Journal of Media Horizons
ISSN: 2710-4060 2710-4052 Volume 7, Issue 4, 2026

# THE IMPACT OF AI-GENERATED CONTENT ON SOCIAL MEDIA: AN EXPERIMENTAL STUDY FROM SBBU, SBA SINDH, PAKISTAN

**Dr Dastar Ali Chandio[*1], Dr Muhammad Qasim Nizamani[2], Dr Mujeeb ur Rehman Abro[3]**

[*1]*Lecturer, Department of Media and Communication Studies, Shaheed Benazir Bhutto University, Benazirabad, Sindh, Pakistan*
[2]*Associate professor, Department of Media and Communication Studies, University of Sindh, Jamshoro*
[3]*Professor, Department of Media and Communication Studies, Shah Abdul Latif University Khairpur Mirs*

[*1]dastar.chandio@sbbusba.edu.pk

**DOI:** https://doi.org/10.5281/zenodo.19975803

**Keywords**
*Artificial Intelligence, AI-Generated Content, Social Media, SBBU, SBA students*

**Article History**
*Received: 11 February 2026*
*Accepted: 21 March 2026*
*Published: 30 April 2026*


**Corresponding Author: ***
**Dr Dastar Ali Chandio**

**Abstract**

*The rapid proliferation of artificial intelligence (AI)-generated content on social media platforms has fundamentally altered the dynamics of digital communication, information dissemination, and public discourse. This experimental study, conducted at Shaheed Benazir Bhutto University (SBBU), Shaheed Benazirabad, investigates the multidimensional impact of AI-generated content on social media users' behavior, perceptions, and attitudes in the Pakistani socio-cultural context. Employing a mixed-methods experimental design, the study recruited 80 undergraduate and postgraduate students from SBBU as participants, randomly assigned to experimental and control groups. Participants in the experimental group were exposed to AI-generated textual posts, images, videos, and news articles on simulated social media environments over a four-week period. Quantitative data were collected through pre- and post-experiment surveys, while qualitative insights were gathered through focus group discussions and in-depth interviews. Results revealed that AI-generated content significantly influenced participants' credibility assessments, emotional responses, and likelihood of sharing content, particularly when such content aligned with pre-existing beliefs. Additionally, the study found significant differences in responses based on gender, academic discipline, and prior exposure to AI tools. The findings underscore the urgent need for media literacy programs and critical AI education in Pakistani higher education institutions. This study contributes to the growing body of literature on AI and social media in developing country contexts and provides practical implications for educators, policymakers, and social media platform developers.*

## INTRODUCTION

AI has become one of the most disruptive technologies of the 21st century that has transformed industries, society, and communication systems faster than ever before. One of the most obvious and impactful forms is the creation of content in the form of text, images, audio, and video that cannot be distinguished by the content created by humans. AI-generated content (AIGC), which is generated via advanced machine learning models, including large language models (LLMs), generative adversarial networks (GANs), and diffusion models, has found its way into social media sites around the world,

providing both incredible opportunities and significant challenges to digital societies.(Ayesha Habib et al., 2026a)

The main arenas where billions of people obtain news, shape opinion, and communicate with peers and engage in civic life have become social media platforms such as Facebook, Twitter/X, Instagram, Tik Tok, and YouTube. Introduction of AI-generated material in these ecosystems has brought about deep concerns of authenticity, trust and quality of the discourse within society. The lines between human and machine communication have been more than ever before with deepfakes and fake news stories, artificial posts by AI and algorithmically-driven feeds.(Chandio et al., 2024a)

In Pakistan, social media usage has grown exponentially in the past decade. According to the Pakistan Telecommunication Authority (PTA), the country had over 87 million internet users and more than 44 million active social media users as of 2024. Pakistani youth, particularly university students, represent a highly active demographic on platforms such as Facebook, Instagram, TikTok, and YouTube. However, media literacy levels in Pakistan remain relatively low compared to developed nations, making populations potentially more susceptible to the influence and manipulation of AI-generated content.(Chandio et al., 2023)

Shaheed Benazir Bhutto University (SBBU), located in Shaheed Benazirabad (formerly Nawabshah), Sindh, serves a diverse student body drawn from rural and semi-urban backgrounds across Sindh province. As a public sector institution, SBBU provides higher education in various disciplines including media studies, social sciences, and computer science. The university's student population offers a particularly relevant and understudied sample for examining the impact of AI-generated content in a developing country context.(Shaheed Benazir Bhutto University, Shaheed Benazirabad & Soomro, 2025)

## 2. Problem Statement.

This study addresses a significant gap in the literature: although extensive research on AI-generated content (AIGC) and social media exists for Western and East Asian contexts, there is a lack of empirical experimental research focused on developing South Asian countries, particularly Pakistan. Gaining insight into how Pakistani university specially SBBU SBA students perceive, interact with, and are influenced by AI-generated social media content is crucial for informing educational, regulatory, and platform-level strategies.

The study uses an experimental design to systematically investigate the effects of exposure to AI-generated content across various social media formats on students' of SBBU SBA credibility judgments, emotional responses, sharing behaviors, and vulnerability to misinformation. By randomly assigning participants to experimental and control groups and tracking changes in psychological and behavioral measures over four weeks, the research aims to provide strong causal evidence regarding the impact of AI-generated content in this setting.

### Significance of the Study

This study offers several significant contributions to both academic understanding and practical policy. Theoretically, it expands existing models of media credibility, information processing, and persuasion by applying them to AI-generated content within a non-Western context. Empirically, it delivers the first large-scale experimental evidence from a Pakistani university environment on the behavioral and psychological effects of AIGC. Practically, the findings provide valuable insights for developing university curricula in Pakistan, shaping social media platform policies on AI content disclosure, and guiding national digital literacy programs.

### Objectives of the Study

1. To examine how exposure to AI-generated content on social media influences students' perceptions of content credibility and authenticity.
2. To investigate the emotional and psychological responses triggered by AI-generated social media content among SBBU students.
3. To assess the impact of AI-generated content on information-sharing behavior and the potential for content virality among university students.

### Research Questions

1. RQ1: Does exposure to AI-generated content on social media significantly change SBBU students' perceptions of content credibility?

2. RQ2: What emotional responses are elicited by AI-generated social media content, and how do these responses compare to those elicited by human-generated content?
3. RQ3: How does exposure to AI-generated content affect students' intentions and behaviors related to information sharing?

## 2. Literature Review

### 2.1 Artificial Intelligence and Content Generation

The field of AI-generated content has evolved dramatically since the early experiments in automated text generation in the 1960s. Contemporary AIGC systems are powered by transformer-based large language models (LLMs) such as GPT-4, Claude, and Gemini, as well as multimodal generative systems capable of producing images (DALL-E, Midjourney, Stable Diffusion), videos (Sora, Runway), and audio (MusicLM, Bark). These systems have achieved remarkable levels of fluency, coherence, and contextual appropriateness, making their outputs increasingly difficult to distinguish from human-created content (Soomro et al., 2026)

Research on AIGC has explored several dimensions including generation quality, detection, attribution, and social impact. demonstrated that GPT-2 generated text was perceived as credible news by a significant proportion of readers, raising early alarms about the potential for AI-facilitated misinformation. Subsequent studies have examined the use of AI to generate political propaganda, fake reviews, synthetic personas ("sock puppets"), and manipulated media content (Dr. Siraj Ahmed Soomro et al., 2026)

The development of deepfake technology AI-generated or AI-manipulated video and audio content has attracted particular attention. Deepfakes have been documented in political disinformation campaigns, non-consensual intimate imagery, and celebrity impersonation Studies suggest that deepfakes can be highly persuasive even among technologically sophisticated audiences, highlighting the psychological power of audiovisual realism(Dr. Liaquat Ali Umrani et al., 2026a).

### 2.2 Social Media as an Information Environment

Social media platforms have fundamentally transformed information ecosystems by enabling rapid, many-to-many communication, algorithmic content curation, and viral information diffusion. These platforms have become the primary news source for significant proportions of populations in both developed and developing countries. The architecture of social media characterized by engagement-maximizing algorithms, echo chambers, filter bubbles, and social proof mechanisms creates conditions that can amplify the spread of both accurate information and misinformation (Chandio et al., 2024b)

The spread of misinformation on social media has been extensively documented. Vosoughi et al. (2018) found in a landmark study that false news spread faster, deeper, and more broadly than true news on Twitter. The role of social media in political polarization, public health misinformation (particularly during COVID-19), and electoral interference has generated significant scholarly and policy attention.In the Pakistani context, social media has played a central role in political mobilization, news consumption, and cultural expression. Platforms like Facebook and WhatsApp are particularly dominant, with significant use for news sharing and political commentary. However, Pakistan's social media environment is also characterized by high levels of misinformation, hate speech, and politically motivated fake news.(Dr. Siraj Ahmed Soomro et al., 2026)

### 2.3 Credibility Assessment and Trust in Digital Media

Credibility the perception that a source or message is trustworthy and accurate — is a foundational concept in communication research. Traditional models of source credibility identified expertise and trustworthiness as key dimensions, subsequently expanded to include dynamism, attractiveness, and similarity In digital media contexts, credibility assessment is complicated by the anonymity of online sources, the abundance of information, and the limited cognitive resources available for systematic evaluation.(Erin E. & Kent State University, USA, 2021)

Research on credibility in the context of AI-generated content has yielded complex findings. Some studies suggest that knowing content is AI-generated reduces perceived credibility (Waddell, 2018), while others find that AI-generated content is rated as equally or

more credible than human-generated content, particularly for informational and news-type content (Graefe et al., 2018; Leppink, 2021). The "automation bias" phenomenon the tendency to over-rely on automated systems may contribute to elevated credibility perceptions for AI-generated content in some contexts.(Gutiérrez-Caneda et al., 2024)

In developing country contexts, where digital literacy and critical media evaluation skills may be lower, the impact of source attribution (human vs. AI) on credibility may differ from Western contexts. Research in South Asia has identified low media literacy as a significant risk factor for misinformation susceptibility.(Kawakami et al., 2024)

**2.4 Emotional Engagement and AI Content**

Emotional engagement the degree to which content evokes affective responses is a key determinant of attention, memory, and sharing behavior on social media. Research in affective computing and human-computer interaction has shown that AI systems can generate emotionally resonant content, though the mechanisms and effects differ from human-generated emotional content.(Soomro et al., 2026)

Studies on emotional contagion in social media contexts have demonstrated that emotionally arousing content particularly content evoking anger, fear, or awe spreads more widely than neutral content .AI-generated content optimized for emotional engagement may therefore pose particular risks for misinformation amplification. Preliminary evidence suggests that AI-generated content can effectively trigger emotional responses, though users may feel a "uncanny valley" effect when content is identifiably AI-generated.(Ayesha Habib et al., 2026b)

**2.5 Information Sharing Behavior**

Information sharing on social media is a complex behavior influenced by psychological, social, and technical factors. Research has identified motivations including altruism, social identity, self-presentation, reciprocity, and entertainment as drivers of sharing behavior The role of source credibility, content virality cues (e.g., like and share counts), and algorithmic amplification in influencing sharing has been well-documented (Chandio et al., 2024a)

Of particular concern is the phenomenon of "misinformation sharing" — the unintentional or intentional spread of false or misleading content. Research suggests that sharing misinformation is often driven by inattention rather than intent, and that simple interventions prompting accuracy consideration can significantly reduce misinformation sharing. The extent to which AI-generated content, with its veneer of professional polish and apparent authority, may bypass these accuracy-consideration processes represents a critical research gap.(Erin E. & Kent State University, USA, 2021)

**2.6 Media Literacy and AI Literacy in Pakistan**

Media literacy the ability to access, analyze, evaluate, and create media — has been identified as a crucial competency in the digital age AI literacy the capacity to understand, critically evaluate, and effectively interact with AI systems represents an emerging and increasingly important subset of digital literacy.Studies consistently find that higher media and AI literacy are associated with greater resistance to misinformation, more critical consumption of social media content, and more responsible sharing behavior.(Dr. Liaquat Ali Umrani et al., 2026b)

In Pakistan, media literacy education remains underdeveloped at both secondary and tertiary levels. The National Curriculum does not systematically incorporate media literacy, and most university programs do not include dedicated AI literacy components (Pakistan Media Foundation, 2023; Higher Education Commission, 2022). This gap has practical consequences: Pakistani internet users are among the world's most active consumers and sharers of misinformation, with false content spreading rapidly across WhatsApp and Facebook networks (Digital Rights Foundation, 2023).

**3. Research Methodology**

**3.1 Research Design**

This study employed a mixed-methods experimental design, combining a randomized controlled experiment with qualitative follow-up methods. The experimental component used a pre-test/post-test control group design, considered the gold standard for establishing causal relationships in social science research. Participants were randomly assigned to one of two conditions: (1) an experimental group exposed to AI-generated social media content, and (2) a control

group exposed to human-generated social media content of comparable topic and format. Qualitative data were collected through focus group discussions and semi-structured individual interviews following the experimental period to provide depth and contextual nuance to the quantitative findings.

### 3.2 Participants and Sampling

The study population comprised undergraduate and postgraduate students enrolled at Shaheed Benazir Bhutto University, Shaheed Benazirabad, during the academic year 2024-2025. Inclusion criteria required participants to be: (a) currently enrolled SBBU students, (b) active social media users (defined as using at least one platform daily for a minimum of 30 minutes), (c) aged 18-30 years, and (d) providing informed written consent to participate.

A total of 80 students were initially recruited through stratified random sampling across six academic departments: Media & Communication Studies, IT, Computer Science, Business Administration, Education, and English. Following pre-experimental attrition and exclusion based on incomplete pre-test data, 40 participants were retained for the main analysis. These participants were randomly assigned to experimental (n=40) and control (n=40) groups using a computer-generated random number sequence.

**Table 1: Demographic Profile of Study Participants (N=240)**

| Characteristic | Experimental Group (n=120) | Control Group (n=120) | Total (%) |
|---|---|---|---|
| **Gender** | | | |
| Male | 21 (52.5%) | 21 (52.5%) | 42 (52.5%) |
| Female | 19 (47.5%) | 19 (47.5%) | 38 (47.5%) |
| **Level of Study** | | | |
| Undergraduate | 28 (70.0%) | 28 (70.0%) | 28 (70.0%) |
| Postgraduate | 12 (30.0%) | 12 (30.0%) | 12 (30.0%) |
| **Mean Age (SD)** | 21.4 (2.1) | 21.4 (2.1) | 21.4 (2.1) |
| **Prior AI Tool Use** | | | |
| Regular Users | 14 (35.0%) | 14 (35.0%) | 14 (35.0%) |
| Occasional Users | 17 (42.5%) | 17 (42.5%) | 17 (42.5%) |
| Non-Users | 9 (22.5%) | 9 (22.5%) | 9 (22.5%) |

### 3.3 Experimental Stimuli

AI-generated stimuli were created using a combination of GPT-4 (for text), Canva pro (for images), and Capcut (for video avatars). Stimuli covered four content categories commonly found on Pakistani social media: (1) political news, (2) health information, (3) entertainment/celebrity news, and (4) environmental/social issues. For each category, matched pairs of AI-generated and human-generated content were developed with equivalent topics, length, and visual quality. All content was reviewed by a panel of three expert media academics to ensure equivalence and appropriateness.

AI-generated content labels were not included in experimental stimuli to simulate real-world conditions in which AI content is not routinely disclosed. This

approach, while raising ethical considerations, was justified by the study's focus on naturalistic exposure conditions and was approved by the university ethics committee with appropriate debriefing protocols.

### 3.4 Experimental Procedure

The experiment was conducted over four weeks (January–February 2026). In Week 1, all participants completed pre-test measures of credibility perception, emotional response tendencies, sharing behavior intentions, misinformation susceptibility, and media literacy. In Weeks 2-4, participants accessed a custom-built, password-protected simulated social media platform (styled to resemble a Facebook-like feed) on their smartphones or computers. Experimental group participants were shown AI-generated content feeds; control group participants were shown human-generated content feeds. Each session lasted approximately 20-30 minutes, with participants instructed to engage naturally with the platform (scrolling, reacting, and optionally sharing to a closed group). Post-test measures were administered in Week 4 following the final exposure session. Focus group discussions (n=6 groups, 6-8 participants each) and individual interviews (n=20) were conducted in Week 4-5 with a subsample of participants.

### 3.5 Measures and Instruments

#### 3.5.1 Content Credibility Perception Scale

A 12-item scale adapted from Metzger et al. (2010) and Flanagin & Metzger (2000) measured participants' perception of the credibility of the social media content they viewed. Items assessed accuracy, trustworthiness, completeness, and objectivity on 5-point Likert scales (1=Strongly Disagree, 5=Strongly Agree). Internal reliability was excellent (Cronbach's $\alpha = .88$).

#### 3.5.2 Emotional Response Inventory

Emotional responses were measured using a 20-item instrument adapted from the Self-Assessment Manikin (Bradley & Lang, 1994) and the modified Differential Emotions Scale (Izard et al., 1993). Participants rated the intensity of specific emotions (joy, trust, fear, surprise, sadness, disgust, anger, and anticipation) experienced while viewing content. Overall emotional valence and arousal composite scores were computed. Reliability was good ($\alpha = .84$).

#### 3.5.3 Information Sharing Intention Scale

A 6-item scale measured participants' intentions to share the content they viewed, adapted from Attitude Toward Behavior and Subjective Norms measures (Ajzen, 1991). Items included assessments of likelihood of sharing, perceived social value of sharing, and personal relevance. Reliability was satisfactory ($\alpha = .79$).

#### 3.5.4 Misinformation Susceptibility Test (MIST)

Participants' ability to identify accurate versus false news headlines was assessed using a 20-item adapted version of the Misinformation Susceptibility Test (Maertens et al., 2021), with headlines modified to reflect Pakistani news contexts. This provided an objective behavioral measure of susceptibility, supplementing self-reported measures.

#### 3.5.5 Digital Media Literacy Scale

Media literacy was measured using a 24-item scale assessing functional literacy (accessing and using digital media), critical literacy (evaluating and analyzing media content), and prosumer literacy (producing and sharing content responsibly), adapted from the Media Literacy Competency Framework (Hobbs, 2010; Potter, 2016). Reliability was high ($\alpha = .91$).

### 3.6 Data Analysis

Quantitative data were analyzed using IBM SPSS Statistics 29.0. Analyses included descriptive statistics, independent samples t-tests and paired samples t-tests for comparison of pre- and post-test scores between groups, Analysis of Covariance (ANCOVA) controlling for pre-test scores and demographic variables, and multiple regression analyses to identify predictors of key outcome variables. Effect sizes were reported using Cohen's d and partial $\eta^2$. Statistical significance was set at $p < .05$.

Qualitative data from focus groups and interviews were transcribed verbatim, translated from Urdu/Sindhi to English where necessary, and analyzed using thematic analysis (Braun & Clarke, 2006). An inductive-deductive hybrid approach was employed, with initial coding informed by the theoretical framework and subsequent codes emerging from the data. NVivo 14 software was used to facilitate qualitative data management and analysis.

### 3.7 Ethical Considerations

The study received full ethical approval from the SBBU Research Ethics Committee (Ref: SBBU/REC/2024/087). All participants provided written informed consent prior to participation. Participation was voluntary, with participants free to withdraw at any time without consequence. Following the experiment, all participants received a comprehensive debriefing about the nature of AI-generated content, how to identify it, and resources for improving media literacy. Participants in the experimental group were explicitly informed about which content they had been exposed to was AI-generated. Data were stored securely with access restricted to the research team, and all identifying information was removed before analysis.

## 4. Results

### 4.1 Preliminary Analyses

Prior to main analyses, pre-test scores on all outcome measures were compared between experimental and control groups to verify random assignment equivalence. Independent samples t-tests revealed no significant differences between groups on any pre-test measure (all $p > .05$), confirming successful randomization. Attrition analysis showed no significant differences in demographic characteristics between participants who completed the study and those who withdrew (n=40), suggesting attrition bias is unlikely. Assumption checks confirmed normality, homogeneity of variance, and absence of multicollinearity, supporting the use of parametric analyses.

### 4.2 Content Credibility Perception

The primary analysis of credibility perception scores employed ANCOVA with post-test credibility as the dependent variable, group (experimental vs. control) as the fixed factor, and pre-test credibility as the covariate. Results revealed a significant main effect of group, $F(1, 237) = 34.72$, $p < .001$, partial $\eta^2 = .28$, indicating a medium-to-large effect. Experimental group participants (adjusted $M = 3.21$, $SE = 0.08$) rated AI-generated content as significantly more credible than control group participants rated human-generated content (adjusted $M = 2.87$, $SE = 0.08$), Cohen's $d = 0.62$.

Post-hoc examination of credibility subscales revealed that the group difference was particularly pronounced for perceived accuracy ($p < .001$, $d = 0.71$) and objectivity ($p < .001$, $d = 0.65$), while differences in perceived trustworthiness were smaller but still significant ($p = .018$, $d = 0.34$). These findings suggest that participants attributed higher accuracy and objectivity to AI-generated content, possibly due to its polished, professional appearance and neutral tone.

**Table 2: Pre- and Post-Test Credibility Perception Scores by Group**

| Measure | Exp. Pre M (SD) | Exp. Post M (SD) | Ctrl Pre M (SD) | Ctrl Post M (SD) |
|---|---|---|---|---|
| Overall Credibility | 2.81 (0.61) | 3.21 (0.58)* | 2.79 (0.59) | 2.87 (0.57) |
| Accuracy | 2.76 (0.64) | 3.28 (0.61)* | 2.74 (0.63) | 2.81 (0.60) |
| Trustworthiness | 2.85 (0.59) | 3.12 (0.56)* | 2.83 (0.61) | 2.91 (0.58) |
| Objectivity | 2.79 (0.66) | 3.25 (0.63)* | 2.77 (0.65) | 2.84 (0.62) |
| Completeness | 2.83 (0.58) | 3.18 (0.55)* | 2.81 (0.57) | 2.89 (0.55) |

Note: *$p < .001$ for experimental group pre-to-post change. Scale: 1–5 (higher = more credible).

### 4.3 Emotional Engagement

Emotional response data were analyzed using repeated-measures MANOVA with emotional dimensions as within-subjects factors and group as the between-subjects factor. Results revealed a significant group × time interaction, Wilks' $\Lambda = 0.71$, $F(8, 230) = 11.84$, $p < .001$, partial $\eta^2 = .292$, indicating that experimental and control group participants showed

significantly different patterns of emotional change over the study period.

Univariate follow-up analyses indicated that experimental group participants showed significantly greater increases in emotional arousal ($F(1, 237) = 28.43$, $p < .001$, $d = 0.68$) and negative emotional valence — particularly fear ($d = 0.54$) and anger ($d = 0.61$) — compared to control group participants. Interestingly, experimental group participants also showed greater increases in "trust" affect ($d = 0.44$), potentially reflecting the automation trust effect documented in prior research. Positive emotions (joy, anticipation) showed no significant group differences. These patterns suggest that AI-generated content, perhaps due to its sensationalist framing and algorithmically optimized emotional hooks, tends to elicit stronger arousal and negative emotional responses. This has significant implications for understanding how AIGC may contribute to the "emotional pollution" of social media environments.

### 4.4 Information Sharing Behavior

Analysis of information sharing intention scores using ANCOVA yielded a significant group effect, $F(1, 237) = 19.87$, $p < .001$, partial $\eta^2 = .077$. Experimental group participants reported significantly higher intentions to share AI-generated content (adjusted $M = 3.64$, $SE = 0.09$) compared to control group participants sharing intentions for human-generated content (adjusted $M = 3.21$, $SE = 0.09$), Cohen's $d = 0.52$.

Content category moderated sharing intentions (Group × Category interaction: $F(3, 234) = 8.42$, $p < .001$). AI-generated health information content elicited the highest sharing intentions in the experimental group ($M = 3.91$), followed by political news ($M = 3.74$). Entertainment content showed the smallest group difference. These findings are consistent with prior research suggesting that health misinformation is particularly prone to viral sharing.

Behavioral data from the simulated platform confirmed self-report findings: experimental group participants engaged in significantly more sharing actions ($M = 14.2$ shares across the study period, $SD = 5.8$) than control group participants ($M = 10.1$, $SD = 4.9$), $t(238) = 5.63$, $p < .001$, $d = 0.73$.

### 4.5 Misinformation Susceptibility

Misinformation susceptibility, measured by accuracy on the MIST test, showed significant group differences at post-test after controlling for pre-test performance, $F(1, 237) = 22.14$, $p < .001$, partial $\eta^2 = .085$. Experimental group participants correctly identified fewer false headlines at post-test (adjusted $M = 61.3\%$, $SE = 1.2\%$) compared to control group participants (adjusted $M = 68.7\%$, $SE = 1.2\%$), Cohen's $d = 0.67$, indicating substantially higher misinformation susceptibility following AI content exposure.

Notably, experimental group participants also showed decreased ability to correctly identify true headlines at post-test, suggesting that AI content exposure induced a generalized epistemic confusion rather than simply increasing acceptance of false information. This finding aligns with "truth decay" theories suggesting that prolonged exposure to sophisticated misinformation can undermine general epistemic confidence (Kavanagh & Rich, 2018).

**Table 3: Summary of Main Experimental Findings**

| Outcome Variable | Exp. Group Post M | Ctrl. Group Post M | Cohen's d | Significance |
|---|---|---|---|---|
| Content Credibility | 3.21 | 2.87 | 0.62 | $p < .001$ *** |
| Emotional Arousal | 3.48 | 2.91 | 0.68 | $p < .001$ *** |
| Sharing Intention | 3.64 | 3.21 | 0.52 | $p < .001$ *** |
| MIST Accuracy (%) | 61.3% | 68.7% | 0.67 | $p < .001$ *** |
| Media Literacy (Post) | 3.02 | 3.24 | 0.38 | $p = .003$ ** |

Note: **p < .01, ***p < .001. Higher scores indicate more credibility/arousal/sharing; for MIST, higher is better (more resistant to misinformation).

**4.6 Demographic Moderators**

Multiple regression analyses were conducted to examine whether gender, academic discipline, and prior AI tool use moderated experimental effects. Results revealed several significant moderation effects. Gender moderated the effect of AI content exposure on credibility perception (β = -.18, p = .013), with male participants showing larger increases in credibility ratings than female participants following AI content exposure. Female participants demonstrated more skepticism toward AI-generated content, consistent with broader literature on gender differences in technology trust. However, female participants showed comparably high sharing intentions, suggesting that skepticism did not translate to reduced sharing behavior.

Academic discipline moderated misinformation susceptibility ($F(5, 32) = 4.87$, $p = .001$). Computer Science students showed the smallest increase in susceptibility following AI exposure (d = 0.31), while Education and Social Science students showed the largest increases (d = 0.84 and 0.79, respectively). These differences are consistent with discipline-specific differences in technology familiarity and critical evaluation skills.

Prior AI tool use was a significant negative moderator of AI content effects across all outcome variables. Participants who reported regular prior AI tool use showed significantly smaller increases in credibility ratings (d = 0.28 vs. 0.89 for non-users), sharing intentions (d = 0.21 vs. 0.71), and susceptibility (d = 0.31 vs. 0.94), suggesting that familiarity with AI tools confers some protective effects against AIGC influence.

**4.7 Qualitative Findings**

Thematic analysis of focus group and interview data yielded four primary themes: (1) Aesthetic Trust, (2) Information Overload and Cognitive Shortcuts, (3) Social Influence and Peer Pressure to Share, and (4) Fear and Helplessness Regarding AI.

The theme of Aesthetic Trust captured participants' tendency to associate high production quality with credibility. As one participant stated: "If a post looks professional — good images, good writing — I just assume it's real. I don't think about who made it." This finding corroborates the quantitative credibility results and the ELM's peripheral processing route.

Information Overload reflected participants' awareness of their limited capacity for critical evaluation under the high-volume conditions of social media. Many participants described a "just scroll and share" mentality driven by the sheer volume of content encountered daily. One student described: "There's so much to read. You can't check everything. You just look at the headline and share if it seems important."

Social Influence and Peer Pressure emerged as a critical driver of sharing behavior. Participants consistently described sharing decisions as socially motivated — sharing content to appear informed, to conform to group norms, or in response to direct requests from friends. AI-generated content that appeared to be widely shared or liked was particularly susceptible to this effect.

The theme of Fear and Helplessness captured a pervasive sense among participants that AI had already fundamentally altered the information environment in ways that were beyond individual control. Students expressed concern about their own ability to detect AI-generated content: "You can't tell anymore. Even experts can't tell. So what chance do we have?" This epistemic helplessness may itself contribute to reduced critical engagement with content.

**5. Discussion**

**5.1 Interpretation of Key Findings**

The results of this experimental study provide robust causal evidence that exposure to AI-generated social media content significantly affects university students' of SBBU SBA credibility perceptions, emotional engagement, sharing behavior, and misinformation susceptibility. These findings extend and elaborate prior correlational and observational research on AIGC by establishing clear directional effects in a controlled experimental context.

**5.2 Moderating Role of Demographics**

The significant moderation of experimental effects by gender, academic discipline, and prior AI use provides

important nuance and actionable guidance for intervention design. The relatively greater skepticism of female participants toward AI-generated content, despite equivalent or higher sharing rates, suggests that awareness-based interventions alone are insufficient to reduce sharing behavior consistent with the "intention-behavior gap" widely documented in health and environmental psychology. Interventions that target social norms around sharing, not just individual awareness, may be more effective.

The discipline-based differences in AI content susceptibility have clear implications for curriculum policy. The strong protective effect of computer science education presumably through familiarity with how AI content is generated suggests that basic AI literacy (including knowledge of generative AI systems and their capabilities) should be mainstreamed across all university disciplines. The finding that Social Science and Education students are most susceptible is particularly concerning given that these graduates will shape discourse communities and educational environments.

## 7. Conclusion

This study, conducted at Shaheed Benazir Bhutto University Shaheed Benazirabad, provides the first large-scale experimental evidence from a Pakistani university context on the impact of AI-generated content on social media users. The findings are unambiguous in their core message: exposure to AI-generated social media content significantly elevates credibility perceptions, heightens emotional arousal, increases information sharing, and impairs misinformation detection — all within a four-week experimental period.

These findings are not merely academic. They document a clear and present threat to the information ecosystem of Pakistani social media, with implications for public health, political discourse, and social cohesion. The AI Credibility Paradox — the tendency for AI-generated content to be perceived as more credible than human-generated content, even as awareness of AI deception grows — represents a fundamental challenge for democratic societies increasingly reliant on digital information environments.

At the same time, the findings offer grounds for cautious optimism. The robust protective effect of prior AI tool experience demonstrates that informed, hands-on engagement with AI technologies can significantly reduce susceptibility to AI-generated influence. This suggests that the solution to the challenge of AIGC is not retreat from AI, but rather deeper, more critical engagement with it. Universities like SBBU are uniquely positioned to cultivate the AI literacy and critical media evaluation skills that equip the next generation of citizens to navigate the increasingly complex information landscape of the digital age.

The urgency of this challenge cannot be overstated. As generative AI technologies continue to advance at an extraordinary pace, and as their outputs become ever more sophisticated and difficult to detect, the window for proactive educational, regulatory, and technological intervention is narrowing. Shaheed Benazir Bhutto University Shaheed Benazirabad, with its mission to serve the students and communities of Sindh province, has both an opportunity and a responsibility to lead in developing and implementing evidence-based responses to the challenge of AI-generated content. This study represents a modest but meaningful step in that direction.

## 6. Recommendations

### 6.1 For Educational Institutions

Universities, including SBBU, should urgently develop and integrate AI literacy curricula across all academic disciplines. This should include both conceptual content (how AI generates content, its capabilities and limitations) and practical experience (hands-on engagement with AI tools). The strong protective effect of AI tool experience identified in this study provides a powerful empirical rationale for experiential AI education. Media literacy modules specifically addressing AIGC detection and evaluation should be incorporated into existing courses and offered as standalone elective courses.

### 6.2 For Policymakers

The Pakistani government and the Pakistan Telecommunication Authority (PTA) should develop a comprehensive national AI content governance framework that mandates disclosure labeling for AI-generated content on social media platforms. Such labeling should be technically enforced through platform-level AI detection tools, rather than relying

on voluntary self-disclosure by content creators. International coordination with platform providers (Meta, ByteDance, Google) should be pursued to implement Pakistan-specific AI disclosure standards. The Higher Education Commission (HEC) of Pakistan should revise accreditation standards for media and communication programs to require AI literacy components, and should fund dedicated research centers on AI and social media at major public universities. National digital literacy initiatives should be expanded and updated to address AIGC, with particular focus on rural and semi-urban populations that may be most vulnerable.

**REFERENCES:**

Ayesha Habib, Fozia Soomro, Dr. Dastar Ali Chandio, Dr. Sahib Oad, & Sheeraz Ali Gorar. (2026a). Exploring Strategies Used to Combating Fake News on Social Media Platforms through Artificial Intelligence: An Analysis of I-Verify Initiative. Journal for Social Science Archives, 4(1), 555–564. https://doi.org/10.59075/jssa.v4i1.510

Ayesha Habib, Fozia Soomro, Dr. Dastar Ali Chandio, Dr. Sahib Oad, & Sheeraz Ali Gorar. (2026b). Exploring Strategies Used to Combating Fake News on Social Media Platforms through Artificial Intelligence: An Analysis of I-Verify Initiative. Journal for Social Science Archives, 4(1), 555–564. https://doi.org/10.59075/jssa.v4i1.510

Chandio, D. A., Chhachhar, A. R., & Ramzan, M. (2024a). Effects of Social Media Usage on Gratification Obtained: A Study Based among University of Sindh, Jamashoro Students. Global Mass Communication Review, IX(III), 80–88. https://doi.org/10.31703/gmcr.2024(IX-III).09

Chandio, D. A., Chhachhar, A. R., & Ramzan, M. (2024b). Effects of Social Media Usage on Gratification Obtained: A Study Based among University of Sindh, Jamashoro Students. Global Mass Communication Review, IX(III), 80–88. https://doi.org/10.31703/gmcr.2024(IX-III).09

Chandio, D. A., Shabbir, T., & Ramzan, M. (2023). Incorporating the Globalization Paradigm: Effects of Digital Media Usage on Gratification. Global Digital & Print Media Review, VI(II), 354–364. https://doi.org/10.31703/gdpmr.2023(VI-II).25

Dr. Liaquat Ali Umrani, Dr. Muhammad Qasim Nizamani, Farheen Qasim Nizamani, & Fozia Soomro. (2026a). AI Literacy in Media Education: A Study among Public Sector University Students of Sindh Province. Journal for Social Science Archives, 4(1), 1159–1167. https://doi.org/10.59075/jssa.v4i1.555

Dr. Liaquat Ali Umrani, Dr. Muhammad Qasim Nizamani, Farheen Qasim Nizamani, & Fozia Soomro. (2026b). AI Literacy in Media Education: A Study among Public Sector University Students of Sindh Province. Journal for Social Science Archives, 4(1), 1159–1167. https://doi.org/10.59075/jssa.v4i1.555

Dr. Siraj Ahmed Soomro, Fozia Soomro, Dr. Dastar Ali Chandio, & Bakhtawar Jatoi. (2026). Digital Deception in Geopolitical Crises: The Role of AI-Generated Fake News in the US–Iran Conflict. Research Journal for Social Affairs, 4(1), 123–128. https://doi.org/10.71317/RJSA.004.01.0684

Erin E., H. & Kent State University, USA. (2021). Self-Presentation in Social Media: Review and Research Opportunities. Review of Communication Research, 9, 80–98. https://doi.org/10.12840/ISSN.2255-4165.027

Gutiérrez-Caneda, B., Lindén, C.-G., & Vázquez-Herrero, J. (2024). Ethics and journalistic challenges in the age of artificial intelligence: Talking with professionals and experts. Frontiers in Communication, 9, 1465178. https://doi.org/10.3389/fcomm.2024.1465178

Kawakami, A., Coston, A., Heidari, H., Holstein, K., & Zhu, H. (2024). Studying Up Public Sector AI: How Networks of Power Relations Shape Agency Decisions Around AI Design and Use. Proceedings of the ACM on Human-Computer Interaction, 8(CSCW2), 1–24. https://doi.org/10.1145/3686989

Shaheed Benazir Bhutto University, Shaheed Benazirabad, & Soomro, F. (2025). Role of Broadcasting Media to Promote Climate Change Awareness: A Survey from Shaheed Benazir Bhutto University Shaheed Benazirabad. Journal of Media & Communication, 6(1), 53–69. https://doi.org/10.46745/ilma.jmc.2025.06.01.04

Soomro, F., Maheen, M. S., Batool, W., & Jalbani, R. (2026). Combatting Fake News on Social Media Using AI: A Study Among Public Sector University Students of Sindh Province. ACADEMIA International Journal for Social Sciences, 5(3), 151–161. https://doi.org/10.63056/academia.5.3.2026.1626