Islamic Ethical Analysis of Deepfakes and Religious Image Manipulation

Hijriatu Sakinah[1*], Arsy Shakila Putri[2], Ainur Rofiq[3]

[1] Universitas Sunan Gresik, Gresik, Indonesia

[2] Istanbul Ticaret University, Istanbul, Türkiye

[3] Universitas Pelita Bangsa, Bekasi, Indonesia

*Corresponding email: h.sakinah@lecturer.usg.ac.id

ARTICLE INFO

ABSTRACT

When images can replace truth, deepfakes test the limits of trust and authenticity in religion. This study aims to explain the relevance of Islamic ethical principles in assessing digital visual manipulation practices, filling a gap in knowledge in the study of Islamic-based technology ethics. This study uses a qualitative approach with a normative-analytical research design, combined with contextual analysis, an explore the meanings, values, and norms contained in Islamic texts and contemporary social practices. The results state that religious deepfakes create an ethical and knowledge crisis because they make it difficult to distinguish authentic fatwās from fake ones, undermine the foundations of Islamic authority in the form of sanad, integrity ('adālah), and clarity of sources, and violate ṣidq and amānah through tadlīs, ifk, and distortion of meaning that has widespread impacts (mafsadah 'āmmah) such as radicalisation, sectarian conflict, and a decline in trust in religious institutions. The novelty of this synthesis positions the phenomenon as post-truth religiosity. It proposes a framework of digital authenticity fiqh that goes hand in hand with digital literacy as farḍ kifāyah, so that practical policy responses need to balance the protection of religion with justice and the right to expression through transparent source verification, accountable platform governance, and tabayyun education.

INTRODUCTION

The development of artificial intelligence over the past decade has brought fundamental changes to how humans produce, represent, and interpret visual reality (Zhang, 2024). One of the most controversial manifestations of this development is deepfake technology, a machine learning-based image and video manipulation technique capable of producing visual and audiovisual representations that closely resemble reality. In Asia, a region with substantial cultural, religious, and religious tradition diversity, deepfakes are not only used for entertainment and creative innovation but are also beginning to enter symbolic and sacred spaces, including representations of religious figures, ritual practices, and religious images that have profound theological and social meanings (Krkić, 2025; Suying, 2022).

This phenomenon raises complex ethical issues, mainly when religious image manipulation is used for disinformation, political propaganda, commercialisation of religious symbols, or even insults to spiritual values. In the context of Asian Muslim societies, this issue becomes even more sensitive because Islam has a strict ethical framework regarding honesty (*ṣidq*), trustworthiness, prohibition of deception (*tadlīs*), and the honour of religious symbols that must not be toyed with.

The urgency of studying deepfakes and religious image manipulation has grown stronger as cases of misuse of this technology increase in various Asian countries. Reports from digital security agencies and media studies show that deepfake content featuring religious figures or Islamic symbols is often used to trigger sectarian conflict, reinforce hate speech, or manipulate public opinion ahead of certain political moments (Hayatullah, Ikhsan, et al., 2025). In several South Asian and Southeast Asian countries, fake videos have circulated showing clerics or religious leaders seemingly making controversial statements, when in fact the content is entirely digitally fabricated (Kaur et al., 2018; Stewart, 2014; Yilmaz et al., 2022).

This phenomenon is not only a technical or legal issue, but also touches on deeper ethical and moral dimensions, as it concerns truth, public trust, and the sanctity of religious representation. Therefore, an Islamic ethical approach is relevant and strategic in providing a normative framework that not only assesses technological right and wrong but also considers the spiritual, social, and cultural implications of using deepfakes in a religious context (Badawy, 2025).

Perspectives from computer science, cybersecurity, and law have so far dominated academic studies on deepfakes (Alanazi et al., 2025). Recent studies have discussed the technical aspects of deepfake detection, the development of algorithms to identify visual manipulation, and legal regulations to protect privacy and prevent the spread of disinformation (Karimullah, 2023). These studies have made significant contributions to understanding how deepfakes work and their impact on social and political stability. However, most of these studies are secular and normative in the framework of positive law, with little attention to the dimensions of religious ethics, especially Islamic ethics. In fact, in Asian regions with large Muslim populations, the public's response to the manipulation of religious images cannot be separated from ethical and theological views derived from Islamic teachings.

Several studies in the field of Islamic digital ethics have begun to emerge and discuss the challenges posed by artificial intelligence from the perspectives of *maqāṣid al-sharīʿah*, the principle of benefit (*maṣlaḥah*), and the prevention of harm (*mafsadah*) (Elmahjub, 2023; Habib, 2025; Harun et al., 2025). These studies generally highlight issues of data privacy, decision automation, and algorithmic justice. Meanwhile, studies that specifically discuss visual representation in Islam tend to focus on the classical debate regarding the prohibition or

permissibility of images and sculptures, as well as the development of visual arts in the history of Islamic civilisation (van Tilborgh, 2023). Research on the manipulation of modern religious images, particularly through deepfake technology, is still minimal and is often only mentioned briefly as part of the discourse on Islamic media or communication ethics.

In the Asian context, several sociological studies and media studies have examined the impact of digital disinformation on religious and identity conflicts (Haq & Kwok, 2024; Neo & Yin, 2023; Octavianne et al., 2024). These studies show that visual manipulation has greater persuasive power than text, making it easier to influence the audience's emotions and beliefs. However, these studies generally stop at analysing the social and political impacts, without delving deeply into how Islamic ethical principles can be used as a normative analytical tool to assess the practice of religious image manipulation (Elmahjub, 2023; Siregar et al., 2025). On the other hand, Islamic theological studies that discuss honesty and the prohibition of deception often fail to integrate the context of digital technology, which is very different from the social conditions of classical times (Adeoye, 2025; Afif et al., 2024). This situation creates a significant knowledge gap, especially in formulating ethical guidelines that are relevant and applicable to Muslim communities in the era of artificial intelligence.

Based on this literature review, the main research gap in this study lies in the lack of comprehensive studies that integrate Islamic ethical analysis with the phenomenon of deepfakes and religious image manipulation in the Asian context. Previous studies tend to be fragmented between technological, legal, social, and theological studies, without adequate synthesis. In addition, there have not been many studies that explicitly discuss how Islamic ethical principles, such as *maqāṣid al-sharī'ah*, *ḥifẓ al-dīn* (protection of religion), *ḥifẓ al-'aql* (protection of reason), and the value of honesty, can be applied to evaluate deepfake practices involving religious symbols and figures. Another shortcoming is the lack of a specific regional focus on Asia, even though the social, cultural, and political contexts in this region strongly shape how deepfakes are used and interpreted.

This study offers an element of novelty by presenting an integrated and contextual Islamic ethical analysis of the phenomenon of deepfakes and the manipulation of religious images in Asia. The novelty of this study lies in its attempt to link classical Islamic ethical principles to the challenges posed by cutting-edge technology, thereby developing a new conceptual framework for assessing the practice of digital visual manipulation. From a theoretical perspective, this study extends the application of Islamic ethics to the realm of artificial intelligence and synthetic media. The focus of this study is directed at a normative and contextual analysis of how Islamic ethics views the use and misuse of deepfake technology in the manipulation of religious images in Asia.

The scope of the study covers relevant Islamic ethical principles, the characteristics of deepfake technology, and the Asian social and cultural context that influences the production and reception of such manipulative content. This study does not focus on the technical development of deepfake detection, but rather on its ethical evaluation and moral implications. The research questions to be answered include how Islamic ethical principles assess deepfake practices involving religious images, the extent to which such manipulation contradicts the objectives of Islamic law, and how the Islamic ethical framework can provide normative guidance in addressing these challenges in the Asian context.

The main objective of this study is to formulate a comprehensive Islamic ethical analysis of the phenomenon of deepfakes and the manipulation of religious images in Asia, and to identify their normative implications for Muslim communities. In particular, this study aims to explain the relevance of Islamic ethical principles in assessing the practice of digital visual

manipulation, fill gaps in knowledge in the study of Islamic-based technology ethics, and offer a conceptual framework that can be used in academic discourse and public policy. The significance of this study lies in its contribution to enriching contemporary Islamic ethical studies and in providing a strong argumentative basis for efforts to protect the integrity of religious symbols in the era of artificial intelligence.

METHOD

This study uses a qualitative approach with a normative-analytical research design, combined with contextual analysis, to understand, interpret, and evaluate the phenomenon of deepfakes and religious image manipulation through an in-depth, reflective Islamic ethical framework. This approach was chosen because ethical issues cannot be reduced to quantitative variables alone and require an exploration of the meanings, values, and norms contained in Islamic texts and contemporary social practices. This study focuses on two main aspects, namely Islamic ethical principles relevant to honesty, the prohibition of fraud, and the protection of religious symbols, as well as the practice of deepfake-based religious image manipulation in the Asian context.

The main instrument of this study is the researchers themselves, who act as key instruments, supported by document analysis guidelines and an ethical categorisation matrix developed systematically based on the Islamic ethical theory framework and previous studies. The instrument development process was carried out through an in-depth literature review and conceptual testing to ensure the clarity of the analysis categories. Data validity is ensured through source and theory triangulation, persistent observation, and analysis-trace audits to maintain consistency of interpretation (Safari et al., 2026). The data collection procedure is carried out in stages, starting with the inventory of literature and normative documents, the identification and selection of cases of deepfake-based religious image manipulation in Asia, and ending with data coding and thematic mapping. Data analysis was conducted using a thematic-normative analysis technique that combines the interpretation of Islamic texts with a critical reading of the social and technological context, as this technique was deemed most appropriate for revealing the relationship between ethical norms and actual practices.

RESULTS AND DISCUSSION

The Battle Between *Hifẓ al-Dīn* (Protection of Religion) and Digital Technology Freedom

The phenomenon of deepfakes, a cutting-edge product of artificial intelligence, has radically shifted the relationship between humans and reality, truth, and symbolic authority. In a religious context, particularly Islam, this technology presents issues that go far beyond technical or positive legal matters, as it touches on sacred areas that have been strictly guarded through theological, ethical, and social norms. Deepfakes featuring clerics, holy figures, or religious symbols create a complex field of tension between the demands of *hifẓ al-dīn,* one of the main objectives of *maqāṣid al-sharī'ah,* and claims of freedom in digital technology innovation, which are often framed as a necessity of the times (Pertamawati et al., 2025; Sianipar et al., 2025). This tension cannot be understood in black and white terms, as it operates in an Asian social space characterised by plurality of schools of thought, diversity of Islamic scholarly traditions, and highly diverse political configurations, ranging from secular states, states with strict control over religion, to states that use Islam as a source of public legitimacy.

Within the framework of *maqāṣid al-sharī'ah, hifẓ al-dīn* occupies a fundamental position because it is related to the preservation of the integrity of teachings, the purity of faith, and the religious authority that serves as a reference for the *ummah* (Ibrahim et al., 2025). The

protection of religion is not only about safeguarding rituals and doctrines from theological deviations, but also about protecting the way religion is represented in the public sphere so that it does not mislead, demean, or manipulate the collective consciousness of the community. Religious deepfakes, which are capable of presenting visuals that appear to be authentic depictions of religious leaders or sacred symbols, have the potential to undermine these foundations because they create an illusion of truth that is difficult to distinguish from reality. In a society that historically places religious authorities as guardians of spiritual interpretation, visual falsification of these figures can shake public trust, create normative confusion, and open the door to the misuse of religion for non-religious interests.

Manipulation of religious images through deepfakes can be categorised under several classical moral concepts, with profound legal and ethical implications. *Tadlīs*, as a form of deception that involves concealing facts or misleadingly presenting information, is a highly relevant category. Deepfakes inherently operate through visual tampering, creating false representations so similar that audiences are convinced they are the real thing. In a religious context, this type of tadlīs not only harms individuals but also the faith community as a whole, as it undermines trust as the foundation of social and spiritual relationships. Visual deception involving religious figures can cause followers to adhere to teachings that were never delivered, or to reject the authority of religious scholars based on false statements attributed to them.

Apart from *tadlīs*, religious deepfakes can also be understood as a form of ifk, the falsification of facts or the deliberate spread of lies to mislead the public. In Islamic tradition, *ifk* carries very serious moral connotations because it involves the defamation of truth and honour, as reflected in Qur'anic stories about the dangers of slander and false news. When deepfakes are used to portray religious scholars or symbols in a context that contradicts Islamic teachings, or to support a particular political agenda, they are no longer merely technical manipulations, but become instruments of *ifk* that can trigger conflict, polarisation, and even religious-based violence. In Asia, where religious sentiment is often intertwined with ethnic and political identity, the impact of such visual ifk becomes even more destructive.

The manipulation of religious images through deepfakes can be categorised as *tahrīf al-maʿnā*, or distortion of spiritual meaning. Unlike *tahrīf lafẓī*, which concerns changes in the text, *tahrīf al-maʿnā* occurs when a spiritual message is distorted or presented in a context that alters its essential meaning. Deepfakes enable visual and symbolic distortion of meaning, for example, by placing religious symbols in narratives that contradict Islamic values, or by making holy figures appear to support views that are contrary to the teachings they have always fought for. This kind of distortion is dangerous because it works on a semiotic and emotional level, so that its impact is not always rationally recognised by the audience. However, the ethical issues surrounding religious deepfakes do not stop at this normative categorisation.

A more complex and problematic question is whether every form of religious deepfake should automatically be considered haram, or whether there is still room for permissibility in specific contexts, such as education, social criticism, or digital da'wah (Karimullah, Rahman, et al., 2023). In Islamic ethics, the moral judgement of an action depends not only on its outward form, but also on the intention (*niyyah*), purpose (*maqṣad*), and impact (*maʾālāt*) it causes. This principle opens up the possibility of ethical differentiation between destructive deepfakes and deepfakes intended for the common good, even though both use the same technology.

In the context of education, for example, deepfakes can be used transparently to simulate historical Islamic figures to help the younger generation understand Islamic history and values, provided that the manipulation is clearly explained and is not intended to deceive (Habib, 2025). From the perspective of *maqāṣid al-sharī'ah,* this kind of use can be considered within the framework of *maṣlaḥah mursalah,* as long as it does not violate the basic principles of honesty and does not demean the sanctity of the figures or symbols represented. However, the line between education and manipulation is often thin, especially when the audience lacks sufficient digital literacy to recognise that the content is fabricated.

In the realm of social criticism, deepfakes can also be seen as a satirical tool used to criticise the abuse of religious authority or deviant religious practices. The tradition of criticism and muḥāsabah in Islam shows that religion is not immune to critical reflection, provided it is conducted with propriety and with the aim of improvement. However, the use of deepfakes for satirical purposes faces a serious ethical dilemma, because visual satire of religious figures has the potential to be interpreted as insulting or abusive, especially in societies that view religious scholars with high respect. This is where there is a tension between freedom of expression and the obligation to maintain the honour of religion, which in *maqāṣid al-sharī'ah* tends to be resolved by prioritising the prevention of mafsadah over the achievement of speculative *maṣlaḥah*.

In the context of digital *da'wah,* some argue that cutting-edge visual technology, including deepfakes, can be used to reach a wider audience and convey Islamic messages creatively. This argument is often based on the principle that the means (*wasā'il*) can change with the times, while the objectives (*maqāṣid*) remain the same. However, the use of deepfakes in da'wah faces fundamental ethical challenges, because *da'wah* in Islam is inherently based on honesty, exemplary behaviour, and authenticity of message (Indriyani, 2023). If the means of da'wah contain elements of visual deception, then there is an internal contradiction between the method and the goal, which can ultimately damage the credibility of *da'wah* itself.

The plurality of schools of thought and political contexts in Asia further complicates the ethical assessment of religious deepfakes. In some countries with strict fiqh traditions and centralised religious authority, the tendency to completely ban the use of religious deepfakes is more dominant, as it is considered a direct threat to social stability and the rule of the *ulama* (Usmonov, 2025). Conversely, in countries or communities with a more contextual approach to fiqh and openness to contemporary *ijtihād,* there are efforts to distinguish between destructive uses and those that are educational or experimental in nature. This difference reflects internal dynamics within Islamic thought itself, where *maqāṣid al-sharī'ah* serves as a flexible framework that allows adaptation without compromising fundamental principles.

The battle between *hifẓ al-dīn* and digital technological freedom essentially reflects a broader tension between value stability and the dynamics of change. Digital technology is often understood as a space of freedom that is relatively autonomous from traditional norms, while Islamic ethics emphasises the attachment of technology to higher moral goals (Raquib et al., 2022). From the *maqāṣid* perspective, freedom of innovation is not an absolute value, but rather an instrumental value that must be subordinate to the goals of protecting religion, reason, life, property, and honour (Mustapha & Malkan, 2025). Therefore, the claim that technological innovation must be left unrestricted for the sake of creativity and progress is not entirely in line with Islamic ethics, especially when such innovation has the potential to undermine society's moral and spiritual order.

At the same time, an overly restrictive and defensive approach to Islamic ethics also risks alienating the ummah from the inevitable reality of technology. If hifẓ al-dīn is narrowly

understood as a total prohibition on all forms of new visual representation, then Islam risks losing its relevance in the digital space, which is increasingly dominant in shaping public consciousness. Therefore, the main challenge is not choosing between protecting religion and technological freedom, but formulating an ethical framework that can productively manage these tensions (Prianto et al., 2024). This framework requires the active involvement of scholars, academics, and technology practitioners to jointly develop normative guidelines that are sensitive to Islamic values while remaining realistic about technological developments.

Within this framework, *hifẓ al-dīn* cannot be reduced to a mere symbolic prohibition; it must be understood as an effort to preserve the meaning, authority, and trust at the core of religious life. Religious deepfakes are problematic not simply because they use advanced technology, but because they have the potential to sever the relationship between symbol and meaning, between authority and trust. When this relationship is damaged, religion risks being reduced to a visual commodity that can be manipulated according to market or power interests. In the Asian context, which is fraught with colonialism, identity conflicts, and the politicisation of religion, this risk becomes even more apparent and urgent.

The debate on religious deepfakes from an Islamic ethical perspective cannot be resolved with a single, simplistic answer. It requires a layered, contextual, and dialogical analysis that acknowledges the complexity of technological reality and the steadfastness of Islamic normative values. The categorisation of religious deepfakes as *tadlīs*, *ifk*, and *tahrīf al-ma'nā* provides a strong ethical basis for criticising destructive manipulation practices. Still, it does not preclude limited *ijtihād* for uses that are truly transparent, educational, and free of deception. This is where *maqāṣid al-sharī'ah* demonstrates its relevance as a dynamic ethical framework capable of navigating the battle between religious protection and digital technology freedom without sacrificing either.

Manipulation of Religious Images as an Instrument of Political Power and Radicalisation

The development of deepfake technology in the contemporary digital landscape has significantly changed the way power is produced, distributed, and legitimised, including in the religious and political spheres. The manipulation of religious imagery through artificial intelligence-based visual engineering is no longer merely a matter of media ethics or individual privacy violations. Still, it has become an effective instrument of power in shaping public perception, directing ideological affiliations, and mobilising collective action.

In the context of Muslim societies in South Asia and Southeast Asia, where religion plays a central role in the construction of social identity and political legitimacy, religious deepfakes pose a grave threat (Anoraga, 2024). This technology enables the creation of false legitimacy through authentic-looking visuals, such as fictitious *fatwās* purportedly issued by prominent clerics or synthetic calls for jihād attributed to religious figures with high symbolic authority. This phenomenon not only manipulates religious perceptions but also blurs the line between truth and falsehood in the public sphere, opening up vast opportunities for the exploitation of religion for the sake of power.

From the perspective of political theory and power studies, symbolic representation is one of the most potent sources of legitimacy. Religion, as a system of meaning that provides a moral and cosmological framework, is often used to justify political actions, both by the state and by non-state actors (Morrison, 2025). Religious deepfakes reinforce this potential by presenting visual evidence that is difficult for the general public to dispute. When a video shows a charismatic cleric seemingly issuing a particular *fatwā* or calling for violent action, its persuasive power far exceeds that of text or oral narratives. In societies with uneven levels of

digital literacy, such visuals are readily accepted as truth without critical verification. As a result, religious deepfakes become a highly effective tool for creating false legitimacy that can be used to strengthen power, justify violence, or undermine legitimate spiritual authority.

An Islamic ethical analysis of this practice of religious image manipulation reveals a serious violation of the fundamental moral principles that underpin religious and social life (Ahmad et al., 2023). The principle of *ṣidq*, which emphasises honesty as a core value in speech and action, is clearly violated by religious deepfakes. Honesty in Islam is understood not only as factual truth, but also as the correspondence between representation and reality. Deepfakes, by deliberately producing representations that do not correspond to reality, undermine this principle at a very fundamental level (Öhman, 2022). When lies are packaged in convincing visual forms and attached to sacred symbols, the impact becomes far more dangerous because it tarnishes the value of honesty that should be upheld in religious preaching and communication.

In addition to violating *ṣidq*, the manipulation of religious images through deepfakes is also a betrayal of the principle of *amānah*, namely the trust the community places in scholars and religious institutions as guardians of Islamic teachings. Scholars in the Islamic tradition are considered the heirs of the prophets, entrusted with conveying the truth and guiding the people. When their image is manipulated to convey messages they never uttered, that amanah is systematically destroyed. Furthermore, the community as the recipient of the message is also betrayed, because their trust is exploited for interests that contradict Islamic values. This betrayal of amanah not only affects specific individuals or groups but also has the potential to undermine the social trust structure that underpins community stability.

From the perspective of *maqāṣid al-sharī'ah*, the practice of religious deepfakes used for false legitimisation and radicalisation clearly causes *mafsadah 'āmmah*, or widespread social damage. This *mafsadah* is not limited to misinformation, but includes the potential for sectarian conflict, religiously-based violence, and social disintegration. In South Asia and Southeast Asia, history shows that the manipulation of religious symbols has often triggered horizontal conflicts between sectarian, ethnic, or political groups. Religious deepfakes accelerate and amplify the potential for such disputes because they spread rapidly through social media and are difficult to control once they go viral (Nandini & Jha, 2025). In this context, the damage caused is not only material, but also spiritual and psychological, as it undermines mutual trust and solidarity within Muslim communities.

The potential of religious deepfakes to trigger sectarian conflict becomes even more apparent when this technology is used to reinforce narratives of exclusivism and the delegitimisation of other groups. Fake videos showing certain clerics condemning other sects or calling for aggressive action can easily be used to mobilise hatred and violence. In societies with latent tensions between groups, deepfakes serve as catalysts that accelerate the escalation of conflict. Moreover, this type of visual manipulation is often designed with a deep understanding of mass psychology and the most sensitive religious symbols, making its impact far more destructive than conventional propaganda.

In addition to triggering sectarian conflict, religious deepfakes also play a significant role in reinforcing extremism. Radical groups utilise this technology to create synthetic calls to *jihād* that appear authentic, as if supported by legitimate religious authorities. In this way, they seek to give religious legitimacy to acts of violence and recruit new followers. In Islamic ethics, the use of religion to justify unlawful violence is a serious deviation from the teachings of *rahmatan lil 'ālamīn*. Deepfakes exacerbate this deviation by creating the illusion of a religious consensus that never actually existed (Matthews & Kidd, 2023). The impact is felt not only by the direct

victims of violence, but also by the global image of Islam, which is increasingly associated with extremism and terrorism (Ayu et al., 2025; Hayatullah, Rohman, et al., 2025; Hibbatulloh et al., 2025).

Another equally important damage is the collapse of public trust in religious institutions and authorities. When the public is repeatedly exposed to fake content showing religious scholars delivering controversial or contradictory messages, they become increasingly sceptical about the authenticity of religious information. This scepticism can develop into cynicism, where all religious authorities are considered untrustworthy. In the long term, this condition is dangerous for religious life, as believers lose credible moral references and become vulnerable to deviant interpretations of religion. From the perspective of *maqāṣid al-sharī'ah*, the collapse of trust in religious authorities is a serious threat to *ḥifẓ al-dīn*, because religion is not only preserved through texts, but also through the authority of its interpreters.

Faced with this threat, countries in South Asia and Southeast Asia often respond with censorship policies, content blocking, and criminalisation of the production and dissemination of religious deepfakes. From a national security perspective, these measures are understood as efforts to maintain social stability and prevent radicalisation. However, this state response raises complex ethical and political dilemmas. In Islamic ethics, the principle of justice (*'adl*) demands that power be exercised proportionally, not arbitrarily, and with respect for fundamental human rights, including the right to freedom of expression and opinion. Widespread censorship and criminalisation without accountability mechanisms have the potential to violate this principle, mainly if they are used to silence legitimate criticism or strengthen political control over religious discourse.

The question that then arises is whether the state has the right to be the sole interpreter of religious truth to maintain stability. In Islamic tradition, religious interpretation has never been monopolised by a single authority but has developed through a pluralistic, dialogical tradition of *ijtihād*. When the state takes over this role exclusively, there is a risk of reducing religion to an instrument of political power. In the context of religious deepfakes, a state that monopolises the determination of what is considered valid or heretical can easily slip into digital authoritarianism, where the pretext of stability is used to limit intellectual freedom and diversity of interpretation.

On the other hand, the absence of adequate regulation also opens the door for destructive actors to continue exploiting deepfake technology without restraint. Therefore, the dilemma between protecting society and respecting freedom cannot be resolved by taking an extreme approach on either side (Latif et al., 2025). From an Islamic ethical perspective, the balance between preventing *mafsadah* (harm) and protecting rights is a key principle. The state has a responsibility to avoid real social harm, but this responsibility must be carried out through mechanisms that are fair, transparent, and participatory. The involvement of scholars, academics, and civil society in formulating policies is crucial to ensure that regulations do not become tools of repression.

In this context, the Islamic ethical approach offers a more balanced normative framework than a security-only approach. The principles of *ṣidq* and *amānah* apply not only to individuals or groups that produce religious deepfakes, but also to the state as the holder of power. The state is required to be honest in its intentions, transparent in its policies, and trustworthy in safeguarding the public interest without abusing its authority. Similarly, the principle of *mafsadah 'āmmah* must be understood comprehensively, not only in the short term but also in terms of its long-term impact on freedom of thought, intellectual creativity, and religious dynamics.

The Crisis of Ulama Authority and Authenticity (*Aṣālah*) in the Age of Synthetic Reality

The development of artificial intelligence-based synthetic reality, particularly deepfake technology, has caused a profound epistemological shock to the structure of Islamic scientific authority, which for centuries has been built on the foundations of transmission, integrity, and clarity of sources. In classical Islamic tradition, the authority of scholars has never been based solely on personal appeal or visual popularity, but rather on scientific legitimacy obtained through a long process of learning, moral testing, and collective recognition by the scientific community. However, when the faces and voices of scholars can be perfectly replicated through synthetic technology, the relationship between knowledge, authority, and trust undergoes a fundamental shift.

The line between authentic *fatwās* and false statements blurs, not because of the weakness of Islamic scholarship, but because the epistemic infrastructure that supports this authority is eroded by visual domination and algorithmic logic. This crisis cannot be understood solely as a technical problem of media manipulation, but rather as a radical transformation in the way religious truth is produced and consumed. In the contemporary digital space, public trust is increasingly shifting from authority grounded in sanad and scientific competence to authority grounded in visibility and performativity.

Deepfakes accelerate this shift by presenting visual simulations that appear more convincing than reality itself (George & George, 2023). When a video features a cleric with a voice, expressions, and speech style identical to the original, the general audience tends to believe it without questioning its origin or context. This phenomenon creates a condition that can be called post-truth religiosity, a situation in which religious truth is no longer determined by scientific and moral validity, but by the power of emotional and instant visual representation.

From an Islamic ethical perspective, this crisis touches the heart of the concept of *aṣālah* or authenticity, which is a prerequisite for the legitimacy of knowledge and religious authority. Islam places scientific authority on three main pillars, namely *sanad* as a traceable chain of transmission, personal integrity (*'adālah*) as a moral guarantee of the bearer of knowledge, and clarity of sources that ensures that the teachings conveyed are based on valid references. *Sanad* functions not only as a historical verification mechanism but also as a symbol of the connection between generations of scholars, guaranteeing the continuity and accountability of knowledge.

Personal integrity ensures that those who hold knowledge do not abuse their authority for worldly or manipulative ends. Clarity of source ensures that religious messages are not detached from their context and the correct methodology of interpretation. Deepfake technology systematically negates these three foundations (Matli, 2024). *Sanad* becomes irrelevant when a message can be fabricated without going through any scientific transmission process. Personal integrity is reduced to a visual image that can be manipulated without the subject's involvement.

Clarity of source disappears because synthetic content is often separated from the context of its production and the intent of its creator. In these conditions, religious authority no longer attaches to process and morality, but rather to technology's ability to mimic reality. As a result, the faithful are faced with a flood of religious content that appears authentic but lacks an epistemic foundation that can be accounted for. This crisis of authority becomes even more complex when linked to changes in the structure of clerical rule in the digital age. Even before

the emergence of deepfakes, Islamic scholarly authority had been fragmented by social media, where anyone can deliver lectures and *fatwās* without strict verification mechanisms.

Deepfakes take this fragmentation to a more extreme level, as they enable the creation of false authorities that are completely detached from real human subjects (Gregory, 2022). In this context, scholars are not only competing with influencers or popular preachers, but also with digital representations of themselves that can be used to convey messages that contradict their views. This situation creates an ethical and epistemological paradox, in which religious authorities lose control over their own representation.

From the perspective of *maqāṣid al-sharī'ah,* this crisis of digital authenticity has direct implications for *hifẓ al-dīn* and *hifẓ al-'aql.* The protection of religion cannot be realised if the faithful are unable to distinguish between authentic teachings and synthetic manipulation. The security of reason is also threatened when critical thinking is replaced by dependence on misleading visual evidence (Williamson & Prybutok, 2024). In a condition of post-truth religiosity, reason is no longer used to assess the validity of sources, but rather to rationalise what is already believed based on visual impressions. This contradicts the Islamic scientific ethos, which holds that reason and verification are integral to faith.

The main controversy arising from this crisis is whether Islamic ethics needs to formulate a new fiqh on digital authenticity, which can be called *fiqh al-aṣālah al-raqamiyyah*, or whether the leading solution lies in increasing the community's digital literacy as a collective moral obligation. The idea of digital authenticity fiqh stems from the awareness that technological realities have given rise to new forms of falsification that are not fully accommodated by classical categories.

This type of fiqh has the potential to formulate normative criteria for what constitutes a valid religious representation in the digital space, the legal status of synthetic content, and the moral responsibility of developers and distributors of such technology. With this approach, Islamic ethics does not merely react to technology, but proactively builds a normative framework capable of guiding its use. However, the normative fiqh approach faces serious challenges in the context of rapid technological development and the diversity of Muslims' social contexts.

Overly rigid legal regulations risk lagging behind technological innovation or even being misused to restrict freedom of expression and *ijtihād*. Furthermore, digital authenticity fiqh without the support of critical capacity among the community has the potential to become ineffective normative formalities in the field. Therefore, an alternative view has emerged that emphasises digital literacy as *farḍ kifāyah*, a collective obligation incumbent on the Muslim community to maintain the integrity of religious life in the digital age.

Digital literacy in the perspective of Islamic ethics cannot be reduced to the technical ability to use devices or applications; rather, it encompasses the epistemic capacity to assess the authenticity of sources, understand the mechanisms of media manipulation, and relate information to the Islamic value framework. Making digital literacy a *farḍ kifāyah* means recognising that maintaining the authenticity of religious teachings in the era of synthetic reality is a collective responsibility, not just the task of scholars or the state. Within this framework, some members of society are required to master technological knowledge and digital epistemology to protect the community as a whole from misguidance and manipulation. However, the digital literacy approach is not without its limitations.

Not all members of society have the same access or capacity to develop critical skills regarding advanced technologies such as deepfakes (Murillo-Ligorred et al., 2023). Educational and information access disparities in many Muslim regions, particularly in Asia,

make the idea of digital literacy as a single solution unrealistic (Karimullah et al., 2022; Wahyudi et al., 2025). Furthermore, placing the entire moral burden on individuals and communities risks obscuring the structural responsibilities of technology actors and the state in regulating the production of synthetic reality (Karimullah, Sugitanata, et al., 2023).

The crisis of clerical authority and authenticity in the era of synthetic reality demands an integrative Islamic ethical approach. *Fiqh al-aṣālah al-raqamiyyah* and digital literacy as *farḍ kifāyah* should not be positioned as mutually exclusive alternatives, but rather as two complementary pillars. Fiqh of digital authenticity provides a normative and principled framework that affirms the ethical boundaries of technology use. In contrast, digital literacy ensures that these principles can be internalised and effectively applied by the *ummah*. In this combination, scholars continue to play a central role as guardians of values and interpreters of norms, while recognising that their authority must be rearticulated in language and media relevant to the digital reality.

At the same time, this crisis demands a rethinking of the very concept of authority in Islam. Scholarly authority can no longer rely solely on traditional symbols such as titles or visual appearances. Still, it must be reinforced through transparency, methodological openness, and active engagement in the digital space. *Ulama* who can explain the process of *ijtihād*, openly disclose their sources, and engage critically with technology will be better able to maintain their authority amid a flood of synthetic representations. In this way, authority is not maintained through control or censorship, but through the strengthening of trust based on authenticity and verifiable competence.

The phenomenon of post-truth religiosity, triggered by deepfakes, also reveals deeper theological challenges regarding the relationship among the senses, reason, and faith (Reczkowski et al., 2025). Historically, Islam has never placed sensory perception as an absolute source of truth, but has always demanded verification through reason and revelation (Lange, 2022). Synthetic reality challenges this hierarchy by creating simulations that deceive the senses, and if not accompanied by epistemic vigilance, weaken the function of reason. In this context, the crisis of clerical authority is actually part of a broader epistemological crisis, in which the faithful are required to revive the ethos of *tabayyun* and verification as religious practices relevant to the times.

The crisis of religious authority and authenticity in the era of synthetic reality is not merely an external threat posed by technology, but also an internal test for Islamic ethics and epistemology. Deepfakes shake the classical foundations of scholarly authority by negating sanad, integrity, and clarity of sources, while creating a dangerous post-truth religious condition for religious life. Responding to this challenge requires bold, reflective, and ethical ijtihad that is not trapped in nostalgia or technological euphoria. Through the development of responsive digital authenticity fiqh and the affirmation of digital literacy as a collective obligation, Islamic ethics has the potential not only to survive but also to make a significant normative contribution to global digital civilisation.

CONCLUSION

Deepfake technology and the manipulation of religious imagery pose a multidimensional ethical challenge that fundamentally undermines the basic principles of Islamic ethics, particularly within the framework of *maqāṣid al-sharīʿah*. Analysis shows that the practice of visual manipulation of religious scholars, symbols, and discourse not only violates the principles of *ṣidq* and *amānah* but also causes *mafsadah ʿāmmah*, which has broad implications for *hifẓ al-dīn, hifẓ al-ʿaql*, and social stability. By integrating classical Islamic ethical studies

with contemporary technological realities, this study enriches and refines previous studies that tended to be fragmented between technological, legal, and theological approaches, while presenting conceptual novelty through the articulation of the crisis of religious scholars' authority, post-truth religiosity, and the urgency of formulating digital authenticity fiqh as a contextual and reflective normative response.

Implicitly, these findings emphasise the need for an integrative Islamic ethical approach in responding to religious deepfakes, which does not rely solely on normative prohibitions or state control, but also demands strengthening digital literacy as a collective responsibility and collaborative dialogue among scholars, academics, policymakers, and technology practitioners. Although a normative qualitative approach limits this study and has not empirically examined the community's concrete responses to religious deepfake content, these limitations open the door to further interdisciplinary research based on comparative case studies across regions. In the future, Islamic ethics studies are expected to function not only as an evaluative framework but also as a source of practical inspiration and policy that can preserve the authenticity of religion without denying the dynamics of technology, so that Islam remains a critical, adaptive, and relevant moral tradition in facing the complexities of global digital reality.

ACKNOWLEDGEMENTS

REFERENCES

Adeoye, M. A. (2025). Hoaxes in Islamic Perspective: Qur'anic Solutions for Building an Anti-Hoax Society. *Al-Karim: International Journal of Quranic and Islamic Studies*, *3*(1), 35–58. https://doi.org/10.33367/al-karim.v3i1.6818

Afif, F., Khasri, M. R. K., & Khasri, M. I. U. A. (2024). Post-Truth And Qur'an: Hacking The Truth Decay Through The Qur'anic Epistemological-Ethical Vision. *Religia*, *27*(2), 162–195. https://doi.org/10.28918/religia.v27i2.8562

Ahmad, S., Qamar, A. J., Bhatti, M. A. A., & Bashir, U. (2023). Integrating Islamic Ethics with Modern Governance: A Comprehensive Framework for Accountability Across Religious, Social, and Economic Dimensions. *Al-Irfan*, *8*(15), 51–79. https://doi.org/10.58932/MULB0043

Alanazi, S., Asif, S., Caird-daley, A., & Moulitsas, I. (2025). Unmasking deepfakes: a multidisciplinary examination of social impacts and regulatory responses. *Human-Intelligent Systems Integration*, 1–23. https://doi.org/10.1007/s42454-025-00060-4

Anoraga, B. (2024). The Future of Artificial Intelligence in/and of Islam: A View from Muslim Southeast Asia. *Journal of Islamic and Muslim Studies*, *9*(1), 115–123. https://doi.org/10.2979/jims.00031

Ayu, R. F., Albab, M. U., Yakin, M. A., & Hibbatulloh, I. (2025). Legal Challenges in Countering Religion-Based Terrorism in Asia. *Jurnal Lentera Insani*, *1*(1), 1–15.

Badawy, W. (2025). Animating the sacred: the ethical and cultural implications of AI-powered awakening of pharaonic symbols using deepfake techniques. *AI and Ethics*, *5*(6), 6171–6184. https://doi.org/10.1007/s43681-025-00807-4

Elmahjub, E. (2023). Artificial intelligence (AI) in Islamic ethics: Towards pluralist ethical benchmarking for AI. *Philosophy & Technology*, *36*(4), 73. https://doi.org/10.1007/s13347-023-00668-x

George, A. S., & George, A. S. H. (2023). Deepfakes: the evolution of hyper realistic media manipulation. *Partners Universal Innovative Research Publication*, *1*(2), 58–74. https://doi.org/10.5281/zenodo.10148558

Gregory, S. (2022). Deepfakes, misinformation and disinformation and authenticity infrastructure responses: Impacts on frontline witnessing, distant witnessing, and civic journalism. *Journalism*, *23*(3), 708–729. https://doi.org/10.1177/14648849211060644

Habib, Z. (2025). Ethics of Artificial Intelligence in Maqāṣid Al-Sharīā's Perspective. *KARSA Journal of Social and Islamic Culture*, *33*(1), 105–134.

Haq, S. U., & Kwok, R. Y. (2024). Encountering "the Other" in Religious Social Media: A Cross-Cultural Analysis. *Social Media+ Society*, *10*(4), 20563051241303364. https://doi.org/10.1177/20563051241303363

Harun, M. S., Abdullah, L., Idris, M. A. H., Meerangani, K. A., & Ali, A. K. (2025). The Concept of Al-Thawābit and Al-Mutaghayyirāt in Technological Innovation According to Maqāṣid al-Sharīʿah. *Millah: Journal of Religious Studies*. https://doi.org/10.20885/millah.vol24.iss2.art1

Hayatullah, M., Ikhsan, M., Ramadhany, S. S. M., Rohman, I. N., & Hidayat, T. (2025). Media Discourse Analysis on Religious Tolerance Issues in Indonesia. *Jurnal Pelita Raya*, *1*(1), 61–75. https://doi.org/10.65586/jpr.v1i1.8

Hayatullah, M., Rohman, M. A., & Gonzales, N. (2025). Analysis of Shiite Political Thought in Iran and Its Influence in Indonesia. *Jurnal Lentera Insani*, *1*(1), 16–31.

Hibbatulloh, I., Najhan, A. S., Dzulfikar, M. L., & Mabrur, R. (2025). The Secularisation of Islamic Criminal Law and Its Implications for the Protection of Human Rights in Indonesia. *Insani: Jurnal Pranata Sosial Hukum Islam*, *1*(1), 17–31. https://doi.org/10.65586/insani.v1i1.7

Ibrahim, Z. S., Karimullah, S. S., Assaad, A. I., Septiani, R., & Okur, H. (2025). Integration of Maqaṣid al-Shari'ah in the Criminal Law Reform to Achieve Justice and Human Dignity. *Jurnal Hukum Islam*, *23*(1), 105–144.

Indriyani, P. I. (2023). Transformation of Islamic Religious Practices in the Digital Era: Opportunities and Challenges for Contemporary Da'wah. *Jurnal Dakwah*, *24*(2), 175–192. https://doi.org/10.14421/jd.2023.24205

Karimullah, S. S. (2023). The Application of Artificial Intelligence in Islamic Law Discovery. *Mutawasith: Jurnal Hukum Islam*, *6*(2), 109–121. https://doi.org/10.47971/mjhi.v6i2.748

Karimullah, S. S., Amin, M., & Munawwarah, U. D. (2022). Strengthening Spiritual Education Based on the Qur'ān in Building Character in the Digital Age. *Devotion Journal of Community Service*, *3*(14), 2785–2794. https://doi.org/10.36418/dev.v3i14.316

Karimullah, S. S., Rahman, R., & Elmurtadho, F. (2023). Reflections of Prophetic Values on Strengthening Da'wah in the Digitalization Era. *Mawaizh: Jurnal Dakwah Dan Pengembangan Sosial Kemanusiaan*, *14*(2), 134–154. https://doi.org/10.32923/maw.v14i2.3279

Karimullah, S. S., Sugitanata, A., & Elmurtadho, F. (2023). Juridical Analysis Of Public Information Disclosure In Government Systems In The Digital Era. *Constitution Journal*, *2*(2), 1–18. https://doi.org/10.35719/constitution.v2i2.64

Kaur, K., Nair, S., Kwok, Y., Kajimoto, M., Chua, Y. T., Labiste, M. D., Soon, C., Jo, H., & Lin, L. (2018). *Information Disorder in Asia: Overview of Misinformation Ecosystem in India, Indonesia, Japan, the Philippines, Singapore, South Korea, and Taiwan*. https://doi.org/10.2139/ssrn.3134581

Krkić, M. (2025). Cultural perspectives on AI usage and regulation in deepfake creation: How culture shapes AI practices. *International Communication of Chinese Culture*, 1–13. https://doi.org/10.1007/s40636-025-00330-5

Lange, C. (2022). Introduction: The sensory history of the Islamic world. In *The Senses and Society* (Vol. 17, Issue 1, pp. 1–7). Taylor & Francis. https://doi.org/10.1080/17458927.2021.2020603

Latif, F. A., Dahhur, M., & Ramadhany, S. S. M. (2025). Protection of Freedom of Religion in Islamic Law in Secular Turkey and Its Relevance for Indonesia. *Insani: Jurnal Pranata Sosial Hukum Islam*, *1*(1), 1–16. https://doi.org/10.65586/insani.v1i1.1

Matli, W. (2024). Extending the theory of information poverty to deepfake technology. *International Journal of Information Management Data Insights*, *4*(2), 100286. https://doi.org/10.1016/j.jjimei.2024.100286

Matthews, T., & Kidd, I. J. (2023). The ethics and epistemology of deepfakes. *The Routledge Handbook of Philosophy and Media Ethics*, 342–354. https://doi.org/10.4324/9781003134749-33

Morrison, T. (2025). Religeopolitics and the Affective Spatialities of Faith: Rethinking Religion in Geopolitical Analysis. *Geography Compass*, *19*(12), e70055. https://doi.org/10.1111/gec3.70055

Murillo-Ligorred, V., Ramos-Vallecillo, N., Covaleda, I., & Fayos, L. (2023). Knowledge, integration and scope of deepfakes in arts education: The development of critical thinking in postgraduate students in primary education and master's degree in secondary education. *Education Sciences*, *13*(11), 1073. https://doi.org/10.3390/educsci13111073

Mustapha, R., & Malkan, S. N. A. (2025). Maqasid Al-Shariah In The Ai Era: Balancing Innovation And Islamic Ethical Principles. *International Journal of Islamic Theology & Civilization (E-ISSN-3009-1551)*, *3*(3), 1–21. https://doi.org/10.5281/zenodo.15381828

Nandini, C. P., & Jha, A. (2025). Shaping Perceptions, Shaping Realities: Social Media Influencers and Human Rights Narratives in Crisis. In *The Asian Yearbook of Human Rights and Humanitarian Law* (pp. 61–106). Brill Nijhoff. https://doi.org/10.1163/9789004748736_004

Neo, R., & Yin, J. (2023). Of social discipline and control: The stratified impact of fake news and disinformation on minorities in Indonesia. *International Journal on Minority and Group Rights*, *31*(1), 110–132. https://doi.org/10.1163/15718115-bja10115

Octavianne, H., Suhariadi, F., Mudzakkir, M. F., Trianto, D., & Chamdan, U. (2024). Identity Politics and Polarization in Contemporary Muslim Countries: The Impact of Elections, Social Media, and Global Dynamics. *MILRev: Metro Islamic Law Review*, *3*(2), 263–286. https://doi.org/10.32332/milrev.v3i2.9909

Öhman, C. (2022). The identification game: deepfakes and the epistemic limits of identity. *Synthese*, *200*(4), 319. https://doi.org/10.1007/s11229-022-03798-5

Pertamawati, T., Rukiyanto, B. A., & Asifarani, E. (2025). Negotiating Religious Identity in State Schools Within the Flow of National Curriculum Policy. *Jurnal Pelita Raya*, *1*(2), 76–91. https://doi.org/10.65586/jpr.v1i2.19

Prianto, A. L., Karimullah, S. S., Idri, I., Ibrahim, Z. S., Nugraha, A. R., & Gönan, Y. (2024). Hijacking of State Power on Religious Freedom by Community Organizations in

Indonesia. *Jurnal Ilmiah Islam Futura*, 24(2), 348–367. https://doi.org/10.22373/jiif.v24i2.17916

Raquib, A., Channa, B., Zubair, T., & Qadir, J. (2022). Islamic virtue-based ethics for artificial intelligence. *Discover Artificial Intelligence*, 2(1), 11. https://doi.org/10.1007/s44163-022-00028-2

Reczkowski, R., Starzyński, S., & Łużyński, W. (2025). Religion in the Face of Cognitive Security Challenges. *Pastoral Psychology*, 1–25. https://doi.org/10.1007/s11089-025-01257-9

Safari, A., Ismael, D., Safari, M., & Freihaut, J. (2026). From comparison to integration: Building energy simulation tool variability and the case for intelligent retrofit workflows. *Energy Reports*, 15, 108880. https://doi.org/10.1016/j.egyr.2025.12.017

Sianipar, G., Wahyunisa, H., Ayu, R. F., Ulfa, M., & Fajri, M. A. (2025). Analysis of Deradicalisation Strategies Through Family Programmes in Indonesia. *Jurnal Pelita Raya*, 1(2), 106–121. https://doi.org/10.65586/jpr.v1i2.21

Siregar, M. H., Abdurrahman, Z., Kamal, M., & Zainal, Z. (2025). Digital fiqh and ethical governance: Negotiating islamic normativity and online narcissism in contemporary indonesia. *JURIS (Jurnal Ilmiah Syariah)*, 24(1), 181–191. https://doi.org/10.31958/juris.v24i1.13218

Stewart, E. R. (2014). Internet Evangelization and the False Dilemma of Innovation and Tradition in the Analysis of Digital Religious Practice. *Arc: The Journal of the School of Religious Studies*, 42, 111–124. https://doi.org/10.26443/arc.v42i.395

Suying, D. L. (2022). Deep fakes and disinformation in Asia. In *Deep Fakes* (pp. 23–49). Routledge. https://doi.org/10.4324/9781003173397-2

Usmonov, M. (2025). From Human Scholars to AI Fatwas: Media, Ethics, and the Limits of AI in Islamic Religious Communication. *Al-i'lam-Journal of Contemporary Islamic Communication and Media*, 5(1). https://doi.org/10.33102/jcicom.vol5no1.125

van Tilborgh, Y. (2023). Civilising Attempts in Art and Islam: Muslim Artistic Performance Facing Social Orders in the UK. In *British Muslims and Their Discourses* (pp. 141–182). Springer. https://doi.org/10.1007/978-3-031-45013-6_7

Wahyudi, Mahrus, Kurniyadi, Munawwarah, U. D., & Khoiruddin. (2025). Carl Rogers' Humanistic Approach in Character Education in Pesantren. *Jurnal Pelita Raya*, 1(1), 31–45. https://doi.org/10.65586/jpr.v1i1.10

Williamson, S. M., & Prybutok, V. (2024). The era of artificial intelligence deception: unraveling the complexities of false realities and emerging threats of misinformation. *Information*, 15(6), 299. https://doi.org/10.3390/info15060299

Yilmaz, I., Saleem, R. M. A., Pargoo, M., Shukri, S., Ismail, I., & Shakil, K. (2022). *Religious populism, cyberspace and digital authoritarianism in Asia: India, Indonesia, Malaysia, Pakistan, and Turkey*. European Center for Populism Studies. https://doi.org/10.55271/5jchdy

Zhang, A. (2024). Virtual Reality Image Creation in the Era of Artificial Intelligence. *International Conference on Human-Computer Interaction*, 118–126. https://doi.org/10.1007/978-3-031-61359-3_9