***Journal of Religion & Society* (JR&S)**
Available Online:
https://islamicreligious.com/index.php/Journal/index
Print ISSN: 3006-1296Online ISSN: 3006-130X
Platform & Workflow by: Open Journal Systems
https://doi.org/10.5281/zenodo.18418195

Deep-fake Exposure, Artificial Intelligence and Online Harassment among Generation-Z in Pakistan: A Quantitative Study of Awareness, Victimization and Psychological Impact

Muhammad Waqas
Criminologist, The University of Lahore-Pakistan
Husnain Hameed Awan
Lecturer, Department of Criminology,
Faculty of Social Sciences, The University of Lahore-Pakistan
Corresponding Author Email: husnain.hameed@crim.uol.edu.pk
Muhammad Atif Nazir
Assistant Professor, M A Raoof College of LAW
Faculty of Law, The University of Lahore-Pakistan
Dr. Muhammad Faisal Khan
Assistant Professor, Department of Criminology
Faculty of Social Sciences, The University of Lahore-Pakistan

Abstract

The rapid development of artificial intelligence has introduced new forms of digital manipulation, among which deep-fake technology represents one of the most serious emerging threats to online safety. Deep-fakes enable the creation of highly realistic but fabricated audio-visual content that can be misused for harassment, identity theft, blackmail and cybercrime. Generation-Z, as the most digitally active generation, is particularly vulnerable to such risks. This study examines the level of awareness, exposure and experiences of Generation-Z with deep-fake technology and its relationship with online harassment and psychological distress in Pakistan.

Using a quantitative survey design, data were collected from 350 Generation-Z social media users aged 18 to 26 years. The study assessed participants' awareness of deep-fake technology, frequency of exposure to manipulated media, experiences of online harassment, psychological impact and knowledge of cybercrime laws. Descriptive statistics and correlational analysis were used to explore patterns of victimization and vulnerability.

Findings indicate that although most respondents had heard of deep-fake technology, only a minority possessed a clear understanding of how it operates or its legal implications. Exposure to deep-fake content was widespread across platforms such as TikTok, Instagram and YouTube and higher exposure was significantly associated with increased experiences of online harassment and psychological distress. Female respondents reported higher levels of fear and victimization compared to male respondents. Awareness of Pakistan's cybercrime laws remained moderate to low, limiting effective reporting and legal recourse.

The study highlights the urgent need for digital literacy programs, targeted awareness campaigns and updated legal frameworks to protect young users from the growing misuse of deep-fake technology. Strengthening education, law enforcement capacity and platform accountability is essential to safeguard Generation-Z in the evolving digital environment.

Keywords: *Deep-fakes, Online Harassment, Cybercrime, Generation-Z, Digital Literacy, Pakistan*

1. Introduction

The rapid advancement of artificial intelligence (AI) has transformed digital communication and media production in unprecedented ways. Among the most controversial innovations of recent years is deep-fake technology, which uses deep learning algorithms to create highly realistic but fabricated images, videos and audio recordings (Chesney & Citron, 2019). While such technology has legitimate applications in entertainment, education and creative industries, its malicious use has become a growing concern for cybersecurity, privacy and social well-being (Kietzmann et al., 2020).

Deep-fakes allow perpetrators to manipulate a person's likeness or voice without consent, enabling new forms of online harassment, identity theft, blackmail, fraud and reputational damage (Citron & Chesney, 2019). Unlike traditional forms of cybercrime, deep-fake manipulation produces synthetic "evidence" that appears authentic, making it difficult for victims to prove falsification and for authorities to verify authenticity (Vaccari & Chadwick, 2020). As a result, deep-fakes represent not only a technological challenge but also a profound social and legal threat.

Generation-Z (Gen-Z), commonly defined as individuals born between the mid-1990s and early 2010s, is the first generation to grow up entirely in the digital era (Pew Research Center, 2019). Gen-Z users rely heavily on social media platforms such as TikTok, Instagram, YouTube and Snapchat for communication, education and identity formation (Anderson & Jiang, 2018). This constant digital engagement makes them particularly vulnerable to manipulated content and online victimization (Patchin & Hinduja, 2018).

Existing research on deep-fakes has largely focused on technical detection methods, political misinformation and high-profile celebrity cases (Maras & Alexandrou, 2019; Westerlund, 2019). Far less attention has been paid to the everyday experiences of young social media users, especially in developing countries such as Pakistan. There remains a significant empirical gap in understanding how frequently Gen-Z is exposed to deep-fake content, how such exposure relates to online harassment and what psychological consequences result from victimization.

This study seeks to address this gap by examining the awareness, exposure and experiences of deep-fake-related harassment among Generation-Z social media users in Pakistan. By providing quantitative evidence from a South Asian context, this study contributes to the growing field of digital criminology and offers practical implications for education, policy and law enforcement.

2. Review of the Literature

2.1 Deep-fake Technology and Digital Manipulation

Deep-fake technology refers to the use of deep learning algorithms, particularly Generative Adversarial Networks (GANs), to generate or manipulate audio-visual content that appears authentic (Goodfellow et al., 2014; Westerlund, 2019). By training on large datasets of images and voice recordings, AI systems can replicate facial expressions, speech patterns and gestures with remarkable accuracy (Maras & Alexandrou, 2019).

Since around 2018, the accessibility of deep-fake tools has increased rapidly, allowing even non-experts to create manipulated content using open-source software and mobile applications (Kietzmann et al., 2020). Scholars warn that the declining cost and increasing realism of deep-fakes pose a serious threat to digital trust and information integrity (Chesney & Citron, 2019).

2.2 Deep-fakes and Online Harassment

Deep-fakes have increasingly been weaponized for malicious purposes, particularly in the form of non-consensual intimate imagery, impersonation and blackmail (Citron & Franks, 2014; Henry & Powell, 2018). Research indicates that women are

disproportionately targeted through sexualized deep-fake abuse, leading to severe reputational harm and psychological trauma (Powell & Henry, 2017).

Online harassment more broadly includes cyberbullying, stalking, impersonation and defamation (Patchin & Hinduja, 2018). Youth and young adults are especially vulnerable due to their high levels of online activity and peer interaction (Livingstone & Smith, 2014). Studies consistently link cyber victimization with anxiety, depression and social withdrawal (Kowalski et al., 2014).

2.3 Gen-Z, Awareness and Digital Vulnerability

Despite being technologically proficient, Gen-Z often lacks formal training in critical media literacy and cyber law (Pew Research Center, 2019). Research suggests that many young users recognize the term "deep-fake" but fail to understand its technical mechanisms or legal consequences (Vaccari & Chadwick, 2020).

In Pakistan, the Prevention of Electronic Crimes Act (PECA) 2016 criminalizes online harassment and image-based abuse, yet does not explicitly address AI-generated content (Government of Pakistan, 2016). Limited awareness of legal protections further weakens the ability of young victims to seek justice (Shah & Ahmed, 2020).

3. Methodology

3.1 Research Design

This study employed a quantitative, cross-sectional survey design to examine awareness, exposure and experiences of deep-fake-related harassment among Generation-Z social media users. The design was descriptive and correlational in nature, aiming to identify patterns and relationships among key variables.

3.2 Sample and Participants

The sample consisted of **350 Generation-Z respondents** aged between 18 and 26 years. Participants were recruited from university students and young professionals in urban and semi-urban areas. Inclusion criteria required active use of at least one social media platform within the past three months.

Demographically, the sample included 40% males, 58% females and 2% non-binary or prefer not to say. The majority were undergraduate students residing in urban areas and belonging to middle-income families.

3.3 Data Collection Instrument

Data were collected using a structured online questionnaire consisting of five sections:

1. Demographic information
2. Awareness of deep-fake technology
3. Exposure to manipulated content
4. Experiences of online harassment
5. Psychological impact and legal awareness

Most items were measured using 5-point Likert scales ranging from "Strongly disagree" to "Strongly agree."

3.4 Ethical Considerations

Participation was voluntary and anonymous. Informed consent was obtained digitally. No identifying information was collected. Participants were informed of their right to withdraw at any time and provided with information on counseling services for distress.

4. Results

4.1 Awareness of Deep-fake Technology

Analysis of awareness levels indicated that **72.3% (n = 253)** of respondents reported having heard of deep-fake technology prior to the survey, while **27.7% (n = 97)** had no prior awareness. However, detailed knowledge was limited. Only **31.4% (n = 110)** of

respondents correctly identified the technical process through which deep-fakes are created, suggesting that awareness was largely superficial rather than technical.

Educational differences were evident. Postgraduate respondents reported significantly higher awareness scores (M = 3.62, SD = 0.71) compared to undergraduate respondents (M = 3.11, SD = 0.84), $t(348) = 4.87$, $p < .001$. Gender differences were also observed. Female respondents reported significantly higher concern regarding the potential misuse of deep-fake technology (M = 3.74, SD = 0.69) than male respondents (M = 3.29, SD = 0.77), $t(346) = 3.92$, $p < .001$.

Awareness of the legal status of manipulated content remained particularly low. Only **24.6% (n = 86)** of respondents were aware that the creation or circulation of manipulated images without consent may constitute a criminal offence under Pakistani law.

4.2 Exposure to Deep-fake Content

Exposure to deep-fake content was highly prevalent within the sample. Approximately **68.9% (n = 241)** of respondents reported encountering manipulated videos or images at least once per month, while **34.2% (n = 120)** reported exposure on a weekly basis.

Social media platforms were the primary sources of exposure. A majority of respondents reported encountering such content on TikTok (61.7%), Instagram (58.9%) and YouTube (54.3%). In contrast, lower exposure was reported on Facebook (32.5%) and Twitter/X (28.1%).

Despite frequent exposure, confidence in detecting manipulated content remained low. Only **19.4% (n = 68)** of respondents reported high confidence in their ability to distinguish between authentic and fake content. Correlation analysis revealed a significant positive association between frequency of exposure and distrust in online information ($r = .41$, $p < .001$), as well as between exposure and anxiety regarding digital identity ($r = .37$, $p < .001$).

4.3 Online Harassment Experiences

With respect to online harassment, **39.7% (n = 139)** of respondents reported having personally experienced at least one form of online harassment involving impersonation, image manipulation or defamation. In addition, **57.1% (n = 200)** reported having witnessed such incidents affecting friends or peers.

Gender-based differences were statistically significant. Female respondents were significantly more likely to report fear of victimization than male respondents, $\chi^2(1) = 12.84$, $p < .001$. Specifically, **46.8% of female respondents** reported fear of becoming a target of manipulated sexual content, compared to **21.3% of male respondents**.

The most commonly reported forms of harassment included impersonation (28.6%), non-consensual image sharing (24.3%) and reputational defamation (19.7%). These findings indicate that deep-fake technology is closely linked to multiple forms of online victimization.

4.4 Psychological Impact

Among respondents who reported direct victimization or close exposure to manipulated content, psychological consequences were substantial. Approximately **44.6% (n = 156)** reported moderate to high levels of anxiety, **38.9% (n = 136)** reported persistent embarrassment and **32.1% (n = 112)** reported social withdrawal.

Regression analysis indicated that exposure to deep-fake content significantly predicted psychological distress ($\beta = .39$, $p < .001$), even after controlling for gender and frequency of social media use. This suggests that deep-fake exposure independently contributes to psychological harm among young users.

4.5 Awareness of Cybercrime Laws

Awareness of cybercrime laws remained limited across the sample. Only **29.1% (n = 102)** of respondents reported familiarity with the Prevention of Electronic Crimes Act (PECA) 2016 and only **17.4% (n = 61)** knew the correct procedure for filing an online complaint with the FIA Cybercrime Wing.

Respondents with higher levels of legal awareness reported significantly lower fear of victimization (M = 2.91) compared to those with low legal awareness (M = 3.47), $t(348) = 4.21$, $p < .001$. This finding indicates that legal knowledge may play a protective role against fear and perceived vulnerability.

5. Discussion

The present study provides quantitative evidence that deep-fake technology has become a significant risk factor for online harassment and psychological harm among Generation-Z in Pakistan. The findings demonstrate that although a majority of respondents were aware of the term "deep-fake," only a minority possessed accurate technical or legal understanding, confirming that awareness remains largely superficial.

The strong association between exposure and online harassment supports previous international research indicating that manipulated content amplifies existing forms of cyber victimization (Chesney & Citron, 2019; Henry & Powell, 2018). The significant correlation between exposure and psychological distress further confirms that deep-fakes represent not merely a technological novelty but a serious public mental health concern.

Gender differences observed in this study are consistent with prior findings that women are disproportionately targeted through sexualized digital abuse (Powell & Henry, 2017). Female respondents reported significantly higher fear of victimization and higher exposure to manipulated intimate content, highlighting the gendered nature of deep-fake misuse.

Low awareness of cybercrime laws further exacerbates vulnerability. Despite the existence of PECA 2016, most respondents were unaware of reporting mechanisms or legal remedies. These finding echoes previous studies in Pakistan indicating weak public knowledge of cyber law and limited trust in formal reporting systems (Shah & Ahmed, 2020).

Taken together, these results suggest that technical innovation has outpaced legal awareness and institutional capacity, creating a regulatory and educational gap that places young users at substantial risk.

6. Conclusion

This study provides robust empirical evidence that deep-fake exposure is significantly associated with online harassment and psychological distress among Generation-Z in Pakistan. Awareness remains superficial, exposure is widespread across major platforms and legal knowledge is critically limited.

The findings underscore the urgent need for:

1. Structured digital literacy programs in universities and colleges.
2. Targeted awareness campaigns on deep-fake risks.
3. Strengthened cyber law education and reporting mechanisms.
4. Greater platform accountability for rapid detection and removal.

Without coordinated educational, legal and technological interventions, deep-fake technology is likely to become a major driver of future cyber victimization.

Reference

Anderson, M., & Jiang, J. (2018). *Teens, social media & technology 2018*. Pew Research Center.

Chesney, R., & Citron, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy and national security. *California Law Review, 107*(6), 1753–1819.

Citron, D. K., & Chesney, R. (2019). Deep fakes: A looming challenge for privacy, democracy and national security. *California Law Review, 107*(6), 1753–1819.

Citron, D. K., & Franks, M. A. (2014). Criminalizing revenge porn. *Wake Forest Law Review, 49*, 345–391.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. In *Advances in Neural Information Processing Systems* (pp. 2672–2680).

Government of Pakistan. (2016). *Prevention of Electronic Crimes Act (PECA) 2016*. Islamabad: Government of Pakistan.

Henry, N., & Powell, A. (2018). Technology-facilitated sexual violence: A literature review of empirical research. *Trauma, Violence, & Abuse, 19*(2), 195–208.

Kietzmann, J., Lee, L. W., McCarthy, I. P., & Kietzmann, T. C. (2020). Deepfakes: Trick or treat? *Business Horizons, 63*(2), 135–146.

Kowalski, R. M., Giumetti, G. W., Schroeder, A. N., & Lattanner, M. R. (2014). Bullying in the digital age: A critical review. *Psychological Bulletin, 140*(4), 1073–1137.

Livingstone, S., & Smith, P. K. (2014). Annual research review: Harms experienced by child users of online and mobile technologies. *Journal of Child Psychology and Psychiatry, 55*(6), 635–654.

Maras, M.-H., & Alexandrou, A. (2019). Determining authenticity of video evidence in the age of artificial intelligence. *International Journal of Evidence & Proof, 23*(3), 255–262.

Patchin, J. W., & Hinduja, S. (2018). Sexting as an emerging concern for adolescent health. *JAMA Pediatrics, 172*(3), 256–257.

Pew Research Center. (2019). *Defining generations: Where Millennials end and Generation Z begins*.

Powell, A., & Henry, N. (2017). *Sexual violence in a digital age*. Palgrave Macmillan.

Shah, R., & Ahmed, M. (2020). Cybercrime and legal responses in Pakistan. *Pakistan Journal of Criminology, 12*(1), 45–62.

Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video. *Social Media + Society, 6*(1), 1–13.

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