame J. on Emerging Tech.*, 6, 1.
- Hussein, K., & Özad, B. (2025). Ai-driven media manipulation: public awareness, trust, and the role of detection frameworks in addressing deepfake technologies. *İnterdisipliner Medya ve İletişim Çalışmaları*, 2(3), 98-133.
- Irfan, A., Abbas, N., Shahid, M. S., Ashraf, J., & Asif, A. (2025). Smartphone addiction and academic procrastination in university students: psychological wellbeing as a moderator. *Qualitative Research Journal for Social Studies*, 2(2), 973-983.
- Khan, P., Shahid, M. S., & Majid, H. A. (2024). Mediating role of online social connectedness in smartphone addiction and online relationship among young adults. *Gomal University Journal of Research*, 40(3), 327-336.
- Kioussis, S. (2001). Public trust or mistrust? Perceptions of media credibility in the information age. *Mass communication & society*, 4(4), 381-403.
- Kulsum, N. M., Ginting, S., & Salim, A. (2025). Revisiting the Public Sphere: Communication, Digital Mobilization, and Misinformation in Indonesiaâ€™s 2019 and 2025 Mass Demonstrations. *Jurnal Komunikasi Ikatan Sarjana Komunikasi Indonesia*, 10(2), 481-491.

- Mammadzada, I. (2021). *Deepfakes And Freedom Of Expression: European Perspective* (Doctoral dissertation, Master's thesis, Tallinn University of Technology]. <https://digikogu.taltech.ee/et/Download/ec6ea9ff-bd47-49e4-986d-3a115c00300b/SvavltsingjavljendusvabadusEuroopaperspekt.pdf>.
- Munusamy, S., Syasyila, K., Shaari, A. A. H., Pitchan, M. A., Kamaluddin, M. R., & Jatnika, R. (2024). Psychological factors contributing to the creation and dissemination of fake news among social media users: a systematic review. *BMC psychology*, 12(1), 673.
- Peña-Alonso, U., Peña-Fernández, S., & Meso-Ayerdi, K. (2025). Journalists' perceptions of artificial intelligence and disinformation risks. *Journalism and Media*, 6(3), 133.
- Sanchez-Acedo, A., Carbonell-Alcocer, A., Gertrudix, M., & Rubio-Tamayo, J. L. (2024). The challenges of media and information literacy in the artificial intelligence ecology: deepfakes and misinformation. *Communication & Society*, 223-239.
- Shahid, M. S., Bashir, S., & Fatima, S. (2024). Social media addiction and aggression in Pakistani young adults: Social connectedness as a moderator. *Journal of Professional & Applied Psychology*, 5(3), 424-433.
- Singh, L. H., Charanarur, P., & Chaudhary, N. K. (2025). Advancements in detecting AI-manipulated content: AI algorithms and future prospects— a review. *Discover Internet of Things*, 5(1), 53.
- Stubbs, J. E., Nicklin, L. L., Wilsdon, L., & Lloyd, J. (2024). Investigating the experience of viewing extreme real-world violence online: Naturalistic evidence from an online discussion forum. *New Media & Society*, 26(7), 3876-3894.
- Tsetsura, K., Murtuza, H. M., Raymond, M., & Joffe, T. (2025). Governing Harms from Deepfakes in Crisis Situations: Comparing Legal and Regulatory Frameworks of G7 Countries and the EU. *Mediascapes journal*, 26(2), 201-221.
- Wang, X. (2025). A Comparative Study of Deepfake Policy in the United States and China. *International Journal of Frontiers in Sociology*, 7(4).
- Zhang, Y. (2024). The impact of new media on traditional media. *Journal of Education, Humanities and Social Sciences*, 28, 691-696.