Wah Academia Journal of Social Sciences
Volume 4, Issue 2, Dec 2025, pp. 1135-1150
ISSN – E 2958-8731 P 2958-8723
https://doi.org/10.63954/WAJSS.4.2.54.2025

Deepfakes and Democratic Discourse: Ethical Implications of AI-Generated Content in Pakistan's Digital Media Landscape

Dr. Taha Shabbir[1], Dr. Muhammad Aftab Madni[2] and Dr. Usman Farooq[3]

Abstract

The emergence of deepfake technology AI-generated synthetic media capable of fabricating convincing audio-visual content of real individuals poses grave challenges to democratic discourse worldwide. In Pakistan, where political polarisation is acute, media literacy is uneven, and digital governance frameworks remain nascent, deepfakes represent a particularly dangerous instrument of disinformation. This paper examines the ethical implications of AI-generated synthetic media for democratic processes, journalistic integrity, freedom of expression, and political communication in Pakistan. Drawing on international scholarship in media ethics, computational propaganda, political communication, and digital law, as well as documented instances of deepfake misuse in Pakistan's turbulent political environment, the paper argues that deepfakes are not merely a technical problem but a socio-political crisis requiring coordinated legal, journalistic, and civic responses. The paper further contends that existing frameworks—PECA 2016, PEMRA regulations, and Pakistan's nascent data protection legislation—are structurally insufficient to address the distinctive epistemic and democratic harms inflicted by synthetic media. A multi-stakeholder framework encompassing legislative reform, platform accountability, media literacy education, and independent forensic journalism infrastructure is proposed as a necessary path forward.

Keywords: deepfakes, synthetic media, democratic discourse, Pakistan, disinformation, PECA 2016, AI ethics, political communication, media literacy, computational propaganda

[1]Department of Computing, Faculty of Engineering Sciences & Technology, Hamdard University Karachi – Pakistan
[2]Department of Media and Communication Studies, Shaheed Benazir Bhutto University (SBBU), Shaheed Benazirabad (SBA) – Pakistan
[3]Department of Media Studies & Design, Faculty of Communication & Design, Indus University, Karachi – Pakistan

Introduction

Democracy depends on a shared epistemic common—a set of broadly agreed facts, verified information, and credible public discourse within which citizens can deliberate, evaluate candidates, and hold power to account. The advent of deepfake technology AI-generated synthetic media that can place real people's faces, voices, and mannerisms into fabricated scenarios with unprecedented verisimilitude strikes at the foundations of this epistemic commons with an intensity no previous disinformation technology has matched. Unlike doctored photographs or edited audio clips, which trained analysts can often detect, high-quality deepfakes generated by generative adversarial networks (GANs) and diffusion models are increasingly indistinguishable from authentic footage, even to expert observers (Chesney & Citron, 2019).

In Pakistan, this challenge is compounded by a constellation of structural vulnerabilities: a fragmented and politically compromised media landscape, acute partisan divisions, a WhatsApp-centric information ecosystem that amplifies viral content without verification, and regulatory institutions whose independence is frequently questioned. The country's history of political crises military interventions, judicial controversies, and electoral disputes provides fertile ground for the strategic deployment of synthetic media. Incidents of alleged deepfake misuse in Pakistani politics, documented in the run-up to and aftermath of the 2024 general elections, suggest that the technology has already entered the practical toolkit of political actors, though attribution remains contested (Human Rights Watch, 2024).

This paper argues that deepfakes in Pakistan's context are not simply a technological curiosity or a distant threat but a present and evolving danger to democratic integrity, journalistic credibility, individual dignity, and freedom of expression. The paper proceeds as follows: Section 2 reviews the technical development and typology of deepfake technology. Section 3 examines deepfakes in the global context of democratic harm. Section 4 analyses Pakistan's specific political and media environment. Section 5 reviews existing legal and regulatory frameworks and their adequacy. Section 6 discusses the ethical dimensions of deepfakes through multiple normative lenses. Section 7 proposes a multi-stakeholder remedial framework. Section 8 concludes.

Understanding Deepfakes: Technology, Typology, and Trajectory

❖ From GAN to Diffusion: The Technical Evolution

Deepfakes are synthetic media artifacts generated by artificial intelligence, specifically through architectures such as generative adversarial networks (GANs), variational autoencoders (VAEs), and, more recently, diffusion models. The term 'deepfake' originated on the Reddit platform around 2017, where a user applied deep learning to superimpose celebrity faces onto pornographic videos (Westerlund, 2019). The underlying GAN architecture, introduced by Goodfellow et al. (2014), involves two neural networks—a generator and a discriminator engaged in an adversarial training process that progressively improves the realism of generated output. By 2020, tools such as DeepFaceLab and FaceSwap had made face-swap deepfakes accessible to users without specialist technical knowledge. By 2023–2024, diffusion-model-based platforms such as Stable Diffusion and Midjourney had extended synthetic media capabilities to photorealistic image and video generation from text prompts, dramatically lowering the barrier to sophisticated forgery.

Audio deepfakes cloned voice recordings generated from brief samples of a target's speech have followed a parallel developmental trajectory. Systems such as ElevenLabs, Microsoft's VALL-E, and various open-source implementations can reproduce an individual's voice with high fidelity from as few as three seconds of audio. The combination of face-swap video and cloned audio produces what researchers term 'talking head' deepfakes: convincing video of a real person appearing to say things they never said (Chesney & Citron, 2019). The implications for political communication where a fabricated press conference, concession speech, or incendiary statement could alter the course of events are immediately apparent.

❖ A Typology of Deepfake Harms

Scholars have proposed several typologies of deepfake-related harms. Chesney and Citron (2019) distinguished between individual harms including non-consensual intimate imagery, reputational damage, and harassment and societal harms, including political disinformation, diplomatic crises, and the erosion of epistemic trust. Paris and Donovan (2019) introduced the concept of the 'liar's dividend': the paradox whereby the mere existence of deepfake technology allows authentic,

genuinely compromising footage to be dismissed as fabricated, providing bad actors with a plausible deniability shield irrespective of whether they have actually produced a deepfake. This epistemic uncertainty is arguably as corrosive as the deepfakes themselves.

For the purposes of this paper, four categories of deepfake harm are particularly relevant to Pakistan's democratic context: (a) electoral manipulation through fabricated statements attributed to candidates or officials; (b) targeted harassment and non-consensual intimate image abuse, disproportionately affecting women in public life; (c) the delegitimisation of authentic media through the liar's dividend; and (d) the amplification of sectarian and ethnic hate speech through synthetic content falsely attributed to community leaders.

Deepfakes and Democratic Harm: The Global Evidence Base

- **Electoral Manipulation**

Concerns about deepfakes in electoral contexts have been extensively theorised, though documented cases of decisive electoral manipulation by deepfakes remain relatively limited in comparison to the scale of the threat (Vaccari & Chadwick, 2020). Documented or credibly alleged instances include a 2018 video in Gabon in which President Ali Bongo's appearance was questioned, contributing to political instability; a 2022 deepfake of Ukrainian President Volodymyr Zelensky appearing to call for Ukrainian military surrender, circulated following Russia's invasion; and various synthetic audio clips attributed to political figures in Slovakia's 2023 election campaign (European Digital Media Observatory, 2023). In each case, the deepfakes achieved their destabilising purpose not necessarily by permanently deceiving audiences but by generating confusion, eroding trust, and consuming the attention and credibility resources of the targeted parties.

Vaccari and Chadwick (2020) conducted experimental research demonstrating that exposure to deepfake political videos increases political uncertainty a measure of citizens' confidence in their ability to determine what is true in politics even when subjects were informed the content might be manipulated. The effect was independent of the quality of the deepfake, suggesting that the category 'deepfake' itself, once salient in public consciousness, becomes a generalised epistemological hazard that diminishes trust in all political video content.

❖ Gender-Based Targeting and the Silencing of Women

The gendered dimension of deepfake harm deserves particular emphasis. Studies consistently show that the vast majority of deepfake content targets women, predominantly through non-consensual intimate imagery (Sensity AI, 2023). While this abuse affects private individuals and celebrities worldwide, women in politics and journalism face a specific threat: the use of deepfake pornographic imagery as a tool of political silencing, designed to damage reputations, induce self-censorship, and deter women from public life. Henry et al. (2021) documented the devastating psychological consequences of non-consensual image-based abuse, including anxiety, depression, social withdrawal, and professional harm. For female politicians and journalists in conservative societies such as Pakistan, where honour-based norms impose severe social and professional penalties for real or perceived sexual transgression, deepfake-facilitated image abuse represents an acutely powerful instrument of suppression.

❖ The Liar's Dividend and Epistemic Breakdown

Perhaps the most insidious democratic harm of deepfakes is the liar's dividend. As deepfake technology becomes publicly known, authentic video evidence of wrongdoing becomes easier to dismiss. In a political environment already saturated with accusations of media bias and fabrication, the existence of deepfake technology provides a ready-made defence for accused parties. This dynamic has particular salience in Pakistan, where politicians routinely dispute the authenticity of recordings whether phone calls, video meetings, or public statements that implicate them in misconduct. The deepfake era potentially converts every such recording into a contested artefact whose evidential value can be undermined by a simple allegation of AI fabrication, regardless of the actual provenance of the content (Chesney & Citron, 2019; Paris & Donovan, 2019).

Pakistan's Digital Media Landscape: Structural Vulnerabilities

❖ Political Polarisation and Media Partisanship

Pakistan's media ecosystem is characterised by intense political polarisation, institutional instability, and a history of state pressure on independent journalism. The country's television news landscape is dominated by private channels whose

ownership is often linked to business interests with political affiliations, producing coverage that critics describe as systematically partisan (Freedom Network, 2023). The Pakistan Electronic Media Regulatory Authority (PEMRA), nominally responsible for broadcast regulation, has been repeatedly accused of partisan enforcement suspending or fining channels critical of incumbent governments while permitting pro-government content to circulate without sanction (Committee to Protect Journalists, 2023).

The political crisis precipitated by the removal of Prime Minister Imran Khan through a parliamentary vote of no confidence in April 2022, his subsequent conviction and imprisonment, and the deeply disputed February 2024 general elections created a period of exceptional political turbulence. This environment characterised by mass mobilisation, intense media warfare between the Pakistan Tehreek-e-Insaf (PTI) and its opponents, and widespread allegations of state interference in the information space is precisely the kind of high-stakes, low-trust context in which deepfakes can achieve maximum destabilising effect.

- **WhatsApp as Primary Information Infrastructure**

A critical structural feature of Pakistan's information ecosystem is the centrality of WhatsApp as a medium for political communication and news sharing. With over 45 million active WhatsApp users in Pakistan and a culture of sharing viral content across family and community groups, information accurate or fabricated can reach mass audiences within hours through private, encrypted channels that evade both platform moderation and journalistic scrutiny (Digital Rights Foundation, 2023). This architecture is deeply inhospitable to fact-checking: content circulates in private groups where professional debunkers have no visibility, corrections rarely travel as far as the original misinformation, and the social trust embedded in family and community networks lends credibility to even implausible claims.

For deepfakes specifically, WhatsApp's infrastructure presents an ideal distribution mechanism. A fabricated video of a politician, cleric, or military figure can be produced cheaply, distributed anonymously through forwarded messages, and reach millions of recipients before forensic analysts can examine its provenance. The platform's end-to-end encryption, while essential for privacy and security, makes it impossible for platform operators or researchers to systematically monitor deepfake circulation, creating a significant accountability gap (Posetti & Matthews, 2018).

❖ Documented Deepfake Incidents in Pakistan

Several incidents involving alleged or confirmed deepfake or synthetic media manipulation have been documented in Pakistan's recent political history. In the lead-up to the 2024 elections, an audio clip purportedly featuring a conversation between a senior judge and a PTI leader was widely circulated; questions about its authenticity including the possibility of AI-assisted voice manipulation were raised without definitive forensic resolution (Geo News, 2023). PTI itself deployed what appeared to be an AI-generated video address by imprisoned former Prime Minister Imran Khan reportedly produced using publicly available footage and AI voice cloning to broadcast a campaign message during the election period, a deployment that simultaneously demonstrated the technology's accessibility and prompted regulatory confusion about whether such content violated election laws (Dawn, 2024).

Non-consensual synthetic intimate images targeting female journalists and activists have also been reported, though underreporting is widespread due to social stigma and inadequate legal redress mechanisms (Digital Rights Foundation, 2023). These incidents collectively suggest that Pakistan's deepfake threat is not hypothetical but emergent, and that the institutional capacity to detect, attribute, and respond to synthetic media manipulation is dangerously underdeveloped relative to the pace of technological change.

Legal and Regulatory Frameworks: Adequacy and Gaps

❖ PECA 2016 and Its Limitations

Pakistan's primary legislative instrument for digital offences is the Prevention of Electronic Crimes Act 2016 (PECA). PECA criminalises the transmission of information through electronic systems that is 'false' and intended to cause 'fear, panic, disorder or unrest' (Section 26), as well as offences related to unauthorised access, cyberterrorism, and dignity violations. Section 21 specifically addresses electronic forgery, making it an offence to produce or circulate forged electronic documents with intent to cause harm or injury. On its face, these provisions are potentially applicable to malicious deepfakes. However, PECA's adequacy for the deepfake era is severely circumscribed by several structural deficiencies.

First, PECA's definitional architecture predates the emergence of AI-generated synthetic media. The Act's references to 'forged' information and 'false' documents do not specifically contemplate content produced by generative AI systems, creating interpretive ambiguity about whether a deepfake which may not involve traditional forgery in the legal sense falls within existing offence categories. Second, the evidentiary burden of proving both the synthetic nature of content and the intent of its producer presents formidable practical obstacles, particularly given the absence of state-funded digital forensics infrastructure capable of deepfake attribution at scale. Third, and most critically, PECA has been systematically applied against journalists, activists, and political opponents rather than against disinformation actors, generating legitimate concerns that any deepfake-specific legislation would similarly become an instrument of political censorship rather than epistemic protection (Reporters Without Borders, 2023).

- **PEMRA and Broadcast Regulation**

PEMRA's regulatory remit covers broadcast and distribution licencees television and radio channels but its jurisdiction does not extend to online platforms, social media, or the encrypted messaging applications through which most deepfake content circulates in Pakistan. Even where broadcast media repeats deepfake content originally circulated online, PEMRA's enforcement record suggests inconsistent application of existing codes. The authority's content codes prohibit the broadcast of content that is 'defamatory, false, or unverified,' provisions that are in principle applicable to channels that broadcast deepfake material without appropriate disclosure. However, PEMRA's track record of applying these provisions on the basis of political rather than editorial criteria significantly undermines confidence in its role as a deepfake governance institution (Freedom Network, 2023).

- **Pakistan's Data Protection Landscape**

A Personal Data Protection Bill has been under deliberation in Pakistan for several years, with multiple drafts circulated and revised without legislative enactment as of early 2026. The absence of comprehensive data protection legislation creates an additional governance gap: the biometric and facial data that underpin deepfake generation can be harvested from social media profiles, electoral databases, and other sources without meaningful legal restriction. Comparative frameworks such as the European Union's GDPR, which constrains the processing of biometric data

as a 'special category' requiring explicit consent, or the proposed EU Artificial Intelligence Act's prohibition on certain high-risk AI applications including real-time biometric identification, provide models for the kind of data governance infrastructure Pakistan currently lacks (European Commission, 2021; Voigt & Von dem Bussche, 2017).

Ethical Dimensions: Multiple Normative Frameworks

- **Consequentialist Analysis**

From a consequentialist perspective, the ethical evaluation of deepfakes turns on an assessment of their net harms and benefits across affected populations. The harms are multiple and severe: erosion of epistemic trust in democratic institutions, targeted psychological harm to individuals depicted without consent, material damage to political processes through fabricated evidence, and the chilling effect on free expression produced by the threat of reputational deepfake attacks. The benefits are minimal in the political disinformation context some researchers note that creative deepfake applications in satire, education, or artistic expression have legitimate value, but these uses are categorically distinct from the weaponised synthetic media that concerns this analysis (Floridi et al., 2020). The consequentialist verdict is unambiguous: deepfakes deployed for political manipulation and non-consensual image abuse produce vast negative welfare consequences that no plausible countervailing benefit can justify.

- **Deontological Considerations**

Kantian and deontological ethics foreground the dignity of persons and the impermissibility of treating individuals merely as means to others' ends. Deepfakes violate the dignity of their subjects in a foundational way: they appropriate a person's likeness, voice, and identity their most intimate personal expressions and deploy these against the subject's will for purposes the subject would reject and that cause them harm. The fabrication of statements, actions, or expressions attributed to real persons without their consent is a profound violation of autonomy that cannot be redeemed by claims of political necessity or artistic license (Nissenbaum, 2010). The deontological prohibition is particularly forceful in cases involving non-consensual intimate deepfakes, where the violation of bodily autonomy and dignity is at its most acute.

Deontological analysis also has implications for the responsibilities of creators, distributors, and platform operators. Kant's universalisability test asking whether the maxim of an action could serve as a universal law produces a clear negative verdict on deepfake production for manipulation: a world in which it is universally permissible to fabricate synthetic media of political opponents would be a world in which democratic deliberation is impossible, a result that self-defeats the political purposes the deepfake producer seeks to advance.

- **Virtue Ethics and Journalistic Integrity**

Virtue ethics asks what a person of good character in the journalistic context, a journalist embodying professional virtues of accuracy, fairness, truth-telling, and public service—would do when confronted with synthetic media. The answer is clear: responsible journalists verify content before publication, disclose uncertainties, protect subjects from unjust reputational harm, and treat audiences as epistemic agents deserving accurate information rather than as targets for manipulation. The proliferation of deepfakes intensifies these professional obligations. News organisations in Pakistan and globally need institutionalised protocols for synthetic media verification, including access to forensic tools, clear editorial standards for the publication of contested audiovisual content, and transparency with audiences about verification processes (Ward, 2015).

- **Ethics of Care and Vulnerability**

An ethics of care perspective emphasises the relational context of harm and foregrounds the vulnerability of those most exposed to deepfake abuse. In Pakistan, several groups are particularly vulnerable: women in public life, subject to gender-based deepfake targeting; religious minorities, whose community leaders could be targeted with fabricated provocative statements to incite sectarian violence; and political dissidents and journalists who already operate in conditions of surveillance and suppression. An ethical framework adequate to Pakistan's context must centre the protection of these vulnerable populations, recognising that deepfake harm is not uniformly distributed but is concentrated in communities that already bear disproportionate burdens of systemic oppression (Noddings, 2013).

Towards a Multi-Stakeholder Remedial Framework

- **Legislative Reform**

Pakistan requires deepfake-specific legislation that clearly defines AI-generated synthetic media, establishes graduated criminal liability for malicious use, and incorporates adequate procedural safeguards against abuse by state actors. Drawing on comparative models—including the European Union's Artificial Intelligence Act's risk-based regulatory approach, Singapore's Protection from Online Falsehoods and Manipulation Act, and the United States' DEEPFAKES Accountability Act (proposed) Pakistani legislators should consider provisions requiring disclosure labelling for synthetic media in political advertising, establishing civil liability for non-consensual intimate deepfakes, and creating an independent digital forensics commission with the technical capacity and institutional independence to assess contested media authenticity (Chesney & Citron, 2019; European Commission, 2021).

Critically, any such legislation must incorporate robust safeguards against politicised enforcement. Independent judicial oversight of investigations, explicit carve-outs for satire and artistic expression, and meaningful whistleblower protections are necessary to ensure that deepfake laws do not become instruments of political censorship in the tradition of PECA's historical application. Civil society organisations, press freedom bodies, and the legal profession must be central stakeholders in the legislative drafting process.

- **Platform Accountability**

Major social media platforms operating in Pakistan Meta (WhatsApp and Facebook), YouTube, TikTok, and X bear significant responsibilities for the deepfake ecosystem. While end-to-end encryption in WhatsApp creates genuine technical barriers to content surveillance, platforms can still contribute to mitigation through investment in deepfake detection tools deployed at upload points for non-encrypted content, mandatory provenance labelling for AI-generated content in political advertising (as Meta has announced for its platforms in various markets), expedited reporting and removal mechanisms for non-consensual synthetic intimate images, and meaningful transparency reporting about synthetic media volumes and enforcement actions in Pakistan specifically (Rini, 2017).

Pakistan's government and civil society should engage platforms through structured multi-stakeholder dialogues to secure context-sensitive commitments that account for Urdu-language and regional language content, which platform moderation systems currently handle with significantly less accuracy than English-language material. The extension of the EU's Digital Services Act framework's requirements for systemic risk assessment to non-EU jurisdictions through diplomatic engagement and the leveraging of market access considerations represents one possible mechanism for securing greater platform accountability.

- **Deepfake Detection and Forensic Journalism Infrastructure**

Pakistan's journalistic sector urgently needs investment in synthetic media literacy and forensic verification capacity. This includes training programs for journalists in the use of deepfake detection tools such as Microsoft's Video Authenticator, the Content Authenticity Initiative's open-source implementations, and academic detection systems developed by research groups including those at Dartmouth, UC Berkeley, and MIT as well as the development of institutional protocols for the verification of contested audiovisual content (Farid, 2022). Collaborative fact-checking networks such as Soch Fact Check and Geo Fact Check have made significant contributions to Pakistan's verification ecosystem but require expanded resources and technical partnerships to address synthetic media at scale.

Academic institutions including universities with computer science and media studies programs have a role to play in building local deepfake detection research capacity. The development of detection models trained on Pakistani political figures' faces and voices, drawing on publicly available video and audio datasets, would be a significant contribution to the country's epistemic resilience. Partnerships between Pakistani universities, international research institutions, and civil society organisations could provide the funding, expertise, and institutional support for such initiatives.

- **Media Literacy and Civic Education**

Sustained investment in media literacy education is a necessary complement to legal and technical interventions. Citizens who understand that synthetic media exists, who have been exposed to examples of deepfakes and their detection, and who have internalised verification habits pausing before sharing, checking source provenance,

seeking corroboration are substantially more resistant to deepfake manipulation (Pennycook & Rand, 2021). Pakistan's formal education system, from secondary through tertiary levels, offers a scalable platform for media literacy integration. The Higher Education Commission of Pakistan and provincial education departments should incorporate digital media literacy including synthetic media awareness—into curricula across disciplines.

Community-based media literacy programmes, particularly those targeting older demographics who are heavy WhatsApp users but may be less familiar with AI capabilities, can complement formal education. Religious institutions, civil society organisations, and community radio networks given their reach in rural and peri-urban areas underserved by formal educational infrastructure—are important partners in such initiatives. The Digital Rights Foundation's Hamara Internet programme and similar civil society initiatives provide models for community-based digital literacy work that can be adapted and scaled to address synthetic media threats specifically.

Conclusion

Deepfakes represent a convergence of technological capability and democratic vulnerability that Pakistan is poorly equipped to manage. The country's fragmented media landscape, polarised political environment, WhatsApp-centric information ecosystem, and underdeveloped regulatory frameworks collectively constitute conditions of exceptional susceptibility to synthetic media manipulation. The documented and alleged incidents of deepfake use in Pakistan's recent political history are harbingers of a more severe and systematic threat that will only intensify as generative AI technologies become cheaper, more powerful, and more widely accessible.

The ethical dimensions of this threat are multiple and serious. Deepfakes violate individual dignity, undermine democratic deliberation, silence marginalised voices, and corrupt the epistemic foundations of public life. No single intervention is adequate to address these harms; what is required is a coordinated multi-stakeholder response integrating legislative reform, platform accountability, forensic journalism infrastructure, and media literacy education. This response must be designed to protect vulnerable populations women, minorities, journalists, and political dissidents who bear the greatest burden of deepfake harm, while

incorporating safeguards against the instrumentalisation of anti-deepfake measures for political censorship.

The integrity of Pakistan's democratic processes, fragile as they are, depends in part on the existence of a shared informational environment in which citizens can access and evaluate evidence, candidates can be held accountable for their actual statements and actions, and journalism can perform its watchdog function without being overwhelmed by a flood of synthetic fabrication. Defending this environment against deepfakes is not a technical challenge that can be delegated to engineers and regulators alone. It is a civic, ethical, and political imperative that demands the sustained engagement of scholars, journalists, civil society, and democratic institutions.

References

Chesney, R., & Citron, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. California Law Review, 107(6), 1753–1820. https://doi.org/10.15779/Z38RV0D15J

Committee to Protect Journalists. (2023). Pakistan: Attacks on the press. CPJ. https://cpj.org/asia/pakistan/

Dawn. (2024, February 7). PTI uses AI-generated video of Imran Khan to broadcast election message. Dawn. https://www.dawn.com/news/1810000

Digital Rights Foundation. (2023). Hamara internet annual report 2023: Mapping digital violence against women in Pakistan. Digital Rights Foundation. https://digitalrightsfoundation.pk/

European Commission. (2021). Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). COM/2021/206 final. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206

European Digital Media Observatory. (2023). Deepfakes and elections: Risks and regulatory responses. EDMO. https://edmo.eu/publication/deepfakes-and-elections/

Farid, H. (2022). Creating, using, misusing, and detecting deep fakes. Journal of Online Trust and Safety, 1(4). https://doi.org/10.54501/jots.v1i4.56

Floridi, L., Cowls, J., King, T. C., & Taddeo, M. (2020). How to design AI for social good: Seven essential factors. Science and Engineering Ethics, 26(3), 1771–1796. https://doi.org/10.1007/s11948-020-00213-5

Freedom Network. (2023). Annual report on media freedom in Pakistan. Freedom Network Pakistan. https://www.freedom.net.pk/

Geo News. (2023, November 12). Audio leak controversy: Questions about authenticity remain unanswered. Geo News. https://www.geo.tv/latest/audio-leaks

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. Advances in Neural Information Processing Systems, 27. https://proceedings.neurips.cc/paper/2014/hash/5ca3e9b122f61f8f06494c97b1afccf3-Abstract.html

Henry, N., Flynn, A., & Powell, A. (2021). Technology-facilitated domestic abuse and coercive control. Violence Against Women, 27(15–16), 2799–2823. https://doi.org/10.1177/1077801220975078

Human Rights Watch. (2024). Pakistan: Widespread interference in general elections. HRW. https://www.hrw.org/report/2024/pakistan-elections

Nissenbaum, H. (2010). Privacy in context: Technology, policy, and the integrity of social life. Stanford University Press.

Noddings, N. (2013). Caring: A relational approach to ethics and moral education (2nd ed.). University of California Press.

Paris, B., & Donovan, J. (2019). Deepfakes and cheap fakes: The manipulation of audio and visual evidence. Data & Society Research Institute. https://datasociety.net/library/deepfakes-and-cheap-fakes/

Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. Trends in Cognitive Sciences, 25(5), 388–402. https://doi.org/10.1016/j.tics.2021.02.007

Posetti, J., & Matthews, A. (2018). A short guide to the history of 'fake news' and disinformation. International Center for Journalists. https://www.icfj.org/our-work/short-guide-history-fake-news-and-disinformation-new-icfj-learning-module

Reporters Without Borders. (2023). Pakistan: Press freedom index 2023. RSF. https://rsf.org/en/country/pakistan

Rini, R. (2017). Fake news and partisan epistemology. Kennedy Institute of Ethics Journal, 27(2), E-43–E-64. https://doi.org/10.1353/ken.2017.0025

Sensity AI. (2023). The state of deepfakes: Landscape, threats, and impact—2023 report. Sensity AI. https://sensity.ai/reports/

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). https://doi.org/10.1177/2056305120903408

Voigt, P., & Von dem Bussche, A. (2017). The EU General Data Protection Regulation (GDPR): A practical guide. Springer.

Ward, S. J. A. (2015). Radical media ethics: A global approach. Wiley-Blackwell.

Westerlund, M. (2019). The emergence of deepfake technology: A review. Technology Innovation Management Review, 9(11), 39–52. https://doi.org/10.22215/timreview/1282

Article Information:

Received	*10-Sept-2025*
Revised	27-Nov-2025
Accepted	10-Dec-2025
Published	15-Dec-2025

Declarations:

Authors' Contribution:

- All Authors **Conceptualization, and intellectual revisions. Data collection, interpretation, and drafting of manuscript**
- The authors agree to take responsibility for every facet of the work, making sure that any concerns about its integrity or veracity are thoroughly examined and addressed

- **Conflict of Interest**: NIL
- **Funding Sources**: NIL

Correspondence:

Dr. Taha Shabbir

taha.shabbir@hamdard.edu.pk