INTERNATIONAL JOURNAL OF SOCIAL SCIENCES BULLETIN
ISSN: (E) 3007-1917 (P) 3007-1909
Volume 4, Issue 5, 2026

DEEPFAKE MEDIA AND THE CRISIS OF VISUAL TRUTH: IMPACT ON POLITICAL TRUST AND MEDIA CREDIBILITY IN PAKISTAN

Shahzad Hussain Shah[*1], Aleena Kanwal[2], Dr Ali Hassan[3], Hira Zahid[4]

*[*1]Copy Writer, Punjab Health and Wellness Radio Station, Health & Population Department, Government of the Punjab, PhD Scholar, Department of Communication & Media Studies, University of Sargodha*
[2]PhD Scholar Department of Communication & Media Studies, University of Sargodha
[3]Lecturer Department of Communication & Media Studies, University of Sargodha
[4]Content Writer Social Media & Marketing, Punjab Health and Wellness Radio Station, Health & Population Department, Government of the Punjab, Department of Communication & Media Studies, University of Lahore

[*1]shahzadshah1130@gmail.com, [2]aleenakanwal07@gmail.com, [3]ali.hassan@uos.edu.pk, [4]Hiraazahid900@gmail.com

Corresponding Author: *
Shahzad Hussain Shah
DOI: https://doi.org/10.5281/zenodo.20052111

Received	Accepted	Published
11 March 2026	21 April 2026	06 May 2026

ABSTRACT

The rapid development of deepfake technology has created substantial epistemological challenges for the political communication research domain, as concerns the forthcoming emergent liberal democratic systems, a category within which Pakistan is located. The research examined how exposure to deepfake media affects Pakistani citizens' political trust and their evaluations of media credibility. The study obtained data using a cross-sectional quantitative survey design, involving 500 adult participants from four Pakistani provinces via stratified random sampling. The researchers used validated measurement tools to assess deepfake exposure, political trust, media credibility, digital media literacy, information skepticism, and participants' sociodemographic characteristics. The researchers applied hierarchical multiple regression analyses, together with the PROCESS Macro Model 1, to assess the moderated regression results. The research results indicated that deepfake exposure decreased both political trust and media credibility assessments, with deepfake exposure as the main factor driving the decline (β = -.36, $p < .001$) and (β = -.34, $p < .001$). The researchers discovered that the moderating variable was digital media literacy whose effect was positive and non-significant in lowering political trust ($\Delta R2$ =.02, p =.003) and that the effect of deepfake exposure was not so strong among individuals with high media literacy. The process of information skepticism functioned as an essential mediating pathway. The research findings demonstrate that Pakistan needs immediate national media literacy initiatives and synthetic media control systems and methods for rebuilding institutional trust in its digital information ecosystem.

Keywords: *deepfake, political trust, media credibility, Pakistan, digital media literacy, misinformation, synthetic media, epistemological crisis*

INTRODUCTION

The introduction of generative artificial intelligence has brought about an unparalleled change in how people understand visual content. Deepfake technology enables users to create synthetic media that utilizes deep learning

algorithms for transforming a person's physical appearance, vocal patterns, and body movements through the application of Generative Adversarial Networks (GANs), while politicians now use this technology as their main tool for producing fake news content (Chesney & Citron, 2019). The process of technological development directly threatens democratic governance because people must trust political communication to believe in democratic systems.

Pakistan provides a crucial setting for carrying out this study. Pakistan is one of the fastest-growing digital marketplaces in South Asia, with an expected 87 million Internet users and over 191 million mobile phone users in 2023, according to the Pakistan Telecommunication Authority [PTA]. People are informed about political news through social media such as Facebook, YouTube, Tik Tok and Whatsapp. The social media platforms operate their content distribution systems using synthetic media technology, which allows them to establish flexible content guidelines. Some of the high-profile deepfake cases have occurred in the country over the last few years, like manipulated video clips of high-profile political leaders, fake media to conduct psychological warfare between opposing political groups, and fake audio recordings to influence people during elections (Abbas et al., 2025).

The researchers explore three related research issues by analyzing three research questions that revolve around the following topics: (1) How much exposure of Pakistani individuals to deepfake political media is associated with their declining levels of political trust? (2) How do deepfake exposure and media credibility perceptions relate to each other? (3) Is digital media literacy an intervener in this relationship? The study of these concepts establishes foundational knowledge that supports both academic research and practical application in media studies, political communication, and Global South governance studies.

Significance of the Study

The international literature now demonstrates, with growing evidence, that synthetic media have negative effects on political attitudes (Vaccari & Chadwick, 2020; Hameleers et al., 2022), yet no empirical research has been conducted in South Asia or Pakistan. The Pakistani media system operates through distinct structural elements, which include a divided partisan press, a history of censorship, weak institutional trust, high smartphone usage, and low official media literacy rates. The study establishes a major gap between current regional and international knowledge of deepfakes and our understanding of their impact on democratic trust.

Literature Review

Deepfake Technology: Origins and Mechanisms

Deepfake is a portmanteau of the terms deep learning and fake, popularized by GAN-generated face-swapping content that became popular on social media platforms in 2017 (Kietzmann et al., 2020). According to Tolosana et al. (2020), deepfakes are developed based on encoder-decoder neural network models to produce photorealistic human facial expressions with Variational Autoencoders (VAEs). The system produces realistic human facial expressions, vocal sounds, and body movements. The Face2Face framework (Thies et al., 2016) and subsequent commercial tools such as DeepFaceLab and Stable Diffusion have democratized the creation of convincing synthetic content, lowering production barriers to the point where a single individual with consumer-grade hardware can produce broadcast-quality fabrications (Westerlund, 2019).

Deepfakes serve as an epistemic weapon that exceeds their value as a technology demonstration. Chesney and Citron (2019) demonstrate that deepfakes introduce a 'liar's dividend' which allows people to dismiss authentic compromising content as fake while they falsely present fabricated content as real. The study identifies synthetic media through its dual mechanism, which creates an 'epistemic double bind' situation for people who examine synthetic media.

Deepfakes in Political Communication

The intersection of deepfake technology and political communication is among the most dangerous uses of AI-generated media content

(Floridi et al., 2018). Empirical studies have documented deepfakes' capacity to distort political reality through multiple mechanisms. The United Kingdom deepfake video experiment conducted by Vaccari and Chadwick (2020) with 2005 participants showed that people who watched deepfake videos of British political leaders lost their ability to trust online political videos. Participants who watched the video but failed to recognise its falsity experienced greater uncertainty because their visual trust in the content had been disrupted.

Hameleers et al. (2022) extended these findings in a cross-national experimental study across four European countries, demonstrating that exposure to deepfake political content reduced perceived credibility of mainstream news media beyond the specific fabricated content encountered. The Pakistani context is particularly affected by the 'spillover' effect, which reduces media trust because Pakistani mainstream media already operate at a low level of credibility (Sheikh et al., 2024).

The South Asian context saw deepfake technology used in Indian elections, according to Kashyap (2025), while Moroojo et al. (2025) presented preliminary evidence of deepfake-driven erosion of trust in Pakistan's 2018 and 2022 elections. The present research study fills this gap in the literature, as prior studies have not provided sufficient quantitative data.

Political Trust: Theoretical and Empirical Foundations

Before citizens can trust political organizations and its members, they must believe that the political entities are good or capable of doing so. Dalton (2004) conducted extensive cross-national research, which showed that Western democracies experienced a permanent decline in institutional trust during the second half of the twentieth century. Trust in Pakistani politics has remained low because military dictatorship, judicial activism, executive corruption scandals, and media sensationalism have created short-lived cycles of trust (Waseem, 2022).

Multiple studies indicate that political trust decreases when the public encounters fake content, including all forms of misinformation, from fake news to deepfakes. The trust-in-media model (Metziger et al., 2003) posits that citizens form their trust in political institutions through their current evaluation of media credibility, resulting in a loss of institutional trust when media credibility declines. The generalized distrust hypothesis (Benkler et al., 2018) asserts that people who frequently see fake political content will develop general epistemic distrust, which makes them unable to trust actual information sources.

The Credibility of Media in the Digital Age

The digital age has brought substantial transformations to two aspects of media credibility, which determine how audiences trust media content, and to the actual media content that people consider trustworthy. The researchers Flanagin and Metzger created a fundamental principle that helps to distinguish between two types of credibility assessment, medium credibility and source credibility, but this principle received substantial criticism because of modern developments in user-generated content, algorithmic curation, and synthetic media. The media's credibility in the Pakistani context is also affected by structural aspects of ownership concentration, pressure from the Pakistan Electronic Media Regulatory Authority (PEMRA), and yellow journalism (Aftab, 2025).

Recent studies indicate a strong negative correlation between exposure to fabricated news and audience ratings of the media's trustworthiness (Lazer et al., 2018; Roozenbeek et al., 2022). The proposed research paper hypothesizes that the most cognitively disruptive type of synthetic media is deepfakes, which have more negative impacts on media credibility and political trust than any other form of fake content.

Digital Media Literacy as a Moderating Factor

Digital media literacy refers to the skill of reading, analyzing, evaluating, producing, and critically interpreting digital media content has always served as a buffer, helping people resist the negative influence of fake information (Pennycook and Rand, 2019). Interventions based

on inoculation theory (Lewandowski et al., 2020) and prebunking (Roozenbeek et al., 2022) have shown that educating individuals on how to manipulate content in synthetic media can make them less susceptible to such information.

Formal media literacy education is underdeveloped in Pakistan and is primarily focused on higher education institutions in urban areas (Wahid, 2024). Such a structural deficit implies that the moderating effect of digital media literacy may be especially significant in the Pakistani context, where people differ markedly in their ability to critically assess digital media.

Mythology and Narrative of Deepfake Disinformation

The existence of preexisting cultural myths and political discourses that affect attitudes toward synthetic content is one such facet of deepfake deception in Pakistan that has not been thoroughly examined. A second-order semiological system that converts transient ideological beliefs into generally accepted social norms is established by Roland Barthes' groundbreaking study of myth (1972). The study shows that deepfake technology's persuasiveness stems not from its sophisticated technological capabilities but rather from its capacity to validate preexisting mythical patterns. Synthetic misinformation, which identifies narrative components that pose a threat to national security and attempts to validate them with false material, thrives in Pakistani political culture.

The relationship between deepfake technology and cultural mythology shows that, because these technologies are employed in specific social and cultural contexts, researchers should take ideology into account. Because it draws into preexisting cognitive schemas (Bartlett, 1932) and confirmation bias processes (Nickerson, 1998), false material that appeals to political mythology can have disproportionately harsh, trust-corrupting effects.

Theoretical Framework

The current research combines three complementary theoretical frameworks to explain the mechanisms by which deepfake exposure affects political trust and media credibility.

Third-Person Effect Theory (TPE)

Third-Person Effect Theory, originally postulated by Davison (1983) and later operationalized as a communication research paradigm by Gunther (1991), holds that people systematically perceive media messages as having more influence on other people (the perceptual component) than on themselves (the behavioral component). Within the deepfake framework, TPE predicts that citizens will perceive deepfakes as more harmful when they target less educated or less politically savvy citizens, but will underestimate their vulnerability to misinformation. This perceived vulnerability asymmetry has significant consequences for political trust: people who think that others are more prone to deepfakes might develop generalized distrust of political communication without subjecting their own information processing to the same level of critical evaluation.

Cultivation Theory

The initial theory, the Cultivation Theory by George Gerbner (1969), posits that excessive television viewing fosters distorted impressions of social reality that align with the television world. In its online amplification (Morgan et al., 2018), cultivation theory proposes that extensive use of social media spaces filled with deepfake posts and synthetic disinformation will foster a sense of a mean world of perceptions of political life, which is the generalized cynicism, distrust of institutions, and the attitude that all political actors are universally deceptive. This theoretical prism offers an explanation of why exposure to deepfakes is not just episodic, but its effects are built on the long-term loss of trust in politics.

Epistemic Trust Framework

Originally created in the spheres of developmental psychology and cognitive science, the Epistemic Trust Framework (Sperber et al., 2010; Fonagy and Allison, 2014) assumes that epistemic trust, the inclination to accept the communicated knowledge as true, relevant, and trustworthy, is a basic cognitive ability that supports social learning

and institutional trust. In the context of deepfakes, the framework indicates that consistent exposure to unproven fake media essentially destroys the epistemic trust processes by which citizens evaluate political information. Once people are no longer able to effectively distinguish genuine from fake political information, the default epistemic trust heuristic (seeing is believing) is fundamentally compromised, resulting in the so-called epistemic hypervigilance condition of endemic uncertainty and distrust that undermines not only media use but also political involvement.

The combined conceptual model, which informs the study's hypotheses, is shown in Figure 1.

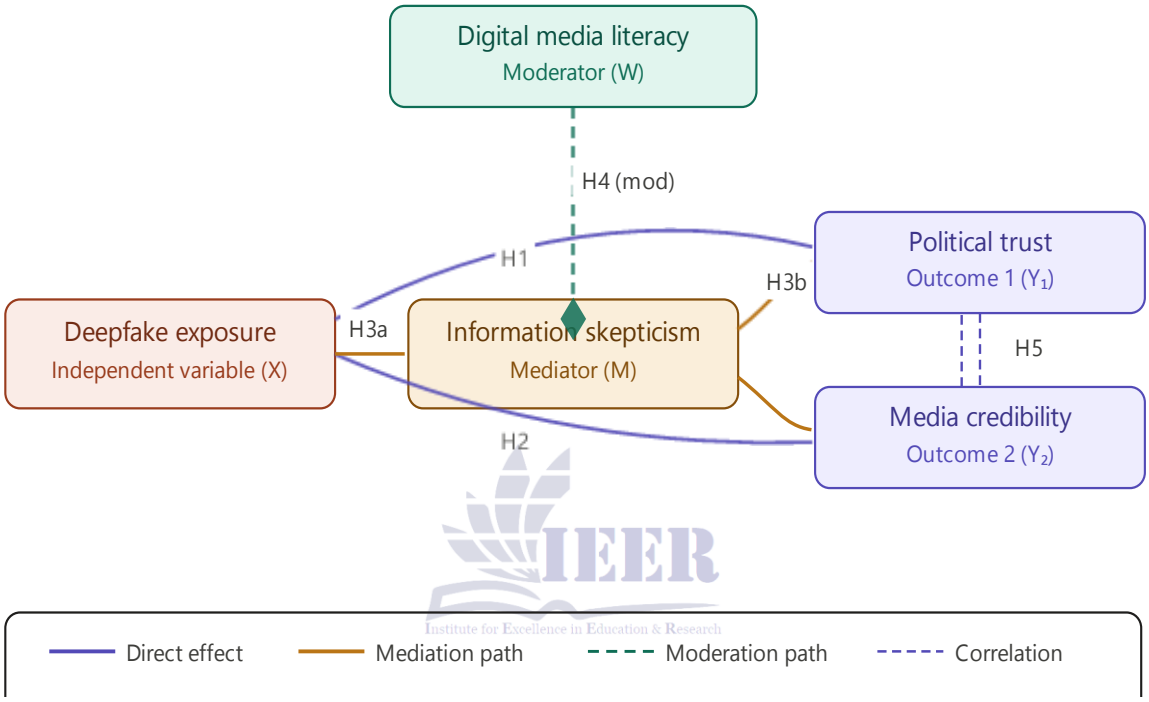

Figure 1. Conceptual Model: Deepfake Exposure → Political Trust & Media Credibility (Moderated by Digital Media Literacy; Mediated by Information Skepticism)

Hypotheses

The theoretical framework and literature review suggest the following hypotheses:

H1: The exposure to deepfakes will have a significant and negative correlation with political trust.

H2: Media credibility perceptions will have a significant and negative relationship with Deepfake exposure.

H3: The relationship between deepfake exposure and political trust will be mediated by information skepticism.

H4: The negative correlation between deepfake exposure and political trust will be mediated by digital media literacy, where an increase in digital media literacy will reduce the negative impact.

H5: Media credibility and political trust will be strongly and positively related.

Methodology

Research Design

The quantitative cross-sectional survey was applied in the study. The researchers employ cross-sectional designs to examine relationships among variables, as they can test their research hypotheses within a single time frame. The survey was conducted over several months, from March to June 2024, using a quantitative cross-sectional research design. These designs are primarily used

to examine relationships among variables, as they allow the researcher to conduct a series of analyses at a given time to test their hypothesis.

Participants and Sampling

The research team used a sampling approach to ensure diversity and representativeness of the sample across major demographic variables, such as age, gender, and education level. The sample was obtained in the four Pakistani provinces-Punjab, Sindh, Khyber Pakhtunkhwa, and Balochistan. They were required to meet the eligibility criteria, which included being a Pakistani citizen aged 18 years or older, having a secondary-level education, and having access to the internet. The team used three different participant recruitment methods, including online panel sites, community organizations, and academic research networks. The study retained N=500 valid survey responses after researchers removed 47 incomplete and suspicious responses, which they identified through attention-check items and response-time analyses. The Results section provides a detailed presentation of the final sample.

An a priori power analysis of the study using G*Power 3.1 (Faul et al., 2009) indicated that 500 participants were needed to achieve 99% power. The analysis identified a medium effect size (f2 = .15) with a = .05, which is correct, as the conventional 0.80 threshold is below this value.

Measures Deepfake Exposure Index (DEI)

The authors sought to gather self-reported data between 2023 and 2023 on the frequency with which people watched deepfake videos and fake political media. The researchers developed a 5-item scale for the study and tested it (e.g., "I have seen video clips"). The items were rated on a 5-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree). The scale demonstrated strong internal consistency, with a Cronbach's alpha of 0.812.

The political trust assessment used an 8-item adaptation of the Political Trust Battery, which Hetherington developed in 1998 and which researchers tested in earlier studies on Pakistan. The questions measure the level of trust that people have in the national government, parliament, judiciary, and political parties (e.g., 'I believe the Pakistani government is generally doing what is best in the interests of ordinary citizens; I believe that elections in Pakistan are a reflection of the will of the people'). Ratings of the items were out of 5 points. The scale showed good consistency (α = .871).

Media Credibility Perception (MCP)

The credibility of the media was assessed using a 7-item scale modified from Flanagin and Metzker (2000) and revised to reflect the digital nature of Pakistani media. Questions measured the perceived accuracy, fairness, and completeness of mainstream news media (e.g., 'Pakistani television news channels tend to present the news in an accurate way; Internet news sources in Pakistan present the political events in a way that is both balanced and fair). Internal consistency was good (α = .849).

Digital Media Literacy (DML)

Digital media literacy was quantified through the 6 items based on the Media Literacy Cognitive Framework proposed by Potter (2004) which included items on critical-evaluation of online sources, awareness of digital-manipulations techniques, and verification behaviors (e.g., I check multiple sources before believing a political video or image I see online; I am aware of how AI can be used to produce realistic fake video). The degree of internal consistency was satisfactory (α =.834).

Information Skepticism Scale (ISS)

An Information Skepticism Scale was used to measure the extent to which the participants hold a persistent sense of skepticism about the veracity of information about politics they encounter online (e.g., I generally do not trust political content I see on social media without checking it out; I assume that political videos and images can be manipulated until proven otherwise). Internal consistency was good (α = .857).

Procedure

The survey was conducted by research assistants under institutional ethics approval (ref: PU-IEC/2024/0117) in Urdu and English versions (back-translated and checked against accepted versions). The online surveys were sent via a pilot-tested Google Forms link, and paper-based surveys were administered in the community to reach participants with limited access to digital devices. Informed consent was given by all subjects. The questionnaire was completed in about 20-25 minutes, and no monetary reward was given to the respondents.

Data Analysis

Data analysis was performed using IBM SPSS Statistics Version 28.0. Initial tests involved checking for missing data, detecting outliers, and checking for parametric distributional assumptions. All primary variables were computed as descriptive statistics and Pearson correlation matrices. All multi-item scales were calculated to determine Cronbach's alpha. Using hierarchical multiple regression, the primary hypotheses were tested. Moderation was also analyzed using Hayes's (2017) PROCESS Macro (Model 1), with 5,000 bootstrap samples to determine the confidence interval. All the confidence intervals reported are 95% level.

Results

Sample Characteristics

Table 1 is an overview of 500 inhabitants of Punjab, the most populated province of Pakistan and their demographic data. The sample was made up of 55.6 percent men and 34.4 percent women with most of them being aged 25-34 years. The largest percentage of the respondents was represented by the ones with undergraduate degrees (43.8 of the respondents who had an education). The greatest percentage of respondents is in Punjab, the most populous province in Pakistan with 39.6.

Table 1
Sociodemographic Characteristics of the Sample (N = 500)

Variable	Category	Frequency (n)	Percentage (%)
Gender	Male	278	55.6
	Female	201	40.2
	Non-binary/Other	21	4.2
Age Group	18–24	138	27.6
	25–34	172	34.4
	35–44	109	21.8
	45–54	56	11.2
	55+	25	5.0
Education	Secondary or below	67	13.4
	Undergraduate	219	43.8
	Graduate (Master's)	158	31.6
	Doctoral	56	11.2
Province	Punjab	198	39.6

	Sindh	131	26.2
	KPK	97	19.4
	Balochistan	49	9.8
	Other/AJK/GB	25	5.0
Total		**500**	**100.0**

Note. Percentages may not sum to 100 due to rounding.

Deepfake Awareness and Exposure

Table 2 presents descriptive findings concerning participants' awareness of and exposure to deepfake technology. A substantial majority of participants (78.4%) reported awareness of deepfake technology, and 69.2% reported having encountered what they believed to be deepfake or manipulated political video content on social media. However, only 31.8% reported being able to identify manipulated media on first viewing, reflecting low detection competency. Concerningly, 44.6% reported having shared political content that they later suspected or discovered to be manipulated.

Table 2
Deepfake Awareness and Exposure Among Participants (N = 500)

Variable / Item	Mean (SD)	Median	% Agree / Yes
Awareness of deepfake technology	3.91 (0.87)	4.00	78.4%
Encountered deepfake political video	—	—	69.2%
Able to identify deepfake on first view	—	—	31.8%
Frequency of social media news consumption (daily)	4.12 (0.79)	4.00	82.6%
Confidence in detecting manipulated media (1–5)	2.58 (1.02)	3.00	—
Shared content later found to be a deepfake	—	—	44.6%

Note. Mean and SD were reported on 5-point Likert scales where applicable. Percentage values reflect affirmative response rates for binary items.

Descriptive Statistics and Internal Consistency

Table 3 presents means, SDs, Cronbach's alpha reliability coefficients, and confidence intervals for all important variables in the research. The average scores for Deepfake Exposure (M = 3.44) and Information Skepticism (M = 3.68) are above average, indicating that participants were relatively exposed, had a stable level of skepticism towards information, and did not change their attitudes towards information significantly. The scales of measurement were satisfactory to strong in internal consistency, with the Cronbach alpha ranging from 0.801 to 0.871.

Table 3
Descriptive Statistics and Internal Consistency of Primary Variables (N = 500)

Scale / Construct	Items	Mean	SD	α	95% CI
Political Trust in Government (PTG)	8	2.74	0.93	.871	[2.66, 2.82]
Media Credibility Perception (MCP)	7	2.61	0.88	.849	[2.53, 2.69]
Deepfake Exposure Index (DEI)	5	3.44	0.82	.812	[3.37, 3.51]
Digital Media Literacy (DML)	6	3.17	0.91	.834	[3.09, 3.25]
Information Skepticism Scale (ISS)	5	3.68	0.86	.857	[3.60, 3.76]
Emotional Distress (Media Anxiety)	4	3.29	0.97	.801	[3.20, 3.38]

Note. PTG = Political Trust in Government; MCP = Media Credibility Perception; DEI = Deepfake Exposure Index; DML = Digital Media Literacy; ISS = Information Skepticism Scale. All scales: 1 (low) to 5 (high). CI = confidence interval. α = Cronbach's alpha.

Correlation Analysis

Table 4 presents the Pearson correlation matrix for all primary variables. Consistent with H1 and H2, deepfake exposure was significantly and negatively correlated with both political trust ($r = -.52$, $p < .001$) and media credibility ($r = -.48$, $p < .001$). Political trust and media credibility were positively and strongly correlated ($r = .61$, $p < .001$), supporting H5. Digital media literacy showed positive correlations with trust and negative correlations with the impact of deepfake exposure and information skepticism, as expected given the hypothesis of a moderating effect.

Table 4
Pearson Correlation Matrix for Primary Study Variables (N = 500)

Variable	1. PTG	2. MCP	3. DEI	4. DML	5. ISS
1. Political Trust (PTG)	—				
2. Media Credibility (MCP)	.61**	—			
3. Deepfake Exposure (DEI)	-.52**	-.48**	—		
4. Digital Media Literacy (DML)	.38**	.41**	-.29**	—	
5. Info. Skepticism (ISS)	-.44**	-.39**	.55**	-.27**	—

*Note. **p < .001. PTG = Political Trust; MCP = Media Credibility; DEI = Deepfake Exposure; DML = Digital Media Literacy; ISS = Information Skepticism.*

Hierarchical Multiple Regression: Political Trust

Table 5 presents the results of the hierarchical multiple regression predicting political trust. In the final model, deepfake exposure emerged as the strongest negative predictor of political trust (β = -.36, p < .001), followed by information skepticism (β = -.30, p < .001). Digital media literacy showed a significant positive relationship with trust (β = .22, p < .001). Gender and education did not reach statistical significance as predictors. The overall model explained 41% of the variance in political trust (R^2 = .41, F(5, 494) = 68.47, p < .001), providing strong support for H1.

Table 5

Hierarchical Multiple Regression Analysis: Political Trust as Criterion Variable (N = 500)

Predictor	B	SE B	β	t	p	95% CI
(Constant)	5.23	0.29	—	18.03	<.001	[4.66, 5.80]
Deepfake Exposure (DEI)	-0.41	0.06	-.36	-6.83	<.001	[-0.53, -0.29]
Digital Media Literacy (DML)	0.28	0.07	.22	4.00	<.001	[0.14, 0.42]
Information Skepticism (ISS)	-0.33	0.07	-.30	-4.71	<.001	[-0.47, -0.19]
Gender (Male = 1)	0.12	0.09	.07	1.33	.184	[-0.06, 0.30]
Education Level	0.09	0.05	.08	1.80	.072	[-0.01, 0.19]
R^2 = .41, Adjusted R^2 = .40, F(5, 494) = 68.47, p < .001						

Note. B = unstandardized regression coefficient; SE B = standard error; β = standardized coefficient; CI = confidence interval. Gender coded 0 = Female, 1 = Male.

Hierarchical Multiple Regression: Media Credibility

Table 6 presents regression results with media credibility as the criterion variable. Deepfake exposure was the strongest negative predictor (β = -.34, p < .001), while political trust was a strong positive predictor (β = .43, p < .001), suggesting a close reciprocal relationship between trust in institutions and trust in the media. Digital media literacy positively predicted credibility perceptions (β = .26, p < .001). The model explained 38% of variance in media credibility (R^2 = .38, F(4, 495) = 75.72, p < .001), providing strong support for H2.

Table 6
Hierarchical Multiple Regression Analysis: Media Credibility as Criterion Variable (N = 500)

Predictor	B	SE B	β	t	P	95% CI
(Constant)	4.98	0.31	—	16.06	<.001	[4.37, 5.59]
Deepfake Exposure (DEI)	-0.37	0.07	-.34	-5.29	<.001	[-0.51, -0.23]
Digital Media Literacy (DML)	0.32	0.07	.26	4.57	<.001	[0.18, 0.46]
Information Skepticism (ISS)	-0.27	0.07	-.26	-3.86	<.001	[-0.41, -0.13]
Political Trust (PTG)	0.44	0.05	.43	8.80	<.001	[0.34, 0.54]
R^2 = .38, Adjusted R^2 = .37, F(4, 495) = 75.72, p < .001						

Note. B = unstandardized regression coefficient; SE B = standard error; β = standardized coefficient; CI = confidence interval.

Moderation Analysis: Digital Media Literacy

The results of the moderation analysis are presented in Table 7 (PROCESS Model 1). The interaction term (Deepfake Exposure × Digital Media Literacy) was noteworthy (B = 0.18, SE = 0.06, t = 3.00, p = .003), indicating that digital media literacy moderated the negative effect of deepfake exposure on political trust. Simple slope analyses revealed that the negative effect of deepfake exposure was significant at low (B = -0.61, p < .001) and moderate (B = -0.43, p < .001) levels of media literacy, but was substantially attenuated at high levels of media literacy (B = -0.25, p = .031). This confirms H4: higher digital media literacy buffers the trust-eroding effects of deepfake exposure.

Table 7
Moderation Analysis: Digital Media Literacy as Moderator of Deepfake Exposure → Political Trust (N = 500)

Term	B	SE	t	p	95% CI [LL, UL]
Deepfake Exposure (X)	-0.43	0.06	-7.17	<.001	[-0.55, -0.31]
Digital Media Literacy (W)	0.29	0.07	4.14	<.001	[0.15, 0.43]
X × W (Interaction)	0.18	0.06	3.00	.003	[0.06, 0.30]
R^2 = .44; ΔR^2 for interaction = .02, p = .003					

Note. The Process Macro Model 1 analysis was carried out (Hayes, 2017) using 5,000 bootstrap samples. Centering of continuous variables in the analysis. X = Deepfake Exposure Index; W = Digital Media Literacy; Y = Political Trust.

Discussion

The aim of the study was to examine the effect of exposure to deepfake media on political trust and media credibility among a stratified sample of Pakistani citizens (N = 500). The results provide strong empirical evidence for the study's main hypotheses and make a valuable contribution to

the growing field of international research on deepfakes and democratic trust.

Deepfake Exposure and Political Trust

The finding that exposure to deepfakes was the most significant negative predictor of political trust (= -.36) aligns with and generalizes previous experimental studies in the West (Vaccari and Chadwick, 2020; Hameleers et al., 2022). The extent of this effect in the Pakistani context is significant and probably indicates the convergence of multiple contextual factors: the already-low level of institutional trust (Waseem, 2022), a politically weaponized information space where deepfakes have been actively used to promote partisan goals (Abbas et al., 2025), and a lack of access to deepfake detection tools and verification resources.

Theoretically, these results are consistent with the Cultivation Theory hypothesis that extended exposure to a biased media environment shapes a biased view of political reality. This sample of Pakistani citizens with high levels of deepfake exposure has been subjected to a media ecology of visual uncertainty that fits within a cultivation framework in which political actors are generally unreliable and political communication is inherently unreliable. The concept of the epistemic trust framework sheds further light on the mechanism: the saturation of deepfakes interferes with the default heuristic of seeing is believing and brings individuals into a state of chronic epistemic vigilance, undermining trust in any political communication, not only in the fake information they have been shown.

Media Credibility and Spillover Effect

The strong negative association between deepfake exposure and media credibility impressions (β = -.34) recreates and generalizes the so-called spillover effect by Hameleers et al. (2022). When participants were presented with deepfake political material, their ratings of the specific source and the media overall decreased. This discovery has grave consequences, especially for the already-challenged legitimate journalism industry in Pakistan. Unless regulators take action to curb the spread of deepfakes, collateral damage to mainstream media credibility, including in outlets involved in producing authentic, verified reporting, might weaken the institutions best positioned to combat disinformation.

The theorized reciprocal relationship between political trust and media credibility is supported by the strong positive correlation (r = .61). This discovery raises the idea that deepfakes are a generalized trust solvent, which both dissolves the faith in political institutions and in the media systems that are supposed to keep the institutions accountable a two-fold corrosive relationship to democratic governance.

The Safeguarding Function of Digital Media Literacy

One of the most policy-relevant implications of the study is the moderation result: digital media literacy mitigated the adverse impact of deepfake exposure on political trust. The negative impact of deepfake exposure on trust was significantly (but not completely) lower among participants who had high digital media literacy scores. This observation aligns with experimental inoculation studies (Roozenbeek et al., 2022) and the theoretical hypothesis that equipping citizens with critical evaluation skills may serve as a cognitive immune system against manipulation of synthetic media.

High media literacy levels do not eliminate the negative association between deepfake viewing and political trust. The media literacy requirement establishes a baseline for trust in deepfake systems, which, in turn, necessitates additional regulatory and technical measures to maintain trustworthiness in deepfake media environments.

Cultural and Contextual Considerations

The high rate of sharing deepfake content (44.6% of participants reported sharing content they later discovered was manipulated) shows how motivated cognition, together with cultural mythology, enables deepfakes to spread. In line with Barthes's (1972) study of myth, deepfakes in Pakistan's political culture seem not to be isolated fabrications but rather the affirmation of already existing political discourses. The reason why synthetic material accuses political opponents of

corruption, betrayal, or incompetence spreads is that it appeals to the already existing ideological frameworks- an insight that has profoundly critical consequences for counter-disinformation strategy. Technical debunking may also not work when deepfakes exploit strongly held political mythologies; successful counter-messaging should seek to address and confront the narrative frames themselves.

Limitations

The current research has several weaknesses that should be acknowledged. The initial aspect of the cross-sectional design does not allow researchers to establish causal relationships, since the theoretical framework and previous experimental findings lead them to certain conclusions, yet the results indicate only statistical correlations. Causation can be determined using longitudinal and experimental research. Second, self-reported exposure to deepfakes is susceptible to social desirability and recall biases; in the future, objective measures should be employed, such as log data and experimental tasks. The sample has a demographic imbalance, as it has a large number of participants who are urban, educated, and internet users, while omitting participants who are rural, less educated, and have different media consumption patterns. The research conducted in 2024 produced results that depended on the political conditions in place during data collection.

Conclusion

This paper has shown that deepfake media exposure is a major and independent risk to media credibility and political trust in Pakistan. The results are theoretically supported by the convergent framework of Third-Person Effect Theory, Cultivation Theory, and Epistemic Trust Framework, and empirically validated by a large, stratified quantitative sample. Exposure to deepfakes was negatively associated with political trust (2 = -.36), and media credibility (2 = -.34), and digital media literacy was a strong but partial buffer.

These repercussions have grave ramifications for practice and policy. At the institutional level, Pakistan's regulatory agencies (PEMRA, PTA) are tasked with developing and implementing comprehensive laws to regulate synthetic media in political communication, such as requiring disclosure of information generated by artificial intelligence and making the use of electoral deepfakes illegal. National media literacy programs must create and distribute instructional materials that prioritize deepfake identification and epistemic hygiene training at secondary and postsecondary educational institutions, according to civil society organizations and education systems. Pakistani IT firms can create sophisticated deepfake identification systems with automated techniques and labeling tools to eliminate fake political content from their platforms.

Future studies should employ experimental designs to determine causal pathways, qualitative research to examine how deepfakes become persuasive through local cultural and political myths in Pakistan, and longitudinal studies to evaluate how exposure to deepfakes affects political trust. Deepfake technology produces visual deception, which is more than just a technical issue; it has become a democratic dilemma that calls for collaboration among many specialists and interest groups to develop solutions.

REFERENCES

Abbas, S., Khalid, N., & Babar, M. K. S. (2025). The Rise of Digital Campaigning: A Comparative Study of Political Parties'social Media Strategies in South Asia. *International Journal of Social Sciences Bulletin*, *3*(7), 728-740.

Aftab, B. (2025). Analysis of issues and challenges of media and its sustainability: A case study of electronic media in Pakistan. *Journal of Political Stability Archive*, *3*(2), 1279-1297.

Barthes, R. (1972). Mythologies (A. Lavers, Trans.). *New York: Hill and Wang*, *117*.

Bartlett, F. C. (1932). Remembering: A study in experimental and social psychology. Cambridge University Press.

Benkler, Y., Faris, R., & Roberts, H. (2018). *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. Oxford University Press.

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

Dalton, R. J. (2004). Democratic challenges, democratic choices: The erosion of political support in advanced industrial democracies. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199268436.001.0001

Davison, W. P. (1983). The third-person effect in communication. Public Opinion Quarterly, 47(1), 1–15. https://doi.org/10.1086/268763

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. Behavior Research Methods, 41(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Flanagin, A. J., & Metzger, M. J. (2000). Perceptions of Internet information credibility. Journalism & Mass Communication Quarterly, 77(3), 515–540. https://doi.org/10.1177/107769900007700304

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

Fonagy, P., & Allison, E. (2014). The role of mentalizing and epistemic trust in the therapeutic relationship. Psychotherapy, 51(3), 372–380. https://doi.org/10.1037/a0036505

Gerbner, G. (1969). Toward 'Cultural Indicators': The analysis of mass mediated public message systems. AV Communication Review, 17(2), 137–148. https://doi.org/10.1007/BF02769085

Gunther, A. C. (1991). What we think others think: Cause and consequence in the third-person effect. Communication Research, 18(3), 355–372. https://doi.org/10.1177/009365091018003002

Hameleers, M., Powell, T. E., de Vreese, C. H., & Lelkes, Y. (2022). A picture paints a thousand lies? The effects and mechanisms of multimodal disinformation and rebuttals disseminated via social media. Political Communication, 39(3), 281–301. https://doi.org/10.1080/10584609.2020.1779872

Hayes, A. F. (2017). Introduction to mediation, moderation, and conditional process analysis: A regression-based approach (2nd ed.). Guilford Press.

Hetherington, M. J. (1998). The political relevance of political trust. American Political Science Review, 92(4), 791–808. https://doi.org/10.2307/2586304

Hobbs, R. (2010). *Digital and media literacy: A plan of action. A white paper on the digital and media literacy recommendations of the Knight Commission on the information needs of communities in a democracy*. Aspen Institute. 1 Dupont Circle NW Suite 700, Washington, DC 20036.

Kalbfleisch, P. J. (2003). Credibility for the 21st century: Integrating perspectives on source, message, and media credibility in the contemporary media environment. *Communication yearbook 27*, 307-350.

Kashyap, S. (2025). The Digital Mirage: India's Evolving Legal Battle Against Deepfake Technology. *SCRIPTed: A Journal of Law, Technology & Society*, *22*(2), 162-226.

Kietzmann, J., Lee, L. W., McCarthy, I. P., & Kietzmann, T. C. (2020). Deepfakes: Trick or treat? Business Horizons, 63(2), 135–146. https://doi.org/10.1016/j.bushor.2019.11.006

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. Science, 359(6380), 1094–1096. https://doi.org/10.1126/science.aao2998

Lewandowsky, S., Ecker, U. K., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the "post-truth" era. *Journal of applied research in memory and cognition*, 6(4), 353-369.

Lewandowsky, S., Smillie, L., Garcia, D., Hertwig, R., Weatherall, J., Egidy, S., ... & Leiser, M. (2020). Technology and democracy: Understanding the influence of online technologies on political behaviour and decision-making.

Morgan, M., Shanahan, J., & Signorielli, N. (2018). Yesterday's new cultivation, tomorrow. Mass Communication and Society, 18(5), 674–699. https://doi.org/10.1080/15205436.2015.1072725

Moroojo, M. Y., Farooq, U., Madni, M. A., Shabbir, T., & Khalil, H. (2025). Algorithmic amplification and political discourse: the role of AI in shaping public opinion on social media in Pakistan. *The Critical Review of Social Sciences Studies*, *3*(2), 2552-2570.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology*, *2*(2), 175-220.

Pakistan Telecommunication Authority. (2023). Annual report 2022–2023. PTA. https://www.pta.gov.pk/en/annual-reports

Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, *188*, 39-50. https://doi.org/10.1016/j.cognition.2018.06.011

Potter, W. J. (2004). *Theory of media literacy: A cognitive approach*. Sage Publications.

Roozenbeek, J., van der Linden, S., Goldberg, B., Rathje, S., & Lewandowsky, S. (2022). Psychological inoculation improves resilience against misinformation on social media. Science Advances, 8(34), eabo6254. https://doi.org/10.1126/sciadv.abo6254

Sheikh, M. A., Ashraf, Z., Mir, B., & Akhtar, S. (2024). Mediatization's Impact on News Media Trust and Credibility: A Comprehensive Analysis of Viewer Perceptions. International Journal of Social Science Archives (IJSSA), 7(2).

Sperber, D., Clement, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. Mind & Language, 25(4), 359–393. https://doi.org/10.1111/j.1468-0017.2010.01394.x

Thies, J., Zollhofer, M., Stamminger, M., Theobalt, C., & Nießner, M. (2016). Face2face: Real-time face capture and reenactment of rgb videos. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2387-2395). https://doi.org/10.1109/CVPR.2016.262

Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). Deepfakes and beyond: A survey of face manipulation and fake detection. *Information fusion*, *64*, 131-148. https://doi.org/10.1016/j.inffus.2020.06.014

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

Wahid, S. (2024). Status of Media Literacy Education in Pakistan. *Journal of Social Sciences and Media Studies*, *8*(2), 21-33.

Waseem, M. (2022). *Political conflict in Pakistan*. Oxford University Press.

Westerlund, M. (2019). The emergence of deepfake technology: A review. Technology *Innovation Management Review*, *9*(11), 39–52.