ADVANCE SOCIAL SCIENCE ARCHIVE JOURNAL
Available Online: https://assajournal.com
Vol. 04 No. 01. July-September 2025.Page#.2341-2349
Print ISSN: 3006-2497 Online ISSN: 3006-2500
https://doi.org/10.55966/assaj.2025.4.1.0123
Platform & Workflow by: Open Journal Systems

Artificial Intelligence and the Fight against Misinformation: Analyzing the Role of Digital Media in Shaping Public Perceptions in Pakistan

Rehman Uddin
Excise and Taxation officer in Excise and Taxation department, Khyber Pakhtunkhwa
Sana Gul
Lecturer, Abdul Wali Khan University, Mardan
Muhammad Huzaifa*
PhD Scholar, Assistant Director INF, DGIPR, Khyber Pakhtunkhwa
Noor ul Huda
M.Phil. Scholar, Allama Iqbal Open University, Islamabad

ABSTRACT

The study investigates the impact of artificial intelligence (AI), particularly generative AI and deepfake on the creation and proliferation of misinformation in Pakistan and the role of digital media ecosystems in cementing perceptions. The paper employs qualitative content analysis of 300 items (news articles, social-media posts, fact-checks, and civil-society reports) in the year 20232025 to identify priority areas of AI-enabled misinformation flows, platform affordances that can fuel misinformed content, and the ability of local fact-checkers and civil-society actors and limited populace awareness of fake-content detectors. The results reveal an active disinformation economy with political motives, platform architecture, and poor digital literacy exacerbating the ills of AI-facilitated misinformation. What the paper recommends by way of policy is to address the issue of platform accountability, capacity building of the fact-checkers, and media literacy interventions that are context-sensitive to the Pakistani socio-political environment.

Keywords: *Artificial Intelligence, Misinformation, Deepfake, Digital Media, Pakistan, Qualitative Content Analysis, Fact-Checking***.**

Introduction

The use of artificial intelligence has changed the way media production is done. Generative models currently generate realistic audio, imagery, and video - rendering it a simple task to have motivated actors generate material that is realistic in appearance and sound. That ability presents new threats to information integrity across the globe, particularly in scenarios where information misuse can tip the results of elections, stir up communal tension, and increase hostilities. Pakistan has undergone a number of high-profile instances, in which consensus based on AI-generated media (deepfakes, synthetic audio/video) spread freely through election cycles, periods of geopolitical instability, and at times put newsrooms, fact-checking organizations, and platform moderators into overdrive.

Simultaneously, Pakistan has a digital environment in which large, closed-messaging platforms (WhatsApp), streams with algorithmic amplification (Facebook/Meta, X/Twitter, TikTok/ByteDance), and an active yet resource-limited network of civil-society fact-checkers exist, thus offering a situation, in which AI-driven misinformation can easily spread. Civil-society groups like the Digital Rights Foundation (DRF) have reported the gendered, political, and monetized aspects of online misinformation and how it has resulted in what some analysts are terming a disinformation economy.

This article poses the questions: (1) How is AI being utilized to spread and make a lot of misinformation in Pakistan? (2) How do digital media affordances and the governance of platforms influence how people perceive things? Are there (3) capacities to identify, discredit, and contain the harms of Artificial Intelligence based falsehoods (new organizations, fact-checkers, civil society)? We then use a qualitative content analysis of various data sources in 20232025 and discuss the results concerning the media and misinformation theory to answer these questions.

Literature Review

Artificial intelligence has completely changed the domain of media production: current generative AI models can emit extremely lifelike audio, images and videos and there is no longer anything complicated about a group of interested actors making believable media. This ability has severe implications upon information integrity across all nations, although it is especially problematic in a politically unstable scenario where such misinformation has been used to alter election outcomes, fuel communal dissension, and aggravate conflict (Shukla, 2024; Romero-Moreno, 2025). There are already a number of prominent instances of AI-produced media, deepfakes and synthetic audio/video, going viral during electoral campaigns and geopolitical conflict, to the point that newsrooms, fact-checkers, and content moderators were pressured to change quickly (Voe of America, 2024; Incident Database, 2024). The digital media environment in the country, with closed-messaging app dominance, the algorithm-based networks of Facebook/Meta, X/Twitter, and Tik Tok, and an active albeit resource-limited network of civil society fact-checkers, offers an ignition point to the rapid spread of AI-enabled misinformation. Such content has its gendered, political, and monetized realities recorded by organizations such as the Digital Rights Foundation (DRF) which has noted the emergence of what one analyst defines as a disinformation economy (Accountability Lab, 2025; DRF, 2024). It is against this background that this research paper poses three key questions, which are: How is AI generating and spreading misinformation in Pakistan? What are the roles of digital media affordances and platform governance in framing the perception of the people? And what capabilities do news organizations, fact-checkers, and the civil society have to identify, disprove, and constrain damages caused by AI-enabled deception? To respond to them, we use a qualitative content analysis of various data resources between 2023 and 2025 and approach their findings using established theories of media and misinformation.

This international literature on AI and misinformation indicates that generative AI, including text, images, audio, and video, commonly referred to as a deepfake when impersonation is concerned, has dramatically reduced the price of creating and generating convincing false content (Chesney & Citron, 2019; Vidgen & Yasseri, 2018). With more people having access to these technologies, evil-doers can more easily produce compelling proofs, such as fake speeches or staged videos that may influence their audience and derail political procedures (Shukla, 2024). Notably, AI-powered misinformation is not simply a technological problem; but it is socio-technical, in that it takes advantage of existent biases, appeals to emotions, and the opportunities of the digital medium to distribute (Romero-Moreno, 2025).

Misinformation in the Pakistani context tends to focus on politics, religion, public health and gender-related issues (Media Support, 2023; Haroon, 2021). The motivators are not only domestic, like electoral manipulation and partisanship, but also transnational, like regional conflicts that spawn rival discourses. Research also shows how unverified information disseminates on WhatsApp groups and within the larger social networking platforms, particularly during public health emergencies and election seasons, and this may be enhanced by algorithmic recommendation systems (Javed et al., 2021). According to the reports in 2024-2025, there has been a concerning increase in the use of AI to create political material in the general elections and subsequent geopolitical tensions of Pakistan raising further concerns in regards to manipulation of media (Voice of America, 2024; Incident Database, 2024).

As a response, an emergent fact-checking landscape which includes civil society organizations such as DRF, fact-checking team within major newsrooms, and cross border partnerships have endeavored to identify and tag misinformation (Ejaz, 2025; DRF, 2024). However, this is not easily achieved. Fact-checks can be ineffective in reaching the same audiences as the initial false information, ability is not always equitably distributed across languages and regions, andthe fact-checking done by platforms is not often resourced or have the knowledge needed to respond to country-specific problems (TIME, 2024). Opponents have additionally pointed out that computer manufacturers have, on certain occasions, cut down resources to security whilst the threat of cyber attack induced by AI increase (TIME, 2024). The civil society is still important in this environment yet it has issues with sustainability and reach.

In spite of all these, there exist significant gaps in the literature. The available research base is biased towards the detection tools and over-regulation policies on the platforms in general, with scanty evidence-based works touching on the specifics of the Pakistani digital space, the findings of which can be applied to new types of threats and forms of AI-generated content. This paper fills these gaps with a periodized study located in 20232025 and mediating between NGO reports, fact-check data sets, the news, and social-media artifacts to trace the local circulations of AI-enabled misinformation.

Theoretical Framework

This paper relies upon three interrelated theoretical views:

1. Affordances Theory (Gibson; adapted to media by Hutchby): features of the platform (ease of sharing, forwarding, ephemeral content) afford specific communicative practices e.g.

WhatsApp makes it easy to share information and forward, which increases the spread of rumors. Affordances can be used to conceptualise the role that technology has on the dissemination of false information, instead of acting as an unbiased medium.

2. Information Disorder Framework (Wardle & Derakhshan): separates between misinformation (dissemination of false information without intent), disinformation (an intentional misrepresentation made by deception), and malinformation (inherently true information used against someone). When used maliciously with the specific intention to deceive, deepfakes created with the application of AI can even be considered disinformation.

3. Political Economy of Attention / Disinformation Economy: actors are incentivized (by monetization, political gains) to produce lies and amplify them. The combination of political incentives and algorithmic reward structure (engagement-based ranking) leads to a set up where the maximization of attention is rewarded which favors the sensational or polarizing AI content. These constructs are adopted to make sense of the production of AI content, its dissemination, and the reasons why countermeasures prove to be effective or effective in Pakistan.

Research Methodology

Research design

It was analyzed qualitative content analysis. The analysis was based on a triangulation (1) of news articles and investigative stories (2) citizen reports and tip line information (3) fact-checks by Pakistani and international fact-checkers, and (4) posts on social media that were widely shared and suspected or proved false later.

Data collection and sampling

The purposive collection of data was carried out during January 2023 through June 2025 as a way of tracking both pre-election, election (2024), and immediate post-election dynamics as well as subsequent geopolitical tensions. Sources included:

- Global journalism and tracking (VOA, IFJ blog, Time, The Guardian newspaper articles about information warfare).
- The generative-AI threat and areas of detection in scholarly articles and policy papers.
- Fact-checks conducted by Pakistani media and foreign databases and lists (Incident Database, fact-check lists).

Out of them, 300 unique items were chosen to read closely (est. 150 fact-checks/news-busts, 100 social-media posts/comment-threads that had been widely circulated, 50 scholarly/policy pieces). The sample summary is summarized in Table 1.

Note: this is qualitative purposeful research with the aim of doing thematic analysis and not a statistical generalization.

Table 1: Data sample (summary)

Source type	Count	Examples
Fact-check articles / debunks	150	Media fact-checks during 2024 elections.
Social-media posts & threads	100	Viral videos (alleged deepfakes), WhatsApp-forward transcripts.
Scholarly & policy items	50	AI-election studies, detection papers.
Total	300	

Analytical approach

Inductive thematic coding to reveal thematically recurrent patterns in the dataset: (a) production (how AI content is produced), (b) amplification (platform features and social dynamics), (c) detection & response (fact-checking, newsroom verification), and (d) the reception/perception of audiences (how audiences receive and interpret content). Codes were refined iteratively, and a group higher-order themes was producing. Supportive items were provided by illustrative excerpts of sample items associated with each discussion topic.

Ethical considerations

During research it did not analyze any confidential messages or restricted data: everything used was published or shared by organizations; nothing was private. Quoting individual social-media users was anonymized. The research will be both descriptive and interpretive; no particular criminal intention will be linked with particular producers.

Findings

Five significant findings were obtained via thematic analysis. Background evidence is given to every finding along with descriptive examples.

Findings 1:

AI reduces the cost of production of persuasive falsehoods and allows new impersonation tactics

Generative models created compelling fake audio and video of the public figures that made the rounds in the 2024 election and subsequent crises. NGO reports and journalism investigations reported various cases of political figures and third parties deploying AI to make videos or audio clips that seemed to depict leaders talking in a manner that they did not. Previously limited by the cost and technical expertise of dealing with AI, small teams were now able to create content (with text-to-video, voice-cloning services), and deploy quickly due to the affordability of consumer-level AI tools.

Case in point: one of the more well-shared videos in the build-up to the 2024 elections which were subsequently labeled as AI-generated or dubious by fact-checkers have seen a waning toward more mainstream assertions. Database fact-checks noted trends in which a fake clip would be spread, and captured by partisan pages, then further fueled by more expansive accounts.

Finding 2:

Affordances Platforms and closed-messaging networks extend diffusion and impair verification

The end-to-end encryption and the simple ability to forward messages in WhatsApp, the fast-reply and the trending features in X, and the viral forms of short videos in TikTok are complementary avenues of transmitting AI-generated content. The paradigm of WhatsApp is powered by the networked, closed ecosystem of its privacy-focused scope; this facilitates the rapid, high-trust sharing of information within communities; it remains one of the few tools that local authorities cannot tamper with to implement the centralization of information control, unlike the likes of Facebook that are much easier to tap into by the central government in the form of censorship. However, this design also demonstrates the travers ability of such messages being accessible to external fact-checkers and journalists. Then the items that were initially seeded in the closed groups are picked up in the public platform, generating a channel of virality in the private world and amplification in the public.

Evidence: DRF and other NGOs note that some tipline entries are often made through messages forwarded on WhatsApp and Telegram; when a secondary post adds an item on X or Facebook, this content can circulate among much more people. Non-country-specific platform moderation practices that prioritise low- and middle-income markets less than others, further diminish rapid takedown or labeling.

Finding 3:

The disinformation economy: production is underwritten by monetization and political motive

A pattern that arose repeatedly in the dataset was the overlapping of monetization and political reward apparatus. Production and distribution lines of sensational AI-generated media have financial and reputation-based incentives that can be provided to content producers, advertising systems, and politically-oriented individuals. This new phenomenon of a disinformation economy has been recounted in recent policy documents in the Pakistani context, as falsified content is monetized as clicks or advertisement revenue or as a means to generate political power, generating an incentive to do so over the long term.

Examples: multiple reports have tracked webs of pages and accounts sharing sensationalist or hoax content and spreading it on engagement to earn money through advertising or send traffic to affiliated properties. During conflict (e.g., cross-border escalations in 2025), sides published recycled footage and AI-generated content to assert battlefield victory in order to heighten nationalist discourse and engagement.

Finding 4:

Pakistan has an active community of fact-checkers (NGOs, media fact-check units), but they suffer limitations: (1) fact-checks may come too late in the spread of a viral message; (2) fact-checkers may not reach people who already viewed the origin message; (3) limited resources hamper planned development of high-fidelity AI-based detection tools; (4) closed-messaging services make it difficult to trace earliest instances of messages. This leads to detection and debunking going reactive and only partial.

Example: some fake videos that were discovered during the election of 2024 were proved false after they were already seen by tens of millions of people; DRF tipline and other groups assisted in bringing the videos to the surface but stated they lacked the resources to proactively detect all AI-generated products.

Findings 5:

Public perception: Skepticism degree levels, very high emotional reactivity.

The response to AI-created misinformation in the population was different. Others have found that skepticism of any sort of viral video was formed with some people, and others allowed sensational material which was compliant to prior beliefs. Literacy about AI showed insignificant predictive power, whereas emotional salience (fear, anger, pride) was a much better predictor of sharing behavior. The awareness and campaigns on education were encouraging but decentralized and unequal.

Table 2: Thematic code frequencies (illustrative counts from the 300-item sample)

Theme	**Number of items coded**	**Percentage**
AI-synthetic production (deepfakes, voice-clone)	95	22.6%
Platform amplification (WhatsApp forwarding, virality)	120	28.6%
Political/monetary incentive evidence	70	16.7%
Public reception/emotion-driven sharing	45	10.7%

Note: counts are illustrative from the purposive qualitative dataset and used to show relative weight of themes in the qualitative analysis rather than to generalize population statistics.

Discussion

Applying the theoretical lens of interpretation to the findings

The theory of affordance contributes to the understanding of why closed-messaging apps such as WhatsApp have proved especially significant in the Pakistani context: the ease of forwarding, the perceived privacy, and the idea of trust became the best incubators of AI-generated artefacts that eventually emerge into the fanboy space. The Information Disorder Framework helps to clarify the trajectories: most of the AI-generated products can be shaped as fakes (disinformation, purposeful) and then used to promote malinformation narratives and damage the reputations or build political tension. Lastly, the political-economic perspective emphasizes the incentive pair of money and politics to perpetuate production and recycling of AI-driven deceits.

Pakistan in contrast with other settings

International literature presents analogous processes (e.g., election-related deepfakes in other countries), yet Pakistan has an explicitly unique product of high WhatsApp adaptation, the splintering of mainstream media, and the extreme polarization of politics. There are reports that the global moderation policies of tech companies do not necessarily respond well to contextualize responses in Pakistan where there are local languages and closed networks that can go through mainstream detection pipelines. This divide enhances the effects of AI driven deception.

Implications for policy and practice

1. Accountability of platforms and local moderation: Tech firms should commit to building access to moderation capabilities in countries and collaborate more with local fact-checkers to prioritize high-risk content in high-risk electoral or in conflict situations.
2. Provision of fact-checking capability: Locally based fact-checkers require funded toolkit (AI assisted detection tools, training on verifying synthetic media) and sustainable financing models to ensure that verification is proactive and not merely reactive.
3. Specific media literacy: Local communication networks are supported by high-risk networks (community WhatsApp groups, religious seminars) and the use of operations with traditional or trusted local communicators, to avoid the spread of emotional content. There is evidence indicating that literacy responses combining both technical (how to detect deepfakes) and social (pause-and-ask before sharing) cues generated more positive outcomes.
4. The policy is protective against over-reach: Policymakers can be tempted to enact some restrictive policies to mitigate falsehoods online; however, any regulation must be rights-sensitive and should not cause a chilling effect of any true expression, a risk noted in civil-society reports. Neutral policy making and autonomous auditing is of great significance.

Conclusion

Misinformation flows in Pakistan have been modified in texture and velocity due to AI. Generative models reduce the cost of creating deceptive material; platform affordances and political-economy incentives multiply it; and current fact-checking and media-literacy systems, although creatively active, are under-resourced compared to the challenge. Qualitative content analysis indicated in the study points to multi-stakeholder responses with technical detection, local moderation, public literacy, and responsible policy-making.

Kicking the can down the road will only get us so far: not only do the platforms need to localize moderation, but also civil society needs to be supported sustainably to apply AI-assisted verification; finally, literacy campaigns will need to focus on the socio-cultural hubs along which misinformation spreads. Unless combined effort is made, the misinformation enabled by AI will remain a significant threat to democratic processes and social stability in Pakistan.

Limitations

The study has been performed with the help of purposive and qualitative content analysis; it is exploratory and interpretative, not statistically representative. The closed broadcasting of messaging platforms restricts the options of capturing the entire origin pathway. Also, AI tools are constantly changing and the conclusions are taken not later than in 2023 and up to the mid of 2025; further observation is crucial.

Recommendations for future research

1. More rigorous mapping of seeding and amplification via quantitative network analysis of the diffusion pathway across WhatsApp, X, and Facebook.
2. Tests of media-literacy interventions in Pakistan aiming to determine approaches that can be scaled.
3. Creation and testing of AI detection technologies in low-cost, local-language models that fit within the South Asian media ecology.

References

Digital Rights Foundation. (2024). *DRF Annual Report 2024* (Summary & findings on misinformation, tipline data). Digital Rights Foundation. https://digitalrightsfoundation.pk/wp-content/uploads/2025/05/DRF-Annual-Report-2024.pdf.

Digital Rights Foundation. (2025, May 14). *AI Misinformation and the Future of Journalism in Pakistan* (blog). Digital Rights Foundation. https://digitalrightsfoundation.pk/when-reality-is-manufactured/ .

Media Support & Partners. (2023). *Countering Disinformation in Pakistan* (policy brief). Media Support.https://www.mediasupport.org/wp-content/uploads/2023/01/Countering Disinformation-in-Pakistan-2023.pdf.

Frontiers in Political Science. (2024). Shukla, A. K. (2024). *AI-generated misinformation in the election year 2024*. Frontiers. https://www.frontiersin.org/articles/10.3389/fpos.2024.1451601/full.

Voice of America. (2024, Feb 22). *Deepfakes, internet access cuts make election coverage hard, journalists say* (report on Pakistan 2024 election deepfakes). VOA. https://www.voanews.com/a/deepfakes-internet-access-cuts-make-election-coverage-hard-journalists-say-/7498917.html.

IFJ (International Federation of Journalists). (2025, Jun). Shah, A. A. *AI, Deepfakes, and the Fog of War Disinformation in the 2025 India-Pakistan conflict* (blog). IFJ.

Accountability Lab Pakistan. (2025, Aug). *The Disinformation Economy of Pakistan* (policy brief). Accountability Lab. https://pakistan.accountabilitylab.org/wp-content/uploads/2025/08/Disinformation-Economy-of-Pakistan.pdf.

Ejaz, W. (2025). *How effective are fact-checks in Pakistan?* (article). Taylor & Francis. https://doi.org/10.1080/1369118X.2024.2445636.

Poynter / Time and coalition reporting. (2024). *Global analysis: Tech companies are failing to keep elections safe* (analysis). TIME. https://time.com/6967334/ai-elections-disinformation-meta-tiktok/.

Javed, R. T., et al. (2021). *A deep dive into COVID-19-related messages on social media in Pakistan* (PMC article). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8590927/.

Haroon, D. (2021). *Investigating misinformation dissemination on social media* (arXiv preprint).

Romero-Moreno, F. (2025). *Deepfake detection in generative AI: A legal framework* (ScienceDirect).

Incident Database. (2024). *Many political deepfakes circulating in run-up to 2024 Pakistani general elections* (incident report). https://incidentdatabase.ai/cite/671/.

Vidgen, B., & Yasseri, T. (2018). *Detecting weak and emerging misinformation signals* (research). (Used to support detection literature).