Proceedings
of International Conference on Religion, Social and Humanities
Volume 4, 2025 | ISSN 3089-3356
Faculty of Ushuluddin, Adab and Dakwah
Universitas Islam Negeri Mahmud Yunus Batusangkar
DOI: https://doi.org/10.31958/proceedingsoficresh.v4i.123

Islamic Digital Ethics in The Era of Deepfake: an Analysis of Technological Mediation

Zira Shafira
Universitas Islam Negeri Sunan Kalijaga Yogyakarta, Indonesia
Correspondence: zirashafira11@gmail.com

Abstract. The development of deepfake technology presents significant challenges to digital ethics in Islam. Its ability to generate human images and voices that closely resemble reality has blurred the boundaries between truth and falsehood. This raises serious ethical concerns regarding fundamental Islamic values such as ṣidq (honesty), amānah (responsibility), and ʿadl (justice). In the digital sphere, visual deception not only distorts information but also undermines social trust and human dignity. This article seeks to examine the phenomenon of deepfake through Peter-Paul Verbeek's theory of technological mediation ethics, which views technology not merely as a tool but as something that participates in shaping the ways humans think and act morally. Using a library research method, this analysis demonstrates that deepfake technology can serve as a useful medium for da'wah (religious propagation) and education when used with good intentions, yet it can also become a source of fitnah (slander) and fasād (corruption) when misused. Therefore, Islamic digital ethics must be directed not only toward assessing the legality of human actions but also toward cultivating awareness that technology itself influences moral decision-making. This awareness of mediation is expected to guide Muslims to uphold the values of truth, justice, and responsibility of technology.

Keywords: Islamic Digital Ethics, Deepfake, Technological Mediation

1. Introduction

The emergence of deepfake technology marks a new phase in the crisis of digital representation in the era of artificial intelligence. With its ability to manipulate faces, voices, and body movements with near-perfect resemblance, deepfake does not merely alter how humans interact with digital imagery, but also shakes the very foundations of ethics and the epistemology of truth.(Chesney & Citron, 2019, pp. 1753–1820) In the Islamic context, this issue extends beyond the technical domain of visual manipulation; it touches the moral values underlying Islamic teachings such as *sidq* (truthfulness), *amanah* (trustworthiness), and *ʿadl* (justice). As the boundary between the true and the false becomes increasingly blurred, the *ummah* is confronted with a new challenge: how to maintain moral integrity within a digital space that is fluid and technologically mediated.

Islamic ethical discourse has long focused primarily on the fiqh-based legal aspects of technology—such as whether certain digital practices are permissible or prohibited—but has seldom examined how technology itself shapes human moral reasoning and behavior.(Al-Qaradawi, Kairo) This is where Peter-Paul Verbeek's theory of technological mediation becomes highly relevant. Verbeek views technology not as a neutral tool, but as an active participant in shaping human moral experience.(Verbeek, 2011, p. 21) In other words, moral action can no longer be understood as purely human choice, but as a relational outcome between human beings and technological artifacts. This perspective opens a new pathway for developing an Islamic digital ethics that does not only evaluate human actions, but also acknowledges technological mediation as part of moral consciousness formation.

Within the context of deepfake, this mediational awareness is crucial to move beyond the simplistic dichotomy between the "good user" and the "bad user." Deepfake can serve as a means for *da'wah*, education, or creative expression when guided by good intentions and the principle of *maslahah* (public benefit). Yet it can equally become an instrument of *fitnah* (deception) and *fasād* (corruption) when misused to deceive, defame, or manipulate truth. Therefore, Islamic digital ethics in the age of deepfake must shift from a merely normative-legal approach toward a reflective-mediatoral one—an ethical awareness that recognizes the active involvement of technology in moral agency.[5] In doing so, Muslims can uphold the principles of truth and responsibility amid the increasingly complex and ambivalent flow of technological progress.

2. Method

This article employs a library research method with a conceptual-normative approach, focusing on the analysis of ethical ideas and principles found in both philosophical and Islamic sources. This approach is chosen because the study does not aim to collect empirical data, but rather to examine the conceptual relationship between Peter-Paul Verbeek's theory of technological mediation and the core Islamic ethical values of *sidq* (truthfulness), *amanah* (trustworthiness), and *ʿadl* (justice). The primary sources include Verbeek's Moralizing Technology (2011) as the main theoretical foundation, as well as classical and contemporary Islamic ethical texts such as the Qur'an, hadith, and the works of Yusuf al-Qaradawi. The analysis is conducted through a descriptive-reflective method, which interprets key concepts and seeks points of convergence between technological morality and human moral responsibility in Islam. This approach allows for a synthesis between theological and philosophical dimensions, thereby constructing an Islamic digital ethics framework that not only evaluates human actions but also understands how technology mediates moral consciousness in the digital era.

3. Results and Discussion

3.1. Deepfake and the Crisis of Digital Representation

The phenomenon of deepfake represents the culmination of a major transformation in how humans produce and interpret digital representation. Through generative artificial intelligence, deepfake technology can fabricate human images and voices with near-perfect realism.(Chesney & Citron, 2019, pp. 1753–1760) This capability blurs the boundary between reality and representation, such that visual truth—once regarded as the highest form of empirical evidence—has become an entity that can itself be engineered.(Floridi, 2020, pp. 1–7) As Jean Baudrillard describes in his theory of

hyperreality, the real and the simulated have now exchanged places.(Baudrillanrd, 1994, p. 27) In this sense, deepfake is not merely a product of technological advancement but a mirror of the epistemological and moral crisis of digital culture: a moment when images no longer refer to reality but instead create their own.

From the standpoint of the philosophy of technology, the issue of deepfake cannot be reduced to a simple moral dichotomy between "good users" and "bad users." Peter-Paul Verbeek, through his technological mediation theory, argues that technology is never neutral.(Verbeek, 2011, p. 5) It mediates human relations with the world—shaping perception, interpretation, and even moral action.(Verbeek, 2005, p. 9) In other words, technology possesses a form of agency that co-constructs how humans "see" and "act" ethically. In the context of deepfake, technology functions not only as an instrument of deception but also as an "actor" that reconstructs the very boundaries between truth and falsity. It introduces a new moral experience that challenges the conventional ethical framework based solely on human intention or will.

According to Verbeek, the human–technology relationship operates through hermeneutic and alterity mediations—that is, technology serves both as a medium for understanding the world and as an "other" that interacts with human beings.(Ihde, 1990, p. 11) Deepfake as a digital artifact functions across both layers of mediation. On one level, it serves a hermeneutic role, mediating perception: humans come to understand reality through the images it produces. On another level, it assumes an alterity role, acting as an "other" capable of influencing, seducing, or deceiving its user. This relationship generates a new ethical condition in which humans must question not only "what is true," but "how truth is mediated." Thus, deepfake exemplifies the moral shift from intention ethics toward relation ethics, an ethical mode rooted in human–technology interaction.

The representational crisis triggered by deepfake also signals a transformation within digital epistemology. In a world mediated by algorithms, truth is no longer determined by indexicality—the direct link between sign and reality—but by credibility: how convincing something appears to be.(Wardle & Derakhshan, 2017, p. 35) This shift results in a "crisis of trust," where society loses its moral foundation for distinguishing between authentic and fabricated information.(Vaccari & Chadwick, 2020, pp. 1–19) Philosophically, this condition demonstrates how technology has produced a new ontology of truth—truth that no longer exists as correspondence but as a technologically performative construct.

Within Verbeek's mediation framework, deepfake cannot be separated from the moral design of technology. Every technology embodies particular values and moral orientations realized through its design, function, and context of use. Hence, moral responsibility does not end with the user but extends to designers, algorithmic systems, and the broader social structures surrounding them.(Vallor, 2016, p. 34) Awareness of such mediation allows humanity to reinterpret morality not as something that exists "beyond" technology but as something that occurs through technology. Therefore, the representational crisis caused by deepfake opens a new space for ethical reflection: how humans can cultivate a resilient moral consciousness amid a world increasingly mediated by artificial artifacts.

3.2. Islamic Digital Ethics: Foundational Values and Principles

In confronting the representational crisis generated by deepfake technology, Islamic ethics offers a moral framework that is not only normative but also ontological—rooted in an understanding of the nature of humanity, truth, and responsibility. Islam does not separate knowledge from morality; *ʿilm* (knowledge) is inherently connected to *akhlaq* (ethics).(al-Ghazālī, 2005, p. 38) In the digital context, this implies that every act of producing or disseminating information is not merely a technical operation, but a moral action that involves one's relationship with God and with others. Consequently, the ethical problem of deepfake is not simply a matter of whether it constitutes "lying," but concerns how humans safeguard the trust of truth (*al-ṣidq*) amid increasingly complex technological mediation.

In classical Islamic epistemology, truth is defined as *muwāfaqat al-ḥukm li al-wāqiʿ*—the correspondence between judgment and reality.(al-Juwayni, 1980, p. 71) Yet in the hyperreal digital world, the boundaries between reality and representation have become blurred. Hence, Islam emphasizes *niyyah* (intention) and *ʿadl* (justice) as moral foundations. A deed is judged not only by its outcomes but by the awareness and moral responsibility of its agent. This principle resonates with the ethics of care in contemporary philosophy of technology, where morality is not grounded solely in formal rules but in empathetic relations and awareness of consequences.(Vallor, 2016, p. 22) In Islam, this relational dimension takes form in the concept of *taqwā*—the consciousness of God's presence in every act, including those within the unseen realm of digital space.

Islamic ethics also contains a hermeneutical dimension that parallels Verbeek's theory of technological mediation. Technology is not an alien entity outside the human being but part of the *amanah khalīfah fī al-arḍ*—the divine mandate of humanity as steward of the earth.(Nasr, 1990, p. 63) Ethical responsibility arises not only because technology may be misused, but because it is an extension of human intellect and will. In this regard, deepfake challenges not merely users' morality but their spiritual awareness of the boundary between *khalq* (creation) and *takhliq* (artificial fabrication). Islam does not view human creative capacity (*takhliq*) as forbidden per se, but as an ethical trial: whether humanity can keep technology within the orbit of *maʿrifatullāh* (knowledge of God), or instead turn it into a tool of deception and domination.(Rahman, 1982, p. 11)

The principle of *ḥifẓ al-ʿaql* (preservation of reason) within the *maqāṣid al-sharīʿah* is equally relevant for assessing deepfake phenomena. The objectives of Islamic law do not only safeguard biological life but also protect the epistemological integrity of the human mind—so that reason remains uncorrupted by falsehood and manipulation.(Kamali, 2008, p. 52) From this standpoint, the deceptive dissemination of deepfakes constitutes a violation of the *maqāṣid*, as it damages reason and distorts public trust (*fitnah*). However, the same technology, when used transparently for education, security, or artistic expression, can be aligned with *maṣlaḥah mursalah* (public interest), provided it does not compromise the principles of *ṣidq* (truthfulness) and *amānah* (responsibility).(Auda, 2008, p. 16)

Moreover, Islamic digital ethics must reinterpret classical moral concepts such as *ghibah* (backbiting) and *iftirāʾ* (false accusation) in algorithmic terms. Whereas these once referred to verbal acts, they now manifest through visual and data manipulation.

Deepfake pornography, for instance, is not merely a violation of privacy but a desecration of human dignity (*ʿird*), which in Islam possesses sacred value.(El Fadl, 2014, p. 57) Thus, any form of digital body exploitation without consent constitutes a moral transgression equivalent to the defilement of honor in the physical world.

Interestingly, the structure of Islamic ethics is dialogical rather than repressive. It opens space for *ijtihād*—ethical reasoning that interprets emerging realities in light of *maqāṣidī* (teleological) understanding. As al-Ghazālī reminds us, genuine ethics does not end at legal compliance but culminates in the cultivation of a *muttaqī* soul—a self that recognizes its inner responsibility behind outward actions.(al-Ghazâlî, 1964, p. 45) In the context of deepfake, this entails fostering a form of digital consciousness that goes beyond mere caution against misinformation, toward understanding how technology itself shapes our moral perception. Philosophically read, Islamic ethics teaches that moral responsibility is not merely about prohibition, but about becoming an authentic human being within an artificial world.

Therefore, Islamic digital ethics does not seek to reject technology but to situate it within a spiritual horizon that upholds truth and justice. It rejects the instrumental logic that reduces technology to a means of domination, replacing it with the logic of *amanah* (trust): that every algorithm, dataset, and digital representation must submit to the values of *taqwā* (God-consciousness) and *ḥikmah* (wisdom).(Bagir, 2017, pp. 89–107) In a world surrounded by simulation, Islam offers an ethics of presence—an inner awareness that is reflective, responsible, and spiritually grounded in every digital act. Thus, the role of Islamic ethics is not merely to supplement legal norms but to serve as a philosophical horizon that reaffirms humanity amidst the proliferation of artificial representation.

3.3. Integrating Verbeek's Technological Mediation and Islamic Ethics as a Moral Design of Deepfake

The convergence between Peter-Paul Verbeek's philosophy of technology and Islamic ethics opens a new horizon for rethinking digital morality. Through his theory of technological mediation, Verbeek argues that technology is never neutral; it actively shapes how humans perceive, evaluate, and act morally.(Verbeek, 2011, p. 23) In his view, technology becomes a co-shaper of morality, where ethical relations occur not only between humans but also between humans and artifacts.(Ihde & Verbeek, 2009, pp. 279–289) This insight disrupts classical ethical paradigms that position technology as a passive tool. Instead, technology takes on an active role in mediating moral perception and decision-making.(Brey, 2014, pp. 957–974)

Within the Islamic framework, this idea resonates deeply with the concepts of *taklīf* (moral responsibility) and *ʿaql* (reason). Humans are endowed with intellect and freedom of choice, yet that freedom operates within a framework of divine consciousness (*taqwā*). Thus, every human-technology interaction becomes a moral testing ground—whether humans employ technology as a means of *ʿibādah* (devotion) or as an instrument of *fasād* (corruption). Islam rejects a rigid dualism between humans and things; every created entity is viewed as an *āyah* (sign of God), containing potential value and meaning.(Nasr, 1990, p. 68) From this perspective, Verbeek's mediation theory may be seen as a Western philosophical attempt to rediscover the ethical dimension of human-made creations—a principle already internalized in Islam through the notions of *amānah* (trust) and *khalīfah fī al-arḍ* (stewardship on earth).(Rahman, 1982, p. 65)

Integrating these two traditions yields a more reflective approach to digital ethics. Verbeek introduces the concept of moral design—the integration of moral values into technological design itself.(Verbeek, 2015, pp. 26–31) In Islam, this aligns with the principles of *ḥusn al-khalq* (good design) and *iḥsān* (acting with excellence and spiritual awareness). Technology should not merely be efficient; it should be *muḥsinah*—a reflection of goodness and public benefit.(al-Ghazâlî, 1964, p. 893) From this perspective, the rise of deepfake technologies is not only a moral threat but also an opportunity to design systems that uphold truth, justice, and human dignity. For instance, deepfake detection algorithms and ethical watermarking can be interpreted as forms of digital *hisbah*—moral oversight embedded in technological systems.(Floridi, 2016, pp. 123–145)

Furthermore, Verbeek's idea of material hermeneutics—that technology functions as an interpretive medium through which humans understand the world—intersects with Islamic epistemology.(Ihde, 1993, p. 77) In Islam, meaning is not confined to the text (*āyāt qauliyyah*), but also found in the *āyāt kauniyyah* (signs within creation), including human technological innovation. Every digital invention thus becomes a new field for ethical interpretation. Technological mediation mirrors an ongoing *ijtihād*—a moral reasoning process in which humans reinterpret divine values amid novel contexts without losing their transcendent orientation.(El Fadl, 2014, p. 473)

Both Islamic ethics and Verbeek's philosophy reject purely rule-based morality. They emphasize contextual, reflective engagement, where moral insight arises through lived technological practice.(Vallor, 2016, p. 117) In Islamic thought, this corresponds to *fiqh al-wāqiʿ*—a jurisprudential understanding grounded in contemporary realities.(Kamali, 2008, p. 723) Integrating the two therefore enables an Islamic digital ethics model that is adaptive, participatory, and mediation-conscious. This model moves beyond simplistic halal-haram binaries and instead asks how technology can be designed to serve the *maqāṣid al-sharīʿah*—the higher objectives of Islamic law, such as the preservation of truth (*ḥifẓ al-ṣidq*), intellect (*ḥifẓ al-ʿaql*), and human dignity (*ḥifẓ al-ʿirḍ*).(Auda, 2008, p. 334) Practically, such integration requires collaboration among scholars, technologists, and designers. Ethics must not remain confined to *fatwā* discourse; it must take the form of design ethics—a creative process ensuring that every layer of digital systems embodies moral values.(Bagir, 2017, pp. 89–107) This approach is especially relevant for addressing the rapid spread of deepfakes, which often outpace moral or legal regulation. A maqāṣidī-based design paradigm can generate technologies that are not only algorithmically intelligent but also morally wise.(Asyraf, 2022, pp. 33–52)

Ultimately, Islamic digital ethics in the deepfake era underscores the importance of ethical presence—a reflective moral awareness amid simulated realities. Verbeek insists that humans and technologies are inseparable; what is needed is a responsible coexistence.(Verbeek, 2011, p. 152) Islam extends this awareness with a transcendental dimension: every digital act is inscribed in *ʿilm Allāh*—the divine knowledge. Hence, moral responsibility is not merely social but cosmic. When technology transcends the boundaries of the real, Islamic ethics does not seek to reject it, but to guide it—ensuring that progress remains within the orbit of *ḥikmah* (wisdom) and *taqwā* (God-consciousness).

4. Conclusion

The rise of deepfake technology reveals that technology is no longer a mere instrument but an active force that shapes how humans perceive and interpret reality. Within the framework of Islamic digital ethics, this phenomenon compels a re-examination of the relationship between morality, truth, and technology. Peter-Paul Verbeek's perspective on technological mediation underscores that moral action never occurs in a neutral space; it is always mediated by technological artifacts that influence ethical perception and decision-making. Thus, cultivating awareness of technological mediation becomes the first step for Muslims to preserve moral integrity in the digital realm. Islamic digital ethics cannot remain confined to the boundaries of prohibition and permissibility; it must foster a reflective awareness that every innovation carries moral consequences. The principles of *sidq* (truthfulness), *amānah* (responsibility), and *ʿadl* (justice) must be reinterpreted within digital contexts to remain relevant in facing visual manipulation and misinformation. The task of Islamic ethics in the age of deepfakes is not to reject technology, but to guide its use toward the realization of *maṣlaḥah*—the common good. Ultimately, moral awareness in the digital world is not merely about choosing between right and wrong, but about safeguarding humanity amid increasingly subtle fabrications. Islamic digital ethics calls upon humankind to remain honest, just, and accountable—so that, in a world saturated with artificial images, the human conscience remains the only representation that can never be faked.

References

Al-Ghazâlî, A. H. M. bin M. (1964). *Mīzān al-`Amal*. ed. Sulaimān Dunyā. Dār al-Ma`ārif.

Al-Ghazâlî, A. H. M. bin M. (2005). *Ihyā' `Ulūm al-Dīn: Vol. I*. Dār al-Ma`rifah.

Al-Juwayni. (1980). *Al-Burhān fī Uṣūl al-Fiqh*. Dār al-Anṣār.

Al-Qaradawi, Y. (Kairo). *Al-Ḥalāl wa al-Ḥarām fī al-Islām*. Maktabah Wahbah.

Asyraf, A. R. (2022). Islamic ethics and artificial intellegence: Toward responsible innovation. *Journal of Contemporary Islamic Studies*, *1*(8).

Auda, J. (2008). *Maqasid al-Shariah as philosophy of Islamic law*. IIIT.

Bagir, Z. A. (2017). Etika teknologi dalam perspektif Islam. *Jurnal Ulumul Quran*, *2*(26).

Baudrillanrd, J. (1994). *Simulacra and simulation*. University of Michigan Press.

Brey, P. (2014). Designing moral mediation. *Science and Engineering Ethics*, *3*(20).

Chesney, R., & Citron, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review*, *107*. https://doi.org/10.15779/Z38RV0D15J

El Fadl, K. A. (2014). *Reasoning with God: Reclaming Shari`ah in the modern age*. Rowman & Littlefield.

Floridi, L. (2016). Faultless responbility: On the nature and allocation of moral responbility for distributed moral actions. *Philosophy & Technology*, *1*(29).

Floridi, L. (2020). Artificial intellegence, deepfakes and a future of epistemic opacity. *Philosophy & Technology*, *1*(33). https://doi.org/10.1007/s13347-020-00419-2

Ihde, D. (1990). *Technology and the lifeworld: From garden to earth*. Indiana University Press.

Ihde, D. (1993). *Postphenomenology essays in the postmodern context*. Northwestern University Press.

Ihde, D., & Verbeek, P. P. (2009). Expanding hermeneutics: Visualism in science and technology. *Philosophy & Technology*, *3*(22).

Kamali, M. H. (2008). *Maqasid al-Shariah Made Simple*. Islamic Foundation.

Nasr, S. H. (1990). *Man and nature: The spiritual crisis of modern man*. Unwin Paperbacks.

Rahman, F. (1982). *Islam and modernity: Transformation of an intellectual tradition*. University of Chicago Press.

Vaccari, C., & Chadwick, A. (2020). Deepfake and disinformation: Exploring the impact on public trust in news. *Digital Journalism*, *2*(8). https://doi.org/10.1177/2056305120903408

Vallor, S. (2016). *Technology and the virtues: A philosophical guide to a future worth wanting*. Oxford University Express.

Verbeek, P. P. (2005). *What things do: Philosophical reflections on technology, agency, and design*. Pennsylvania State University Press.

Verbeek, P. P. (2011). *Moralizing technology: Understanding and designing the morality of things*. University of Chicago Press.

Verbeek, P. P. (2015). Beyond interaction: A short introduction and mediation theory. *Interaction*, *3*(22).

Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policymaking*. Council of Europe.