Sapientia Diversalis Journal of Human Interaction and Social Studies
Vol 1 No 2 May 2026, Hal 106-116
ISSN: 3125-1269 (Print) ISSN: 3125-0696 (Electronic)
Open Access: https://researchfrontiers.id/sapientiadiversalis

Deepfake AI and the Protection of Privacy Rights: An Islamic Perspective in Muslim-Majority Countries of Southeast Asia

Asyifa Namira[1*], Naylla Faradillah[2], Keisha Atiqah Delia[3], Bimo Mukti Alghyfari Akbar[4], Arjuna Ridlallah Bimaulidits Tsani[5], Abdul Rahman Zaky[6], Muhammad Haidar Ali[7]
[1-7] Universitas Brawijaya, Indonesia
email: muhaidaral@gmail.com[1]

Article Info :
Received:
22-04-2026
Revised:
07-05-2026
Accepted:
12-05-2026

Abstract

The emergence of deepfake artificial intelligence (AI) has intensified concerns regarding privacy violations, identity manipulation, and reputational harm in digital environments. These challenges are particularly significant in Muslim-majority countries of Southeast Asia, where privacy and personal honour are closely connected to Islamic ethical and legal values. This study examines the implications of deepfake technology for privacy rights from an Islamic perspective, focusing on Indonesia, Malaysia, and Brunei Darussalam. Using a systematic literature review and qualitative-normative analysis, the study synthesizes scholarly publications, legal documents, policy reports, and Islamic jurisprudential sources related to deepfake governance and privacy protection. The findings reveal that existing regulatory frameworks remain fragmented and insufficient to address deepfake-related harms effectively. Islamic principles, particularly hifz al-'ird, la darar wa la dirar, sidq, and amanah, provide a strong normative foundation for protecting privacy and human dignity. The study proposes a more integrated approach that combines contemporary privacy regulation with Islamic ethical principles to strengthen digital rights protection in Muslim-majority societies.

Keywords: *Depfake AI, Digital Ethics, Islamic Law, Privacy Rights, Southeast Asia.*

©2022 Authors.. This work is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License.
(https://creativecommons.org/licenses/by-nc/4.0/)

INTRODUCTION

Over the past decade, advances in generative artificial intelligence (AI) have given rise to a phenomenon now widely known as deepfake—a technology capable of manipulating a person's face, voice, and gestures so convincingly that the results are often indistinguishable from authentic recordings. The rapid evolution of Generative Adversarial Networks (GANs) has significantly enhanced the realism and accessibility of synthetic media production, transforming deepfakes from experimental tools into widely available digital technologies (De Souza et al., 2023). As deepfake capabilities continue to expand, concerns have shifted from technical innovation toward broader societal risks involving misinformation, identity manipulation, cybersecurity threats, and violations of personal privacy (Ali et al., 2022). Contemporary debates on AI governance increasingly emphasize that issues of fairness, explainability, accountability, and human rights must be addressed alongside technological advancement to ensure responsible deployment of generative systems (Zhao et al., 2023). These developments have become particularly significant in Muslim-majority countries of Southeast Asia, where accelerating digital transformation intersects with evolving legal frameworks and religiously informed understandings of human dignity and privacy.

A substantial body of scholarship has examined the technological capabilities and societal risks associated with deepfake systems, yet the literature increasingly suggests that privacy violations represent one of the most severe and enduring consequences of synthetic media proliferation. Early investigations focused primarily on the detection and classification of manipulated content, while more recent studies have shifted attention toward the broader implications of identity exploitation, reputational harm, and psychological victimization enabled by AI-generated fabrications (Deeptrace, 2020). Research on technology-mediated abuse demonstrates that digital platforms facilitate new forms of coercion and victimization by extending harmful conduct beyond physical environments into networked social spaces where content can be replicated and disseminated indefinitely (Bailey et al., 2024). Scholars examining image-based sexual abuse have further argued that deepfake-enabled exploitation magnifies existing vulnerabilities because fabricated content can create reputational

damage even when no underlying conduct ever occurred (Rigotti & McGlynn, 2022). Legal and human-rights-oriented analyses similarly contend that privacy, freedom of expression, and reputation are increasingly entangled within complex regulatory debates surrounding AI-generated media, making simplistic legal solutions inadequate for addressing emerging harms (Prasad & Jain, 2025). Contemporary discussions have consequently expanded beyond questions of technological feasibility and toward the normative foundations required to govern synthetic media in ways that preserve both innovation and human dignity.

Despite this growing scholarly attention, significant fragmentation persists in the way deepfake-related privacy concerns are conceptualized and addressed across disciplines. Existing regulatory studies frequently approach deepfakes through the lenses of cybersecurity, digital governance, or data protection without adequately considering how cultural and religious value systems shape understandings of privacy and personal dignity (Kishwar et al., 2025). Comparative legal scholarship highlights substantial disparities among national responses to AI-generated harms, reflecting divergent assumptions regarding autonomy, consent, and accountability in digital environments (Doing et al., 2025). Research focused on Muslim societies has identified broader socio-legal challenges associated with AI adoption, yet these analyses often remain detached from concrete discussions of privacy rights and synthetic media abuse (Sarwar & ul Abideen, 2025). At the same time, emerging literature on digital ethics recognizes the need for context-sensitive normative frameworks but frequently treats religious traditions as peripheral rather than constitutive sources of ethical reasoning (Rabiu et al., 2025). The result is a conceptual gap in which technological harms are extensively documented while the normative foundations for evaluating those harms within Muslim-majority contexts remain insufficiently theorized.

This gap becomes particularly consequential in Southeast Asia, where rapid digital transformation has proceeded alongside ambitious national strategies for technological development. Malaysia's AI roadmap reflects a strong commitment to accelerating digital innovation and integrating advanced technologies into national development agendas (MDEC, 2021). Similar modernization trajectories are evident across the region, including in Brunei Darussalam, where institutional efforts continue to address barriers to digital transformation while expanding participation in the digital economy (bin Haji Zaini, 2025). Yet the acceleration of technological adoption has not been matched by an equally comprehensive framework for addressing the ethical and legal implications of synthetic media manipulation. Evidence from Indonesia demonstrates that technology-facilitated gender-based violence remains a persistent concern, particularly for women whose images are circulated or manipulated without consent (Komnas Perempuan, 2022). The increasing availability of deepfake tools raises the prospect that privacy violations may evolve into more sophisticated forms of harm that challenge existing legal institutions, social norms, and mechanisms of victim protection.

Within Islamic intellectual traditions, the protection of privacy cannot be understood merely as an individual preference or statutory entitlement because it is closely connected to broader objectives concerning human dignity, social order, and the prevention of harm. Classical and contemporary scholarship on maqasid al-shariah identifies the preservation of honour, reputation, and personal integrity as essential dimensions of human welfare that legal and ethical systems are expected to safeguard (Ibn Asyur, 2006). Islamic legal thought further emphasizes that rights and responsibilities operate within a moral framework designed to minimize injury and maximize collective well-being, creating a potentially valuable perspective for evaluating emerging technological risks (Kamali, 2008). Contemporary Islamic ethical analyses increasingly argue that AI governance requires normative benchmarks capable of addressing issues that extend beyond technical compliance and into questions of moral legitimacy and social justice (Elmahjub, 2023). Similar arguments maintain that digital transformation in Muslim societies should be guided by ethical principles that remain responsive to technological innovation while preserving foundational religious values (Waseem & Rahim, 2025). Existing religious guidance concerning online interaction also underscores the importance of truthfulness, respect, and the avoidance of harmful digital conduct, principles that bear direct relevance to deepfake-related privacy violations (MUI, 2017). Contemporary Islamic legal discussions have likewise stressed the necessity of applying established jurisprudential principles to novel technological contexts in order to prevent harm and protect vulnerable individuals (Jum'ah, 2019).

Against this backdrop, the present study positions itself at the intersection of human–technology interaction, privacy studies, Islamic legal thought, and digital governance by addressing a critical gap

in the existing literature concerning the normative and regulatory implications of deepfake AI in Muslim-majority countries of Southeast Asia. Rather than treating privacy exclusively as a legal construct or exclusively as a moral concern, this research develops an integrated analytical framework that examines how Islamic conceptions of dignity, honour, and harm can enrich contemporary discussions on privacy protection in the age of synthetic media. More specifically, this article seeks to map the forms of privacy threats generated by deepfake technologies, examine the Islamic ethical and jurisprudential principles relevant to these threats, evaluate the preparedness of regulatory frameworks in Indonesia, Malaysia, and Brunei Darussalam, and identify unresolved challenges that require further scholarly attention. The study contributes theoretically by bridging the divide between AI governance scholarship and Islamic ethical reasoning, while its methodological contribution lies in constructing a comparative socio-legal analysis capable of capturing the interaction between technological development, regulatory responses, and religiously informed understandings of privacy rights in contemporary Muslim societies.

RESEARCH METHODS

This study adopts a non-empirical research design employing a Systematic Literature Review (SLR) integrated with a qualitative-normative approach. The SLR method was selected because it enables a transparent and rigorous synthesis of fragmented knowledge while facilitating the identification of emerging research trends and unresolved scholarly debates (Snyder, 2019). Data were derived exclusively from secondary sources, including peer-reviewed journal articles, academic books, policy reports, legal documents, fatwas, and regulatory frameworks relevant to deepfake technology, privacy rights, Islamic ethics, and digital governance in Southeast Asia. Literature searches were conducted through Scopus, Web of Science, Google Scholar, and the Islamic Research Foundation International (IRFI) databases using combinations of keywords such as “deepfake,” “AI privacy,” “Islamic ethics technology,” “digital privacy Muslim,” and “Southeast Asia data protection.” The selection process applied predefined inclusion and exclusion criteria covering publication year, language, source credibility, thematic relevance, and full-text accessibility. Following a multi-stage screening procedure, the final corpus consisted of scholarly and policy materials considered most relevant to examining the intersection between deepfake AI, privacy protection, and Islamic legal perspectives.

The review process followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework to enhance transparency, consistency, and replicability in source selection (Moher et al., 2009). Data analysis was conducted through two complementary stages. First, thematic content analysis was employed to identify recurring patterns, dominant narratives, and conceptual developments concerning deepfake-related privacy threats, regulatory responses, and ethical concerns within contemporary scholarship. Second, a comparative normative analysis was undertaken to examine the relationship between existing legal frameworks in Indonesia, Malaysia, and Brunei Darussalam and Islamic jurisprudential principles, particularly the concepts of maqasid al-shariah, hifz al-‘ird, and the legal maxim la darar wa la dirar. Analytical rigor was strengthened through systematic coding, iterative comparison across sources, and triangulation between academic literature, legal instruments, and Islamic normative texts. This approach ensured that interpretations were grounded in diverse yet complementary sources, thereby increasing the credibility and coherence of the study’s findings.

RESULTS AND DISCUSSION

The systematic literature review identified a collection of studies addressing the intersection of deepfake technology, privacy protection, artificial intelligence governance, and Islamic ethical perspectives. The selected literature reveals that scholarly attention has progressively shifted from technical discussions of synthetic media generation toward broader concerns involving privacy violations, reputational harm, regulatory inadequacies, and ethical governance. This shift reflects growing awareness that deepfake-related risks extend beyond cybersecurity and increasingly affect social relations, human dignity, and fundamental rights within digital environments.

Analysis of the reviewed sources indicates three dominant themes. The first concerns the emergence of deepfake as a threat to privacy and personal security, the second focuses on Islamic legal and ethical principles relevant to digital harms, and the third examines the adequacy of regulatory

responses in Muslim-majority countries. Although these themes have been explored independently, studies integrating all three dimensions within the Southeast Asian context remain limited. Table 1 summarizes the principal sources that form the analytical foundation of this study.

Table 1. Summary of Selected Literature

Author (Year)	Method	Focus	Key Findings
Al-Qaradawi (2020)	Normative Jurisprudential Analysis	Islamic legal principles and digital harm	The principles of *hifz al-'ird* and *la darar wa la dirar* provide a strong basis for preventing technological practices that damage dignity and reputation.
Deeptrace (2020)	Research Report	Deepfake landscape and online abuse	Most deepfake content circulating online is pornographic and disproportionately targets women.
MDEC (2021)	Policy Analysis	National AI governance	Ethical AI governance has become a strategic priority, although implementation challenges remain significant.
Komnas Perempuan (2022)	Institutional Report	Online gender-based violence	Digital violence has increased substantially, with image-based abuse becoming an important concern.
Ali et al. (2022)	Conceptual Analysis	Security and privacy threats	Deepfakes present serious risks to privacy, public trust, and national security.
Rigotti & McGlynn (2022)	Juridical Analysis	Image-based sexual abuse	Existing legal mechanisms remain inadequate for addressing technology-enabled exploitation.
De Souza et al. (2023)	Literature Review	GAN development and image generation	Advances in generative models have significantly improved the realism of synthetic content.
Elmahjub (2023)	Ethical-Normative Analysis	AI and Islamic ethics	Islamic ethical values can contribute to accountable and human-centered AI governance.
Bailey et al. (2024)	Qualitative Analysis	Social technologies and abuse	Digital platforms facilitate new forms of surveillance, coercion, and reputational harm.
Kishwar et al. (2025)	Comparative Regulatory Analysis	Global deepfake regulation	Regulatory frameworks continue to lag behind the rapid evolution of synthetic media technologies.

The literature summarized in Table 1 demonstrates a consistent concern regarding the implications of deepfake technologies for privacy protection and human dignity. While existing studies have extensively examined technological risks and regulatory shortcomings, comparatively little attention has been devoted to integrating Islamic legal principles into discussions of privacy governance in the age of artificial intelligence. This gap is particularly relevant in Muslim-majority countries of Southeast Asia, where legal frameworks, cultural values, and religious norms collectively influence public understandings of privacy rights and acceptable technological conduct. The following sections discuss these issues through three interconnected dimensions: privacy threats posed by deepfake

technology, Islamic normative responses, and regulatory challenges in Indonesia, Malaysia, and Brunei Darussalam.

Deepfake AI as an Emerging Threat to Privacy Rights in Muslim-Majority Southeast Asia

The literature reviewed in this study indicates that deepfake technology has transformed the landscape of privacy threats in contemporary digital societies. Initially developed as an application of generative artificial intelligence, deepfake systems are capable of producing highly realistic synthetic images, videos, and audio recordings that can imitate real individuals with increasing accuracy (De Souza et al., 2023). The sophistication of these technologies has expanded rapidly due to advances in machine learning architectures and the wider availability of computational resources. As a result, the misuse of synthetic media is no longer confined to isolated technological experiments but has emerged as a growing social and legal concern. Existing scholarship increasingly associates deepfake proliferation with threats to personal autonomy, identity protection, and informational integrity.

Current studies suggest that privacy violations resulting from deepfake technology differ from conventional forms of digital misconduct because the manipulated content often appears authentic to viewers. The ability to replicate an individual's facial features, voice patterns, and behavioural characteristics enables malicious actors to fabricate events that never occurred while preserving a convincing appearance of reality (Ali et al., 2022). Such manipulation undermines trust in digital information and complicates efforts to distinguish authentic content from fabricated material. The consequences extend beyond individual victims because widespread exposure to synthetic misinformation can weaken confidence in digital communication systems. Concerns regarding these developments have intensified as deepfake production tools become more accessible to ordinary users.

The reviewed literature further reveals that deepfake-related harms are multidimensional and affect privacy in several interconnected ways. Some forms of abuse involve the unauthorized use of personal images, while others involve identity impersonation, reputational attacks, or the creation of fabricated intimate content. Legal scholars increasingly argue that privacy should not be interpreted solely as control over personal data but also as protection against the unauthorized reconstruction of personal identity in digital spaces (Prasad & Jain, 2025). This perspective is particularly relevant in the context of artificial intelligence because synthetic media technologies blur the distinction between authentic and fabricated representations of individuals. Consequently, the concept of privacy is expanding to encompass protection against algorithmically generated forms of identity exploitation.

Table 2. Categories of Deepfake-Related Privacy Threats

Threat Category	Description	Privacy Implications
Non-Consensual Intimate Imagery (NCII)	Creation of synthetic sexual content without consent	Violation of bodily privacy and personal dignity
Identity Fraud	Use of AI-generated images or voices to impersonate individuals	Unauthorized exploitation of personal identity
Reputational Manipulation	Fabrication of misleading audiovisual content	Damage to honour, credibility, and social standing
Political Disinformation	Manipulated content involving public figures	Distortion of public trust and informational integrity

As illustrated in Table 2, non-consensual intimate imagery constitutes one of the most damaging forms of deepfake misuse. Evidence compiled by Deeptrace (2020) indicates that sexually explicit content accounts for the overwhelming majority of deepfake materials distributed online. The victims of such practices are predominantly women whose images are manipulated and circulated without permission. The resulting harm frequently extends beyond privacy invasion because victims may experience social stigma, psychological distress, and long-term reputational consequences. These outcomes demonstrate that deepfake abuse often combines technological manipulation with broader patterns of gender-based violence.

.

The significance of these harms becomes even more apparent within Muslim-majority societies, where personal honour and social reputation occupy an important position in both cultural and religious life. Reports concerning online gender-based violence in Indonesia indicate a continuing increase in technology-facilitated abuse targeting women and vulnerable groups (Komnas Perempuan, 2022). Similar concerns have been identified in broader studies examining the relationship between social technologies and digital victimization. Online platforms can amplify harmful content by enabling rapid dissemination across multiple networks while reducing opportunities for effective intervention (Bailey et al., 2024). In this context, privacy violations frequently generate consequences that extend beyond individual victims and affect families, communities, and social relationships.

The Southeast Asian context presents additional challenges because digital transformation is progressing at a pace that frequently exceeds regulatory adaptation. Indonesia alone recorded more than 215 million internet users in 2023, reflecting the extensive integration of digital technologies into everyday social and economic activities (APJII, 2023). The expansion of online participation creates valuable opportunities for communication and development, yet it simultaneously increases exposure to emerging forms of digital harm. Recent regulatory scholarship suggests that legal frameworks across many jurisdictions continue to struggle in responding effectively to synthetic media technologies due to their speed of evolution and technical complexity (Kishwar et al., 2025). The literature therefore portrays deepfake not merely as a technological innovation but as a structural challenge requiring coordinated responses from legal institutions, policymakers, technology developers, and civil society.

Islamic Legal and Ethical Foundations for Protecting Privacy Against Deepfake Abuse

The findings of this review indicate that Islamic legal and ethical traditions provide a comprehensive normative framework for assessing the challenges posed by deepfake technology. Unlike contemporary privacy theories that frequently emphasize individual autonomy and control over personal information, Islamic jurisprudence approaches privacy as an integral component of human dignity, moral responsibility, and social order. The protection of personal honour occupies a particularly significant position within Islamic legal thought because reputation and dignity are regarded as essential elements of human welfare (Ibn Asyur, 2006). From this perspective, the unauthorized manipulation of an individual's image, voice, or identity constitutes more than a technical infringement upon personal data. It represents a form of harm that directly affects values protected by the objectives of Sharia.

A central concept emerging from the reviewed literature is hifz al-'ird, or the protection of honour and reputation. Although classical formulations of maqasid al-shariah focused primarily on the preservation of religion, life, intellect, lineage, and property, later scholars expanded the framework to include honour as an independent objective deserving explicit protection (Ibn Asyur, 2006). Deepfake technologies challenge this objective because fabricated content can damage a person's reputation regardless of whether the underlying allegations are true. The consequences become particularly severe when manipulated content is widely disseminated through digital platforms. In such circumstances, reputational harm may persist even after the falsity of the content has been demonstrated. The literature therefore suggests that deepfake abuse directly conflicts with one of the fundamental normative objectives recognized within Islamic legal theory.

The reviewed studies also emphasize the relevance of Islamic legal maxims in evaluating emerging technological harms. Among the most frequently cited principles is la darar wa la dirar, which establishes that harm should neither be inflicted nor reciprocated in social relations. Contemporary Islamic scholars argue that this maxim remains fully applicable within digital environments because technological innovations do not alter the ethical obligation to prevent harm and protect human welfare (Al-Qaradawi, 2020). Deepfake misuse clearly falls within this framework because it frequently generates psychological, social, and reputational injuries without the consent of the affected individuals. The principle therefore provides a strong normative justification for both preventive measures and legal intervention. Its continued relevance demonstrates the adaptability of Islamic jurisprudence in responding to technological developments that were not envisioned within classical legal contexts.

.

Table 3. Islamic Principles Relevant to Deepfake Privacy Violations

Principle	Source	Relevance to Deepfake Abuse
Hifz al-'Ird	Maqasid al-Shariah	Protects honour, reputation, and personal dignity from manipulation and defamation
La Darar wa La Dirar	Legal Maxim	Prohibits actions that cause harm to others, including digital harms
Dar' al-Mafasid Muqaddam 'ala Jalb al-Masalih	Legal Maxim	Prioritizes the prevention of harm over the pursuit of potential benefits
Sidq (Truthfulness)	Islamic Ethics	Rejects deception, fabrication, and the dissemination of false information
Amanah (Trustworthiness)	Islamic Ethics	Requires responsible and ethical use of information and technology

The principles summarized in Table 3 collectively demonstrate that Islamic legal reasoning possesses substantial resources for addressing privacy threats associated with artificial intelligence. The ethical concern is not limited to the consequences of deepfake abuse but also encompasses the processes through which such content is created and disseminated. The values of sidq and amanah place particular emphasis on truthfulness and responsible conduct in communication. Deepfake technologies become ethically problematic when they facilitate deception, misrepresentation, or the exploitation of personal information for harmful purposes. These concerns reflect a broader commitment within Islamic ethics to preserving trust within social interactions and public life. The literature consequently suggests that the evaluation of AI technologies should extend beyond technical functionality to include moral accountability.

Recent scholarship on artificial intelligence and Islamic ethics has increasingly argued that technological governance requires normative benchmarks capable of addressing complex social consequences. Discussions of AI ethics frequently focus on fairness, accountability, transparency, and explainability, yet these concepts are often articulated within predominantly secular frameworks (Elmahjub, 2023). Islamic ethical perspectives contribute an additional dimension by emphasizing human dignity, moral responsibility, and the prevention of harm as foundational principles guiding technological development. Similar arguments have been advanced in studies examining digital transformation within Muslim societies, where technological innovation is expected to remain aligned with broader ethical commitments (Waseem & Rahim, 2025). Research on ethics in the digital age likewise highlights the importance of value-based approaches that can respond to emerging technological risks while preserving social well-being (Rabiu et al., 2025). These perspectives indicate that Islamic ethical traditions may offer valuable insights for contemporary debates on responsible AI governance.

The practical relevance of these principles is reflected in existing Islamic guidance concerning online conduct and digital communication. Regulatory and religious institutions have increasingly recognized that digital interactions can generate harms comparable to those occurring in physical environments. The Indonesian Council of Ulama has emphasized that the dissemination of false, defamatory, and reputation-damaging content through digital platforms violates Islamic ethical standards and should be avoided (MUI, 2017). Similar positions appear in contemporary Islamic legal opinions that apply established jurisprudential principles to emerging technological contexts (Jum‘ah, 2019). The findings of this review therefore suggest that Islamic law does not view deepfake abuse as a novel ethical category requiring entirely new principles. Rather, it interprets such practices through an established normative framework that prioritizes dignity, truthfulness, responsibility, and the prevention of harm in both physical and digital environments.

.

Regulatory Responses and Future Governance Challenges in Indonesia, Malaysia, and Brunei Darussalam

The literature reviewed in this study reveals that regulatory responses to deepfake-related privacy violations in Southeast Asia remain fragmented and largely reactive. Although governments increasingly recognize the risks posed by artificial intelligence and synthetic media, existing legal frameworks were generally developed before the emergence of deepfake technologies and therefore lack provisions specifically designed to address AI-generated manipulation. This regulatory mismatch creates uncertainty regarding liability, victim protection, and enforcement mechanisms. Comparative studies indicate that many jurisdictions continue to rely on general provisions concerning privacy, defamation, cybercrime, or harmful content when responding to deepfake-related cases (Doing et al., 2025). Such approaches provide a degree of legal protection but often fail to address the unique characteristics of synthetic media abuse.

Indonesia has made notable progress in strengthening privacy protection through the enactment of Law No. 27 of 2022 concerning Personal Data Protection. The legislation introduces important principles such as consent, purpose limitation, accountability, and individual control over personal information. These provisions are relevant because deepfake creation frequently involves the unauthorized collection and processing of personal images and biometric data. Nevertheless, existing analyses suggest that the law does not explicitly regulate the use of personal data for synthetic media generation, leaving significant interpretative gaps in enforcement (Doing et al., 2025). Questions concerning ownership of facial data, AI-generated identity replication, and algorithmic manipulation remain insufficiently addressed within the current regulatory framework.

Malaysia has similarly developed a legal structure for personal data protection through the Personal Data Protection Act (PDPA) 2010 while simultaneously promoting technological innovation through national AI initiatives. Policymakers have increasingly acknowledged the need to balance digital transformation with ethical governance and privacy protection (MDEC, 2021). Despite these efforts, the literature suggests that existing legislation remains primarily focused on conventional data processing activities rather than emerging forms of AI-enabled identity manipulation. The rapid development of synthetic media technologies has exposed limitations in regulatory provisions that were drafted before deepfake abuse became a significant public concern. Consequently, scholars have called for more adaptive legal approaches capable of responding to evolving technological risks while preserving innovation and economic development.

Table 4. Comparative Regulatory Frameworks on Deepfake and Privacy Protection

Country	Main Regulatory Framework	Strengths	Identified Gaps
Indonesia	Personal Data Protection Law 2022	Strong personal data protection principles and individual rights	No explicit provisions addressing deepfake generation and biometric manipulation
Malaysia	Personal Data Protection Act 2010 and National AI Roadmap	Established privacy framework and growing AI governance agenda	Limited regulation of synthetic media and AI-generated identity misuse
Brunei Darussalam	Personal Data Protection Order 2021 and Sharia-based legal framework	Integration of privacy protection with religious legal principles	Absence of comprehensive AI-specific legislation
Regional Context	ASEAN digital governance initiatives	Growing awareness of AI governance challenges	Lack of harmonized regional standards on deepfake regulation

.

The comparison presented in Table 5 highlights a common pattern across the three jurisdictions. Each country possesses legal instruments that can be applied indirectly to privacy violations arising from deepfake misuse, yet none has established a comprehensive framework specifically addressing synthetic media technologies. This situation reflects a broader global trend in which technological innovation advances more rapidly than legal adaptation (Kishwar et al., 2025). Existing regulations often focus on the consequences of harmful conduct rather than the technological mechanisms that facilitate such conduct. As a result, enforcement authorities may encounter difficulties when attempting to classify and prosecute deepfake-related offenses under conventional legal categories.

Brunei Darussalam occupies a distinctive position within this regulatory landscape because its legal system incorporates Islamic principles more directly than those of many neighboring countries. Existing regulations concerning personal data protection and digital governance provide an important foundation for addressing emerging technological challenges. At the same time, scholarship examining digital transformation in Brunei identifies institutional and regulatory obstacles that continue to affect the development of effective governance frameworks (bin Haji Zaini, 2025). The absence of dedicated AI legislation creates uncertainty regarding how synthetic media harms should be regulated within a system that places considerable emphasis on moral and religious values. This challenge highlights the need for legal approaches capable of integrating technological expertise with established normative principles.

The governance challenges associated with deepfake technology are not solely legal in nature but also involve broader questions concerning ethics, accountability, and public trust. Studies examining law and society in the era of artificial intelligence emphasize that emerging technologies frequently generate complex social consequences that extend beyond the reach of traditional regulatory instruments (Sarwar & ul Abideen, 2025). Privacy violations facilitated by deepfake systems often intersect with issues of misinformation, discrimination, reputational harm, and unequal power relations. Effective governance therefore requires coordination among legislators, technology developers, educational institutions, religious authorities, and civil society organizations. A purely punitive approach may be insufficient if it is not accompanied by preventive and educational strategies.

Contemporary discussions on AI governance increasingly stress the importance of transparency, fairness, and explainability as mechanisms for strengthening accountability in automated systems. These principles are particularly relevant because deepfake technologies often operate in ways that obscure the processes through which synthetic content is generated and disseminated (Zhao et al., 2023). Greater transparency may enhance public awareness while improving the ability of regulators and institutions to identify harmful practices. Similar concerns are reflected in broader debates regarding the relationship between technological innovation and fundamental rights, where privacy and reputation are increasingly recognized as central governance considerations (Prasad & Jain, 2025). Such perspectives suggest that regulatory responses should incorporate both technical safeguards and human-centered protections.

The findings of this review indicate that future governance frameworks in Muslim-majority Southeast Asian countries would benefit from a more integrated approach that combines legal regulation, ethical guidance, and regional cooperation. Existing studies consistently identify the need for adaptive policies capable of responding to rapidly evolving forms of synthetic media manipulation. Regulatory reform should not merely react to technological harms after they occur but should establish preventive mechanisms that reduce opportunities for abuse while protecting fundamental rights. Greater collaboration among Southeast Asian countries may also facilitate the development of shared standards concerning privacy protection and responsible AI governance. Such efforts would be particularly valuable in addressing cross-border digital harms that increasingly transcend national jurisdictions.

CONCLUSION

This study demonstrates that deepfake AI has emerged as a multidimensional challenge that extends far beyond the domain of technological innovation, generating profound implications for privacy rights, personal autonomy, and human dignity in Muslim-majority countries of Southeast Asia. The findings reveal that the threats posed by deepfake technologies are not limited to data misuse or digital deception but encompass broader forms of identity exploitation, reputational manipulation, and image-based abuse that can inflict long-lasting social and psychological harm. Within the Southeast Asian context, these risks are intensified by rapid digital transformation and the growing integration of

online technologies into everyday life, while existing legal frameworks continue to struggle to keep pace with the evolving capabilities of synthetic media. The study further finds that Islamic legal and ethical traditions provide a robust normative foundation for addressing such challenges through the principles of hifz al-'ird, la darar wa la dirar, truthfulness (sidq), and trustworthiness (amanah), all of which emphasize the protection of human dignity and the prevention of harm. These findings suggest that discussions of deepfake governance in Muslim societies cannot be confined to technical regulation alone but must also incorporate ethical and religious considerations that shape social understandings of privacy, honour, and responsibility.

The comparative analysis of Indonesia, Malaysia, and Brunei Darussalam indicates that although important legal instruments for privacy protection have been established, significant regulatory gaps remain regarding the governance of AI-generated synthetic media. None of the examined jurisdictions has yet developed a comprehensive framework specifically addressing deepfake-related harms, resulting in fragmented responses that often rely on legal provisions not originally designed for such technologies. Theoretically, this study contributes to the emerging literature by integrating deepfake governance, privacy rights, and Islamic normative perspectives within a single analytical framework, thereby extending contemporary debates on responsible AI governance beyond predominantly secular approaches. Methodologically, the study demonstrates the value of combining systematic literature review techniques with qualitative-normative analysis to examine complex socio-technical issues situated at the intersection of law, ethics, religion, and digital technology. Future research should expand this discussion through empirical investigations of deepfake impacts on Muslim communities, particularly women and other vulnerable groups, while policymakers should pursue more adaptive and harmonized regulatory strategies capable of balancing technological innovation, privacy protection, and ethical accountability in the rapidly evolving digital environment.

REFERENCES

Ali, A., Ghouri, K. F. K., Naseem, H., Soomro, T. R., Mansoor, W., & Momani, A. M. (2022, October). Battle of deep fakes: Artificial intelligence set to become a major threat to the individual and national security. In *2022 International Conference on Cyber Resilience (ICCR)* (pp. 1-5). IEEE. https://doi.org/10.1109/ICCR56254.2022.9995821

Al-Qaradawi, Y. (2020). Fiqh al-Wasatiyyah al-Islamiyyah wa al-Tajdid. Doha: Markaz al-Qaradawi.

APJII. (2023). Indonesian Internet User Survey Report 2023. Jakarta: Asosiasi Penyelenggara Jasa Internet Indonesia.

Bailey, L., Hulley, J., Gomersall, T., Kirkman, G., Gibbs, G., & Jones, A. D. (2024). The networking of abuse: Intimate partner violence and the use of social technologies. *Criminal Justice and Behavior*, *51*(2), 266-285. https://doi.org/10.1177/00938548231206827

bin Haji Zaini, M. I. (2025). Understanding The Obstacles in Digital Transformation for Islamic Social Finance: Insights from Brunei Darussalam. *iEco| Islamic Economics Journal*, *3*(1), 101-115. https://unissa.edu.bn/journal/index.php/ieco/article/view/1335

De Souza, V. L. T., Marques, B. A. D., Batagelo, H. C., & Gois, J. P. (2023). A review on generative adversarial networks for image generation. *Computers & Graphics*, *114*, 13-25. https://doi.org/10.1016/j.cag.2023.05.010

Deeptrace. (2020). The State of Deepfakes: Landscape, Threats, and Impact. Amsterdam: Deeptrace Labs.

Doing, M., Fitrian, Y., Alkadrie, S. M. R. R. M., Warriyodi, W., Fahlevi, M. R., & Aprieyanti, A. (2025, July). Legal Challenges in Combating Deepfake Abuse: A Comparative Study of AI Regulation in Privacy Protection and Digital Security. In *Proceeding International Conference Restructuring and Transforming Law* (Vol. 4, No. 1, pp. 633-645). https://proceedings.ums.ac.id/icrtlaw/article/view/5772

Elmahjub, E. (2023). Artificial Intelligence (AI) in Islamic Ethics: Towards Pluralist Ethical Benchmarking for AI: E. Elmahjub. *Philosophy & Technology*, *36*(4), 73. https://link.springer.com/article/10.1007/s13347-023-00668-x

Ibn Asyur, M. T. (2006). Maqasid al-Syariah al-Islamiyyah. Amman: Dar al-Nafa'is.

Jum'ah, A. (2019). Al-Fatawa al-Islamiyyah fi Masa'il Mu'asirah. Cairo: Dar al-Muqattam.

.

Kamali, M. H. (2008). Shari'ah Law: An Introduction. Oxford: Oneworld Publications.

Kishwar, S. D., Tripathi, A., Khatoon, S., Poddar, D., & Khurana, B. (2025). Regulating deep fakes and synthetic media: Privacy, policy and global regulatory challenges. *Journal of Data Protection & Privacy*, *8*(1), 78-96. https://doi.org/10.69554/HBGG8150

Komnas Perempuan. (2022). Annual Report on Violence Against Women 2022. Jakarta: Komnas Perempuan.

MDEC. (2021). Malaysia's National Artificial Intelligence Roadmap 2021–2025. Cyberjaya: Malaysia Digital Economy Corporation.

MUI. (2017). MUI Fatwa No. 24 of 2017 on Legal Guidelines for Social Media Interaction. Jakarta: Majelis Ulama Indonesia.

Prasad, N. L., & Jain, I. D. (2025). Technological Progress vs. Human Rights: Privacy, Free Expression, And Reputation in the Era of AI And Deepfakes. *Journal of Legal Research and Polity*, *2*(2), 89-102. https://doi.org/10.64322/JLRP.2025.2207

Rabiu, A. A., Merican, A. M. M. N., & Al Murshid, G. (2025). Ethics in the digital age: Exploring the ethical challenges of technology. *Journal of Information Systems and Digital Technologies*, *7*(1), 29-50. https://doi.org/10.31436/jisdt.v7i1.555

Rigotti, C., & McGlynn, C. (2022). Towards an EU criminal law on violence against women: The ambitions and limitations of the Commission's proposal to criminalise image-based sexual abuse. *New Journal of European Criminal Law*, *13*(4), 452-477. https://doi.org/10.1177/20322844221140713

Sarwar, M. W., & ul Abideen, Z. (2025). Law and Society in the Artificial Intelligence Era: Socio-Legal Challenges in Muslim Countries and Youth Populations. *Al-Kashaf*, *5*(01), 48-61. https://alkashaf.pk/index.php/Journal/article/view/220

Waseem, U., & Rahim, A. (2025). Islamic Ethical Perspectives on AI and Digital Transformation in the 21st Century. *Wah Academia Journal of Global Religions*, *1*(3), 31-54. https://doi.org/10.63954/4sz9d580

Zhao, Y., Wang, Y., & Derr, T. (2023, June). Fairness and explainability: Bridging the gap towards fair model explanations. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 37, No. 9, pp. 11363-11371). https://doi.org/10.1609/aaai.v37i9.26344