

AI-Generated Photo Manipulation and Visual Disinformation Among Muslim University Students in Indonesia

Marlena¹, Septi Gumindari²

^{1,2} Universitas Islam Negeri Siber Syekh Nurjati Cirebon, Indonesia

Email: lenamrlnn@gmail.com; septigumindari@uinssc.ac.id

Proceedings of the Kalijaga International Conference on Education 2026

Abstract. The rapid development of generative artificial intelligence has intensified the production of highly realistic manipulated images, increasing the spread of visual disinformation and raising critical concerns beyond technical domains. While existing studies largely focus on detection technologies, limited attention has been paid to the psychological and moral dimensions of AI-generated visual manipulation, particularly within religious higher education contexts. This study explores the forms of AI-generated photo manipulation encountered by Muslim university students, the psychological responses elicited by exposure to such content, and the ethical reasoning underpinning their judgments and coping strategies. Employing an exploratory qualitative approach, data were collected from five undergraduate students at UIN Siber Syekh Nurjati Cirebon through in-depth interviews and semi-structured online questionnaires. The findings indicate that AI-manipulated images predominantly involve fabricated interactions with public figures, altered representations of celebrities and political actors, and subtle visual anomalies signaling artificial construction. Exposure to these images generates psychological effects including cognitive ambiguity, anxiety, fear of identity misuse, and declining epistemic trust in digital information. Drawing on Islamic ethical principles, participants largely interpret deceptive AI-generated photo manipulation as a violation of shiddiq (truthfulness) and amanah (moral responsibility), while expressing conditional acceptance of parody or entertainment when transparency, consent, and non-harm are ensured. To mitigate disinformation risks, students adopt integrated coping strategies encompassing individual verification practices, peer-based clarification, and critical digital literacy informed by moral awareness (tabayyun). This study contributes to global discussions on AI ethics and education by highlighting the need for value-based digital literacy frameworks that integrate psychological resilience and moral reasoning in responding to AI-driven visual disinformation within higher education.

Keywords: *AI-generated images; digital literacy; Islamic ethics; psychological impact; visual disinformation.*

Abbreviations: AI (Artificial Intelligence); UIN (Universitas Islam Negeri / State Islamic University); UU ITE (Undang-Undang Informasi dan Transaksi Elektronik / Law on Electronic Information and Transactions).

INTRODUCTION

Social media ideally serves as a normative public sphere governed by truthfulness (shiddiq) and visual authenticity, wherein uploaded materials are expected to reflect empirical reality. However, this normative ideal has been severely disrupted by the rapid advancement of generative artificial intelligence (AI), which has democratized the creation of hyper-realistic visual manipulations (deepfakes). Deepfake technology enables the synthesis and manipulation of visual and auditory content to fabricate highly realistic depictions of individuals saying or performing actions that never occurred in reality [1]. On a global scale, the malicious weaponization of synthetic media has destabilized democratic discourse, fueled political disinformation, targeted public figures, and systematically eroded the boundaries separating empirical reality from synthetic fabrication. This shift transcends purely computational and algorithmic concerns; it has evolved into a profound epistemological crisis that challenges human cognitive faculties to authenticate visual evidence in digital environments.

Within higher education contexts, the friction between hyper-realistic synthetic imagery and the imperative for verifiable truth induces severe cognitive disruption while eroding epistemic trust. University students frequently report an acute loss of confidence in digital information, leading to pervasive skepticism and systemic "trust issues" across social platforms.

Prior scholarship has examined digital hoaxes and information behavior among students. Nurrahmi and Syam [2] investigated undergraduate information behavior regarding social media hoaxes, demonstrating that students often exhibit passive verification behaviors and require structured media literacy interventions. However, their inquiry focused primarily on conventional textual and verbal hoaxes, leaving the affective, psychological, and moral dimensions of hyper-realistic AI-driven visual fabrications largely

unexplored. In another domain, Laksmi and Andry [3] addressed the misuse of generative AI in gender-based online violence, emphasizing how deepfake technologies disproportionately target women. While their work highlighted digital vulnerability, it focused primarily on gendered harassment rather than generalized psychological impacts—such as epistemic skepticism, cognitive dissonance, and broader religious coping frameworks. Furthermore, Puspitasari [4] conducted a normative, theological examination of Quranic communication ethics—specifically honesty (*shiddiq*) and justice—to mitigate slander and disinformation. Nonetheless, theological inquiries remain predominantly theoretical and lack empirical grounding in the lived experiences and behavioral responses of Muslim students confronting synthetic media.

To address these empirical and theoretical gaps, this study investigates how Muslim university students experience, interpret, and navigate AI-generated photo manipulation. Specifically, this research pursues four core objectives: (1) to identify the typologies of AI-manipulated photographs circulating on social media that carry visual disinformation risks; (2) to describe the psychological impacts (e.g., ambiguity, anxiety, epistemic skepticism) elicited among students; (3) to analyze students' ethical perceptions regarding AI visual alterations from an Islamic moral perspective; and (4) to explore the adaptive coping strategies employed by students to mitigate visual disinformation while preserving moral integrity. The primary novelty of this research lies in establishing an integrated "Triadic Value-Based Coping Framework" that synthesizes individual forensic verification (*khadzr*), peer-deliberative validation (*musyawarah*), and Quranic epistemological inquiry (*tabayyun*) in mitigating generative AI visual disinformation within higher education ecosystems.

METHODS

Study area

This research was conducted at Universitas Islam Negeri (UIN) Siber Syekh Nurjati Cirebon, located in Cirebon, West Java, Indonesia. As Indonesia's designated pilot Cyber Islamic University (CIU), this institution provides an exceptionally strategic empirical environment where students actively engage with advanced digital learning systems while being firmly anchored in classical and contemporary Islamic epistemological and ethical traditions.

Procedures

Sampling and Informant Selection

Methodologically, this study was structured as an in-depth exploratory case study designed to uncover the nuanced cognitive, psychological, and moral lived experiences of students confronted with synthetic media [5]. Primary data were gathered through purposive sampling involving five key undergraduate informants selected based on rigorous criteria: (1) high digital literacy and daily engagement on visual social media platforms exceeding four hours; (2) active participation in digital content production or online academic discourse; and (3) verified direct encounters with AI-generated visual disinformation on social media. Although qualitative case studies frequently operate with focused cohorts, the sample size of five informants was strictly governed by the principle of data saturation, whereby iterative data collection from the fifth informant yielded informational redundancy and generated no new thematic codes regarding visual typologies, psychological states, or Islamic ethical responses. In accordance with methodological standards [6], primary empirical data were complemented by secondary sources, including peer-reviewed literature, statutory frameworks (e.g., the Indonesian UU ITE), and public deepfake case records.

Data Collection Techniques

Methodological triangulation was employed to ensure data richness and structural validity [7], [8]. Data collection incorporated two primary modalities:

In-Depth Qualitative Interviews: Administered to three key informants (informants MF, SAB, and AUA) to elicit contextual, detailed narratives concerning psychological disruption, visual discernment processes, and ethical reasoning.

Semi-Structured Online Questionnaires: Deployed via Google Forms to two additional informants (including informant MAN) to collect targeted, consistent comparative responses regarding ethical judgments and coping behaviors.

Documentary Analysis: Documenting relevant regulatory frameworks, institutional guidelines, and published literature to corroborate empirical field data.

Data analysis

Qualitative data were analyzed using an interactive analysis model adhering to the Spradley, Miles, and Huberman framework [9], executed across three iterative, interdependent stages:

Data Reduction: Raw interview transcripts and questionnaire responses were systematically filtered, coded, and categorized according to the four core research themes: visual manipulation typologies, psychological responses, Islamic moral reasoning, and coping strategies.

Data Display: Structured thematic matrices and narrative descriptions were formulated to organize students' empirical accounts coherently.

Conclusion Drawing and Verification: Emergent patterns and conceptual relationships were cross-checked and contextualized against theological doctrines, psychological literature, and secondary documentation to ensure rigorous credibility, objectivity, and descriptive validity.

RESULTS AND DISCUSSION

Typologies of AI-Generated Photo Manipulation on Social Media

The empirical findings reveal that undergraduate students at UIN Siber Syekh Nurjati Cirebon frequently encounter various forms of AI-manipulated photographs across social media platforms including TikTok, Instagram, and X (formerly Twitter). Based on the primary data, these visual manipulations are classified into three primary typologies.

Fabricated Interactions with Public Figures and Idol Culture

The most common form of AI image manipulation identified by informants involves the synthetic insertion of ordinary individuals into fabricated, intimate settings with public figures, K-Pop idols, or romantic interests. Informants noted that while these images frequently originate as lighthearted social media trends, they carry inherent risks of deception when circulated without context. Informant SAB described this phenomenon: "There are K-pop idols... girls make them just for fun on TikTok. Her photo was edited hugging a K-pop idol, and another friend edited herself with her crush." (Interview with female undergraduate student SAB, 3rd Semester, October 17, 2025).

Informants observed that despite being intended as humorous or expressive content, these fabrications are frequently misinterpreted by other users as genuine visual documentation, leading to reputational damage or non-consensual use of personal likenesses.

Manipulated Representations of Celebrities and Political Figures

Informants highlighted a more concerning category involving the deliberate manipulation of political figures and celebrities to engineer public perception. In the socio-political realm, AI is utilized to manufacture synthetic imagery depicting political actors performing religious rituals for superficial political branding, or conversely, engaging in controversial acts to provoke sectarian outrage. Informant MF stated: "...politics is just as prevalent, especially when political tensions are high; many people fabricate scandals from edited photos of political figures." (Interview with female undergraduate student MF, 3rd Semester, October 17, 2025).

Corroborating this observation, informant AUA noted: "These images usually relate to celebrities, political figures, and religious issues. They are frequently used to spread hoaxes, generate sensationalism, or provoke public anger." (Interview with female undergraduate student AUA, 7th Semester, October 19, 2025).

Physical Anomalies and Visual Aesthetic Discrepancies

Despite advances in photorealism, informants reported that AI-generated imagery consistently displays detectable visual anomalies and artificial artifacts. Students evaluate fine-grained details to identify artificial construction, specifically scrutinizing anatomical proportions, background stability, and skin texture consistency. Informant MF remarked: "Usually in AI photos, there is something slightly off—for example, the hands or the background keeps morphing inconsistently." (Interview with female undergraduate student MF, 3rd Semester, October 17, 2025).

Informant SAB highlighted the unnatural visual smoothness characteristic of generative models: "Sometimes it looks like an exaggerated filter, unnaturally smooth." (Interview with female undergraduate student SAB, 3rd Semester, October 17, 2025).

Additional visual cues reported by respondents include distorted peripheral objects, irregular light reflections, and illegible pseudo-text on backgrounds. Table 1. Typologies of AI-Manipulated Imagery Identified by Students

Typology Category	Manifestation & Content Modality	Perceived Disinformation Risk	Key Visual Indicators (Artifacts)
Parasocial & Idol Manipulation	Synthetic photos depicting fabricated physical proximity or affection with celebrities and peers.	Misrepresentation of interpersonal relationships, non-consensual likeness appropriation.	Contour blurring around subjects, inconsistent lighting between figures.
Socio-Political & Religious Fabrications	Staged political scandals, synthetic piety displays, and fabricated transgressive behavior.	Public polarization, weaponized sectarian outrage, institutional delegitimization.	Contextual incongruities, spatial background warping, inconsistent shadows.
Anatomical & Aesthetic Anomalies	Hyper-realistic generative synthetic portraits and artificial scene renderings.	Deception of sensory perception, epistemic confusion regarding visual truth.	Incoherent hand/finger anatomy, unnatural skin smoothing, unreadable background text.

Psychological Impacts of Exposure to AI-Manipulated Imagery

Exposure to hyper-realistic AI-manipulated photographs induces complex cognitive and affective reactions among students, progressing from perceptual confusion to heightened anxiety and systemic epistemic skepticism.

Cognitive Ambiguity and Initial Astonishment

Informants reported an initial reaction of shock and disorientation upon confronting manipulated images whose realism closely matches authentic photographs. The inability to immediately distinguish authentic photos from synthetic fabrications causes cognitive friction. Informant MF expressed: "...it looks so identical that it creates confusion. Initially, I was shocked, like 'wow, how is this possible?' Over time, I've had to learn how to distinguish them." (Interview with female undergraduate student MF, 3rd Semester, October 17, 2025).

Heightened Anxiety, Fear, and Identity Vulnerability

Beyond cognitive disorientation, informants experienced genuine anxiety and emotional vulnerability regarding the security of their personal digital identities. The realization that personal photographs can be hijacked, altered, or weaponized for cyberbullying and extortion creates a climate of apprehension. Informant MF shared: "Yes, I feel afraid because AI is increasingly used for malicious acts; I fear becoming a victim myself." (Interview with female undergraduate student MF, 3rd Semester, October 17, 2025).

This anxiety is notably pronounced among female respondents, who expressed acute concern regarding the safety of women's personal portraits in digital spaces. Informant SAB noted: "What makes me deeply uneasy is seeing photos of women alone on social media, fearing they might be maliciously edited into inappropriate content." (Interview with female undergraduate student SAB, 3rd Semester, October 17, 2025).

Erosion of Epistemic Trust (Visual Skepticism)

Repeated exposure to manipulative AI imagery causes a decline in students' confidence in digital information, fostering a pervasive state of skepticism toward online visual content. Informant SAB described this condition: "...nowadays, it is so difficult to believe anything; it is like having trust issues with information." (Interview with female undergraduate student SAB, 3rd Semester, October 17, 2025).

Informants MF and AUA corroborated this erosion of trust, noting that visual skepticism induces constant hesitation and fear of making false moral assumptions (*su'udzan*). Informant MF stated: "It really makes me doubt; I become deeply afraid of misunderstandings and unjustly suspecting others." (Interview with female undergraduate student MF, 3rd Semester, October 17, 2025).

Informant AUA affirmed: "Yes, it makes me far more doubtful of any photo circulating on social media." (Interview with female undergraduate student AUA, 7th Semester, October 19, 2025).

Table 2. Psychological Dimensions Elicited by AI Visual Disinformation

Psychological Dimension	Primary Trigger/Mechanism	Subjective Manifestation	Behavioral/Social Consequence
Cognitive Ambiguity	Inability to immediately distinguish authentic from synthetic visual media.	Mental disorientation, astonishment, perceptual friction.	Cognitive hesitation, delayed processing of breaking visual news.
Digital Anxiety & Fear	Awareness of personal vulnerability to non-consensual image manipulation.	Fear of identity misuse, gendered apprehension regarding personal portraits.	Self-censorship, withdrawal of personal photographs from social media.
Epistemic Distrust	Proliferation of unverified, deceptive visual media across digital platforms.	Chronic visual skepticism ("trust issues"), fear of unjust suspicion (<i>su'udzan</i>).	Broad rejection of visual evidence, generalized institutional cynicism.

Islamic Ethical Perspectives on AI Photo Manipulation

Students evaluated generative visual manipulation not merely as a technical phenomenon, but through the ethical frameworks of Islamic moral philosophy.

Manipulation as Deception: Violations of Shiddiq and Amanah

Informants evaluated unclarified and deceptive AI photo manipulation as an explicit violation of Islamic ethical principles, categorizing it as a form of falsehood (*kizb*). Presenting a synthetic construct as empirical reality is viewed as an infringement on truthfulness (*shiddiq*) and moral responsibility (*amanah*). Informant MF asserted: "If the intention is to deceive or post without clarification, that is undeniably a form of falsehood." (Interview with female undergraduate student MF, 3rd Semester, October 17, 2025).

Respondents underscored that digital image generation carries moral accountability, as fabricating false representations conceals truth and risks generating communicative sins and perpetual transgression (*dosa jariyah*).

Conditional Exceptions: Entertainment, Parody, and Memes

While deceptive manipulation is condemned, informants made explicit moral exceptions for content created strictly for parody, creative entertainment, or memes, provided specific ethical boundaries are respected. Informant MF explained: "If it is strictly for entertainment, it remains tolerable, unless it is generated without clarification or maliciously." (Interview with female undergraduate student MF, 3rd Semester, October 17, 2025).

Informants emphasized that humor remains morally acceptable only when accompanied by transparent disclaimers and explicit consent from the individuals depicted. Informant MAN stated: "It is permissible provided one does not steal another's photo, and if using someone's likeness, explicit permission must be obtained." (Questionnaire response with female undergraduate student MAN, 5th Semester).

Moral Responsibility and Stewardship of Muslim Students

Informants affirmed that their status as Muslim university students confers an ethical duty to uphold integrity and actively counteract digital slander (*fitnah*). Informant AUA elaborated: "...In Islam, upholding truth, avoiding slander (*fitnah*), and safeguarding human dignity are core imperatives. Therefore, Muslim university students bear a vital responsibility as role models in utilizing technology wisely." (Interview with female undergraduate student AUA, 7th Semester, October 19, 2025).

This moral consciousness functions as an internal barrier preventing students from producing, consuming, or sharing manipulative synthetic media.

Adaptive Coping Strategies Against AI-Generated Disinformation

To mitigate the risks of visual disinformation and manage psychological distress, students implement a multi-tiered coping framework combining technical scrutiny, social verification, and value-based digital literacy.

Individual Verification Practices (Cross-Checking)

Informants primarily rely on individual verification routines upon encountering suspicious visual media. These actions include checking the credibility of the uploading account, analyzing public comments, performing reverse image searches, and cross-referencing claims across multiple platforms. Informant SAB shared: "I check the source and review the account that uploaded it. Then I read the comments to see if someone pointed out that it is AI. I once saw an image that looked completely real, but the comment section revealed it was AI." (Interview with female undergraduate student SAB, 3rd Semester, October 17, 2025).

Informant MF detailed cross-platform information retrieval habits: "When I encounter sensitive political or religious news, I research it multiple times across GPT, Gemini, and Google News." (Interview with female undergraduate student MF, 3rd Semester, October 17, 2025).

Communal Clarification and Peer Deliberation

Informants emphasized that individual scrutiny is reinforced by peer networks and social circles. Collaborative deliberation allows students to pool technical expertise and verify dubious images collectively before drawing conclusions. Informant MF noted: "Friends really need to help each other out, especially if someone is overly convinced by an AI photo that has become a public issue." (Interview with female undergraduate student MF, 3rd Semester, October 17, 2025).

Informant AUA affirmed: "Peers serve as discussion spaces to exchange information and perspectives, especially when certain friends possess stronger technical literacy or access to credible sources." (Interview with female undergraduate student AUA, 7th Semester, October 19, 2025).

Value-Based Digital Literacy: Operationalizing Tabayyun

Informants identified comprehensive digital literacy coupled with internalized moral awareness (tabayyun) as the most effective defense against visual deception. Informant SAB stated: "There must be structured education along with individual moral awareness." (Interview with female undergraduate student SAB, 3rd Semester, October 17, 2025).

Informant AUA concluded: "Learning digital literacy is extremely vital, especially for university students who constantly confront diverse streams of information on social media." (Interview with female undergraduate student AUA, 7th Semester, October 19, 2025).

Table 3. Multi-Layered Coping Strategy Framework Adopted by Students

Strategy Level	Operational Mechanism	Specific Verification Actions	Underlying Islamic Value
Individual Verification	Technical scrutiny and source provenance analysis.	Inspecting uploader profile and metadata. Reading public comment threads for debunking cues. Cross-querying search engines and AI tools.	Shiddiq (Truthfulness) Khadzr (Vigilance)
Peer/Social Deliberation	Communal epistemic triangulation.	Discussing suspicious imagery within peer networks. Seeking input from technically skilled peers. Communal clarification of viral content.	Musyawah (Deliberation) Ta'awun (Mutual Support)
Value-Based Digital Literacy	Cognitive discernment anchored in moral principles.	Applying critical evaluation to digital messages. Exercising restraint against impulsive sharing. Safeguarding personal digital identity assets.	Tabayyun (Systematic Verification) Amanah (Moral Stewardship)

Discussion

The rapid evolution of generative artificial intelligence (AI) has initiated a profound paradigm shift in how digital visual media is produced, perceived, and authenticated. While conventional discussions surrounding AI manipulation have predominantly centered on computational detection models, the empirical findings of this study underscore that visual disinformation operates at the intersection of technological realism, psychological vulnerability, and moral-theological reasoning. By analyzing the lived experiences of Muslim university students at UIN Siber Syekh Nurjati Cirebon, this research elucidates how hyper-realistic synthetic media disrupts human cognition, destabilizes epistemic trust, and activates distinctive faith-based coping mechanisms.

Technological Realism vs. Visual Epistemology: The Degradation of Photographic Truth

Historically, photographic images served as indexical evidence of empirical reality—a visual testament that "what was captured had indeed occurred". However, the emergence of generative adversarial networks (GANs) and diffusion models has severed this ontological link between imagery and truth. As demonstrated in the empirical data, students frequently encounter manipulations ranging from parasocial fan edits to fabricated political and religious spectacles. The weaponization of synthetic media to simulate religious piety or orchestrate political scandals highlights how AI is systematically leveraged for ideological manipulation, plain-folks propaganda, and public polarization [10], [11].

Although students currently rely on identifying visual artifacts—such as morphological inconsistencies in hands, hyper-smoothed skin textures, unreadable text, and disparate pupil lighting reflections [12], [13]—this reliance constitutes a fragile epistemic defense. As neural rendering algorithms continuously refine anatomical precision and physics-based lighting consistency, heuristic detection will become increasingly obsolete [14]. In digital epistemic theory, this vulnerability is recognized as an 'algorithmic epistemic bubble' where sensory intuition is systematically bypassed by hyper-realistic neural rendering [15]. Consequently, relying solely on human sensory perception risks leaving audiences vulnerable to sophisticated deceptive campaigns, emphasizing the urgent need to transition from passive visual observation to active, multi-source digital forensics [16], [17].

The Cognitive, Affective, and Gendered Dimensions of Digital Insecurity

The psychological trajectory reported by the participants demonstrates that exposure to synthetic disinformation is not merely an intellectual dilemma, but a visceral affective crisis. The progression from initial astonishment to cognitive ambiguity illustrates the cognitive overload individuals experience when

forced to continuously evaluate ambiguous sensory stimuli [18], [19]. When individuals can no longer instinctively distinguish fact from fabrication, mental fatigue ensues, making them susceptible to either uncritical acceptance or chronic cynicism [19], [20]. As theorized in information disorder paradigms [21], deepfake technology weaponizes uncertainty, causing citizens to disengage from verifiable news ecosystems altogether.

Furthermore, the study uncovers an acute sense of digital vulnerability and anxiety (khauf), particularly concerning non-consensual identity manipulation and cyberbullying [22], [23]. The expressed fear that personal portraits—especially those of female students—could be misappropriated for malicious or sexually explicit deepfakes reflects a well-founded awareness of gendered digital violence [3], [23]. This constant apprehension forces students into digital self-censorship, restricting their self-expression and agency in digital spaces.

On a broader epistemic level, repeated exposure to manipulated media induces systemic "trust issues" (su'udzan). This chronic skepticism extends beyond suspicious accounts to credible journalism and institutional channels [24], [20]. When epistemic trust collapses, the shared foundation of public discourse is eroded, creating an environment where truth becomes fragmented and public cynicism thrives [14], [21].

Islamic Epistemology and Communication Ethics: Theorizing Shiddiq, Amanah, and Parody

A central contribution of this study is the articulation of how Muslim undergraduate students utilize Islamic normative frameworks to evaluate synthetic media. Rather than treating AI manipulation as a morally neutral technological novelty, participants evaluated it through foundational Islamic ethical virtues: shiddiq (truthfulness), amanah (trustworthiness and moral responsibility), and the avoidance of kizb (falsehood) and fitnah (slander/defamation) [4], [25].

In Islamic communication ethics, the intentional presentation of a synthetic depiction as empirical reality is viewed as an explicit violation of shiddiq, as truthfulness demands strict congruence between an external assertion and objective reality [4], [26]. From a theological standpoint, this moral mandate is rooted in the Quranic doctrine of sensory accountability in Surah Al-Isra' [19]: 36, which explicitly dictates: "And do not pursue that of which you have no knowledge. Indeed, the hearing, the sight, and the heart—about all those one will be questioned." This verse establishes that visual perception and cognitive endorsement are not morally inconsequential acts; rather, accepting and propagating unverified synthetic imagery incurs direct spiritual culpability [4], [24]. Moreover, because digital media propagates exponentially, distributing deceptive AI content carries the severe ethical consequence of perpetual transgression (dosa jariyah), wherein the disseminator incurs continuous moral liability for every individual misled by the falsehood [4], [26].

Concurrently, the empirical data reveal an ethical distinction between malicious deceit and transformational satire or entertainment. Participants recognized that humor, memes, and parody do not inherently violate Islamic ethics, provided three non-negotiable criteria are met:

Explicit Transparency: The content must be clearly identified as synthetic/parody to prevent misleading the audience.

Prior Consent: The likeness of individuals must not be misappropriated without authorization [27], [28].

Absence of Harm (Lā Darar): The media must not inflict reputational damage, mockery, or harassment (fitnah) upon the depicted subjects [27], [25].

This ethical differentiation demonstrates that Muslim students do not adopt an anti-technological stance; rather, they advocate for a principled balance that accommodates creative expression while preserving human dignity (karamah) and social harmony.

Operationalizing Tabayyun: Towards an Integrated Moral-Cognitive Literacy Paradigm

To mitigate the risks of visual disinformation, participants developed an adaptive, multi-tiered coping strategy that bridges technical verification with socio-religious principles. The primary verification practices identified—source provenance checking, reverse image lookup, and cross-platform verification via large language models and search engines—align with modern media literacy models [2], [17], [29]. However, this empirical inquiry reveals that individual scrutiny is significantly reinforced by communal deliberation (musyawarah) and peer support networks [30], [31]. Peer communities function as informal verification bodies where technical knowledge is pooled to evaluate dubious content collectively.

Most significantly, students' coping strategies are anchored in the Quranic principle of tabayyun (rigorous, systematic inquiry and verification before accepting or disseminating information). Tabayyun transforms standard technical literacy from a purely pragmatic skill into a sacred moral duty [4], [24]. By instilling tabayyun, individuals are ethically motivated to pause, investigate metadata, corroborate sources,

and suppress impulsive sharing behaviors [20], [29]. This synthesis of technological literacy (itqān) and moral accountability (amanah) offers an alternative to secular digital literacy models, anchoring critical cognitive discernment in transcendent ethical values [29], [32].

Pedagogical and Policy Implications for Higher Education Institutions

The findings of this study have direct implications for higher education curricula and institutional policy:

Curricular Integration: Universities—particularly Islamic Higher Education Institutions (PTKIN)—must move beyond rudimentary computer literacy by establishing comprehensive digital ethics courses that integrate AI forensics, cognitive bias awareness, and Islamic communication ethics (tabayyun, shiddiq).

Institutional Verification Hubs: Higher education institutions should establish peer-led digital verification clinics and fact-checking student communities that actively monitor, analyze, and debunk viral synthetic disinformation circulating in academic networks [30], [31].

Psychosocial and Cyber-Safety Support: Given the heightened anxiety regarding identity theft and gendered visual exploitation, student affairs and counseling units must implement targeted psychological support and clear legal aid mechanisms (referencing UU ITE and personal data protection regulations) to protect students against non-consensual deepfake harassment [1], [23].

In sum, responding effectively to AI-generated visual disinformation requires a multi-layered framework that unites technological detection, psychological resilience, peer deliberation, and ethical conviction. Anchoring digital literacy within the normative values of shiddiq, amanah, and tabayyun provides a resilient paradigm to safeguard the epistemic integrity of the digital public sphere.

CONCLUSION

This study demonstrates that AI-generated photo manipulation introduces multifaceted challenges that disrupt visual epistemology, induce affective vulnerabilities, and demand principled moral responses among Muslim university students at UIN Siber Syekh Nurjati Cirebon. Empirical evidence reveals that students frequently encounter synthetic imagery spanning parasocial interactions with public figures, politically and religiously staged fabrications, and hyper-realistic portraits marked by subtle anatomical anomalies. Exposure to these hyper-realistic fabrications triggers acute psychological distress, progressing from initial cognitive ambiguity and disorientation to persistent digital anxiety regarding identity misuse and a generalized erosion of epistemic trust across digital platforms. Evaluated through the lens of Islamic moral philosophy, deceptive AI manipulation is categorically condemned as an infringement on truthfulness (shiddiq) and moral stewardship (amanah), while conditional tolerance is afforded to creative parodies provided they uphold transparency, consent, and non-harm. To mitigate these disinformation risks, students employ an integrated coping framework that combines technical source verification, communal peer deliberation, and a value-based digital literacy paradigm fundamentally anchored in the Quranic imperative of tabayyun.

Several limitations characterize this investigation, notably the focused empirical scope involving five undergraduate informants within a single Islamic higher education institution, which limits broader generalizability across diverse student demographics. Furthermore, the study prioritized subjective psychological and ethical perceptions rather than technical algorithmic detection benchmarks. Future research should implement mixed-method and longitudinal frameworks across diverse academic institutions to evaluate the empirical effectiveness of institutionalizing tabayyun-based digital literacy curricula, thereby fostering cognitive resilience and ethical integrity in the era of generative artificial intelligence.

ACKNOWLEDGMENTS

The authors express their sincere appreciation to the academic leadership of Universitas Islam Negeri Siber Syekh Nurjati Cirebon and to all undergraduate student informants who contributed their time and insights to this research endeavor.

REFERENCES

- [1] M. T. Darmawan, A. Junaidi, and A. Khaerudin, "Penegakan hukum terhadap penyalahgunaan deepfake pada pornografi anak di era Artificial Intelligence di Indonesia," *J. Serambi Huk.*, vol. 18, no. 1, pp. 42–54, 2025.
- [2] F. Nurrahmi and H. M. Syam, "Perilaku informasi mahasiswa dan hoaks di media sosial," *Communicatus J. Ilmu Komun.*, vol. 4, no. 2, pp. 129–146, 2020.
- [3] R. Laksmi and S. Andry, "Potensi kekerasan gender berbasis online pada penyalahgunaan teknologi kecerdasan buatan bagi perempuan di media," *J. Netnografi Komun.*, vol. 2, no. 2, pp. 91–104, 2024.

- [4] Y. Puspitasari, "Etika komunikasi tentang kejujuran dan keadilan dalam perspektif al-Qur'an," *Tabayyun J. Komun. Penyiaran Islam*, vol. 4, no. 1, pp. 17–26, 2023.
- [5] Y. Yusanto, "Ragam pendekatan penelitian kualitatif," *J. Sci. Commun.*, vol. 1, no. 1, pp. 1–13, 2020.
- [6] U. Sulung and M. Muspawi, "Memahami sumber data penelitian: primer, sekunder, dan tersier," *J. Edu Res.*, vol. 5, no. 3, pp. 110–116, 2024.
- [7] D. Susanto, Risnita, and M. S. Jailani, "Teknik pemeriksaan keabsahan data dalam penelitian ilmiah," *Qosim J. Pendidik. Sos. Hum.*, vol. 1, no. 1, pp. 53–61, 2023.
- [8] B. Arianto and Rani, *Teknik Wawancara dalam Metoda Penelitian Kualitatif*. Kalimantan: Borneo Novelty Publishing, 2024.
- [9] Qomaruddin and H. Sa'diyah, "Kajian teoritis tentang teknik analisis data dalam penelitian kualitatif: perspektif spradley, miles dan huberman," *J. Manag. Account. Adm.*, vol. 1, no. 2, pp. 77–84, 2024.
- [10] F. P. P. Mewdy and N. W. Ramalis, "Analisis kriminologi visual terhadap konten hoaks politik di Indonesia dengan 'teknologi ai' pada web turnbackhoax.id tahun 2023-2024," *J. Inf. Technol. Comput. Sci.*, vol. 8, no. 1, pp. 289–308, 2025.
- [11] B. S. Situmeang, I. Y. Silitonga, R. F. Silaen, and T. Siringo-ringo, "Pengaruh artificial intelligence terhadap tingkat kasus deep fake pada selebritis di twitter," *J. Device*, vol. 14, no. 1, pp. 80–91, 2024.
- [12] D. Maulana, N. S. Azzahrah, R. A. Syach, and S. Syauri, "Identifikasi konten visual buatan ai dengan resnet dan fine-grained feature extraction," *JITET J. Inform. dan Teknol. Elektro Terap.*, vol. 13, no. 3S1, pp. 1248–1256, 2025.
- [13] H. E. Nawawi, R. Rahmadewi, and L. Nurpulaela, "Pendeteksian wajah hasil dari manipulasi ai pada citra digital," *J. Ilm. Wahana Pendidik.*, vol. 2338, no. 2, pp. 13–34, 2022.
- [14] D. Fitri, A. N. Hidayah, A. Putri, N. H. Tanjung, and S. Izzati, "Deepfake dan krisis kepercayaan: analisis hukum terhadap penyebaran konten palsu di media sosial," *Jiic J. Intelek Insan Cendekia*, vol. 2, no. 6, pp. 11556–11568, 2025.
- [15] P. D. Lestari, D. S. Kahfi, and W. Kuncoro, "Persepsi mahasiswa terhadap pemberitaan di media online instagram pada akun harian bhirawa," *J. Penelit. Komun.*, vol. 3, no. 4, pp. 45–50, 2023.
- [16] Alfitriah and M. D. M. Alyatalathaf, "K-Pop idol standard dan praktik komodifikasi pekerja dalam video clip 'iu-celebrity,'" *Interaksi J. Ilmu Komun.*, vol. 12, no. 1, pp. 19–35, 2023.
- [17] A. Fahrudin and S. Arif, "Modeling manajemen informasi dan verifikasi berita menjelang pemilu dan pemilihan 2024: sebuah studi preliminary tentang upaya mencegah black campaign dan hoax," *Electoral Gov. J. Tata Kelola Pemilu Indones.*, vol. 4, no. 2, pp. 212–232, 2023.
- [18] Kuiliem et al., "Persepsi visual dan pengolahan informasi digital di kalangan generasi muda," *J. Psikol. Kognitif*, vol. 5, no. 2, pp. 112–124, 2023.
- [19] I. Rusdiana, "Kognisi pembaca berita palsu (fake news) di media online," *Kodifikasia*, vol. 2, no. 1, pp. 186–194, 2018.
- [20] A. Syukur, A. Fadilla, H. N. Kifli, I. T. A. Selti, M. D. A. N. Mubaraq, and N. Hidayati, "Peran literasi media dalam memerangi berita hoax pada media sosial," *Pediaqu J. Pendidik. Sos. dan Hum.*, vol. 4, no. 1, pp. 823–836, 2025.
- [21] K. R. Nana, F. Mas, and S. B. Gemian, "Etika media sosial dan implikasinya bagi individu dan masyarakat," *JIMMI J. Ilm. Mhs. Multidisiplin*, vol. 2, no. 3, pp. 288–299, 2025.
- [22] A. Agustini, M. R. Ramadhan, and A. Rahmawati, "Mengatasi kecemasan melalui spiritual emotional freedom technique (SEFT)," in *Proc. Conf. Psychol. Behav. Sci.*, 2024, vol. 3, no. 1, pp. 90–105.
- [23] Suhartati and Jennifer, *Hukum Perlindungan Anak dan Perempuan di Indonesia dalam Berbagai Isu dan Realitas*. Makassar: PT Nas Media Indonesia, 2025.
- [24] M. Z. Al-Ayyubi, "Etika bermedia sosial dalam menyikapi pemberitaan bohong (hoax) dalam perspektif hadis," *J. Studi Ilmu-Ilmu Al-Qur'an dan Hadis*, vol. 19, no. 2, pp. 148–166, 2019.
- [25] D. Daryanto and F. Ernawati, "Integrasi moral dan etika dalam Pendidikan Agama Islam," *Dinamika J. Kajian Pendidik. dan Keislaman*, vol. 9, no. 1, pp. 15–31, 2024.
- [26] A. Caniyago and Yulianti, "Manipulasi wajah menggunakan filter pada sosial media dalam perspektif islam," *AT-TAJDID J. Pendidik. dan Pemikir. Islam*, vol. 7, no. 1, pp. 230–235, 2023.
- [27] A. Farhan and Cindy, "Etika dan perilaku masyarakat Indonesia dalam menggunakan media sosial," *J. Serina Sos. Humaniora*, vol. 1, no. 1, pp. 139–145, 2023.
- [28] A. Salsabila, W. Franciska, and M. Candra, "Memahami legalitas remix & parodi di sosial media mengkaji ambiguitas uu hak cipta terhadap konten berbasis kebebasan berkarya," *J. Ris. Ilm.*, vol. 1, no. 1, pp. 15–18, 2020.
- [29] A. Rahim and M. Indah, "Pentingnya pendidikan literasi digital di kalangan remaja," *Sabajaya J. Pengabd. Kpd. Masy.*, vol. 2, no. 2, pp. 51–56, 2024.
- [30] A. Marlia, M. I. Darmawan, I. Triana Sari, D. Rendrahadi, A. Nosa, A. Agustina, and P. W. Kusmara, "Peran guru BK Dan PAI dalam mengatasi problematika terhadap trend media sosial remaja SMP Shailendra Pelembang," *Significant J. Res. Multidiscip.*, vol. 2, no. 2, pp. 299–311, 2023.
- [31] N. S. Husna, R. Octaviani, Z. Sahara, and U. Usiono, "Penerapan metode diskusi untuk meningkatkan kemampuan komunikasi peserta didik Kelas III di Mis Al-Wardah," *Khazanah Pendidik.*, vol. 18, no. 1, pp. 53–58, 2024.
- [32] F. A. Rahman, M. Rohmah, S. Rustiani, I. Y. Fatmawati, N. Alisda, and D. Sofianatul, "Pendidikan karakter dalam era digital: bagaimana teknologi mempengaruhi pembentukan moral dan etika," *J. Creat. Stud. Res.*, vol. 1, no. 6, pp. 294–304, 2023.