

ISLAMIC ETHICS AND THE CRISIS OF DIGITAL INTEGRITY: A CRITICAL ANALYSIS OF TRUTH, DECEPTION AND RESPONSIBILITY IN THE AGE OF GENERATIVE AI

¹Sri Andayani Mahdi Yusuf, ²Idariani, ³Miftakhul Falah, ⁴Nurfuadina

¹²³⁴Pascasarjana, Universitas Muhammadiyah Aceh, Indonesia

¹sri.andayani@unmuha.ac.id, ²idariani@unmuha.ac.id, ³mitakhul.falah@unmuha.ac.id, ⁴nurfuadina@unmuha.ac.id

*Corresponding Author: Sri Andayani Mahdi Yusuf (sri.andayani@unmuha.ac.id)

Received: 8 August 2026; Revised: 14 Sept 2026; Accepted: 20 Oct 2026; Published: 01 Nov 2026

Vol: 4, Issue 3 (2026)

Doi: 10.5281/zenodo.23252464

Abstract

Keyword:

Islamic ethics;
generative
artificial
intelligence;
digital
integrity;
deepfakes;
tabayyun; *sidq*;
amanah;
maqasid al-
shariah;
misinformatio
n; moral
responsibility



Generative artificial intelligence (AI) can now produce fluent text, realistic images and convincing audio at almost no cost, and it does so without any commitment to truth. This has created a crisis of digital integrity: fabricated content is hard to distinguish from authentic content, machine outputs present invented claims as fact, and it is unclear who should answer for the resulting harm. Mainstream AI ethics has responded with principles such as transparency and accountability, but these frameworks draw mainly on Western philosophical traditions and treat deception largely as a technical or regulatory problem. Islamic scholarship on AI has grown quickly, yet most of it addresses AI in general through *maqasid al-shariah*, and its treatment of misinformation relies almost entirely on the receiver's duty of *tabayyun* (verification). This concept paper critically analyses truth, deception and responsibility in generative AI through Islamic ethics. Using conceptual analysis and theory adaptation, it synthesises the empirical literature on synthetic media with Qur'anic, prophetic and juristic concepts. It argues that Islamic ethics offers an integrated account built on three pillars: *sidq* (truthfulness) as a duty owed by every speaker and publisher, the prohibition of *ghish*, *tadlis* and *kadhib* (deception, concealment and falsehood) as limits on design and deployment, and *amanah* and *mas'uliyah* (trust and answerability) as a model of responsibility that always returns to a human agent. The paper proposes a Digital Integrity Framework that distributes these duties across developers, deployers, users and regulators, and offers suggestions for policy, education and future research.



This is an open-access article under the CC BY-SA license.

Doi: 10.5281/zenodo.23252464

Introduction

Generative AI has changed the economics of falsehood. Large language models and image, audio and video generators allow anyone to produce persuasive content in seconds, in any language and in any person's likeness. The information environment was already fragile before these tools arrived: a study of about 126,000 news cascades on Twitter found that false stories travelled farther, faster and more widely than true ones, and that people rather than bots were mainly responsible for spreading them (Vosoughi et al., 2018). Wardle and Derakhshan (2017) described this condition as "information disorder", made up of misinformation, disinformation and malinformation. Generative AI adds scale, personalisation and realism to that disorder, and this paper refers to the result as a crisis of digital integrity: a weakening of the shared ability to trust that what is seen, heard and read online is what it claims to be.

The crisis has three connected dimensions. The first concerns truth. Language models generate text by predicting plausible sequences of words, not by checking claims against reality, so they regularly produce confident statements that are false (Bender et al., 2021; Ji et al., 2023). The second concerns deception. Synthetic faces are now rated as indistinguishable from real ones and even as more trustworthy (Nightingale & Farid, 2022), and readers cannot reliably tell AI-written text from human writing (Kreps et al., 2022; Spitale et al., 2023). The third concerns responsibility. When a system that no one fully controls produces a harmful falsehood, it becomes difficult to say whether the developer, the deploying company or the user should answer for it, a problem long described as the "responsibility gap" (Matthias, 2004; Santoni de Sio & Mecacci, 2021).

The global response has been to issue ethical principles. A review of 84 AI ethics guidelines found convergence around transparency, justice and fairness, non-maleficence, responsibility and privacy, but also found that the guidelines came overwhelmingly from economically developed countries, with other regions and traditions barely represented (Jobin et al., 2019). The AI4People framework and the UNESCO Recommendation follow a similar principled approach (Floridi et al., 2018; UNESCO, 2022). Critics point out that principles alone do not guarantee ethical practice, because AI development lacks the shared aims, professional duties and accountability mechanisms that make principles effective in fields such as medicine (Mittelstadt, 2019). Elmahjub (2023) adds that Western ethical theories dominate the debate and argues for pluralist benchmarking that includes Islamic ethics. For the world's Muslim communities, this is a practical concern, because rules that do not connect with their moral vocabulary are unlikely to shape everyday conduct.

Islamic scholarship has started to fill this space, but a clear gap remains. Existing studies mostly address AI in general, using *maqasid al-shariah* or virtue ethics to discuss privacy, bias, autonomy and human dignity (Kannike & Fahm, 2025; Mohadi & Tarshany, 2023; Raquib et al., 2022). Where misinformation is discussed, the dominant response is *tabayyun*, the Qur'anic duty to verify news before acting on it, which was developed for social media and places the burden on the receiver (Mohamad & Mat Isa, 2022). Work on deepfakes is recent and still at the level of general frameworks (Uriawan et al., 2025). Three questions are therefore left open. Can a tradition that locates lying in the intention of a morally accountable speaker deal with a machine that has no intention at all? What duties fall on those who design and deploy systems that generate falsehood, as distinct from those who receive it? And how is responsibility to be assigned when the harm passes through many hands?

The current debate makes these questions urgent. One side warns that generative AI will flood public life with cheap, convincing disinformation and allow the guilty to dismiss real evidence as fake, the so-called liar's dividend (Chesney & Citron, 2019). The other side argues that these fears are overstated, because the demand for misinformation is limited and its reach depends more on audiences and distribution than on production capacity (Simon et al., 2023). A further dispute concerns language: describing false AI output as "hallucination" suggests an occasional malfunction, whereas Hicks et al. (2024) argue that such systems are indifferent to truth by design. This paper takes a position within that debate. It argues that the deeper harm is the erosion of trust,

that Islamic ethics has well-developed resources for exactly this harm, and that those resources need to be extended from the receiver of information to its producers. The paper's objectives are to analyse truth, deception and responsibility in generative AI through Islamic ethical concepts, to identify where existing Islamic and secular frameworks fall short, and to propose an integrated framework with practical suggestions.

Literature Review

The first body of literature documents how convincing synthetic content has become. In experiments with AI-generated news articles, Kreps et al. (2022) found that readers judged the machine-written stories to be credible and were largely unable to separate them from human-written ones. Spitale et al. (2023) reported a double-edged result: GPT-3 produced accurate information that was easier to understand than human-written posts, but it also produced more compelling disinformation, and participants could not tell which posts came from the model. In the visual domain, Nightingale and Farid (2022) showed that AI-synthesised faces were indistinguishable from real faces and were rated as more trustworthy. Taken together, this evidence shows that human judgement is no longer a dependable filter, which matters for any ethical approach that relies mainly on the individual's capacity to verify.

The second body of literature debates what this capability actually does to society. Chesney and Citron (2019) argue that deepfakes threaten privacy, democratic discourse and national security, and they identify the liar's dividend, in which the mere existence of fakes lets wrongdoers deny authentic evidence. Experimental work gives a more specific picture. Vaccari and Chadwick (2020) found that a political deepfake was more likely to leave viewers uncertain than to deceive them outright, and that this uncertainty lowered trust in news on social media. Simon et al. (2023) take the sceptical position that fears about generative AI and misinformation are overblown, since consumption of misinformation is driven by demand and by existing beliefs more than by supply. The two sides agree on one point that is central to this paper: the main damage may not be that people believe false things, but that they stop believing true things.

The third body of literature concerns the relationship between language models and truth. Bender et al. (2021) describe such models as "stochastic parrots" that assemble linguistic forms according to statistical patterns without reference to meaning. Ji et al. (2023) survey the resulting phenomenon of hallucination, in which generated text is fluent but unfaithful to its source or to fact. The scale of the problem is measurable: in one study, 55% of the bibliographic citations generated by GPT-3.5 and 18% of those generated by GPT-4 were fabricated (Walters & Wilder, 2023), and universities have raised parallel concerns about academic integrity (Cotton et al., 2024). Hicks et al. (2024) argue that "hallucination" is the wrong word. Drawing on Frankfurt's (2005) distinction between the liar, who knows the truth and conceals it, and the bullshitter, who does not care whether a statement is true, they hold that these systems fall into the second category. The debate is unresolved, but it shows that false output is a structural feature of the technology, not a rare fault.

The fourth body of literature addresses responsibility and governance. Matthias (2004) argued two decades ago that learning systems create a responsibility gap, because their makers and operators cannot fully predict or control their behaviour. Santoni de Sio and Mecacci (2021) refine this into four separate gaps, covering culpability, moral accountability, public accountability and active responsibility, and they contend that the gaps can be reduced by designing systems for meaningful human control. On the governance side, Floridi et al. (2018) propose five principles for a good AI society, and Jobin et al. (2019) show that guidelines worldwide converge on a similar short list. Mittelstadt (2019) doubts that such lists can do the work expected of them, since AI developers do not share a common professional ethic or face the kind of sanctions that give principles force. The evidence therefore points to an enforcement and motivation problem: principles exist, but the inner commitment and the clear chain of answerability needed to apply them are weak.

The fifth body of literature is the emerging Islamic ethics of AI, which contains its own debate about method. Raquib et al. (2022) propose a virtue-based framework grounded in the *maqasid*, arguing that AI should be assessed by its effect on human moral formation and not only by its outputs. Elmahjub (2023) builds his account on *maslahah* (public good) and shows that Islamic jurisprudence can support both welfare-maximising and duty-based readings of ethical AI. Auda's (2008) systems approach to the *maqasid* supplies much of the theoretical ground for such work, since it treats the objectives of the law as open and purposive. Applied studies extend the discussion to work ethics (Ghaly, 2024), to privacy and manipulation (Mohadi & Tarshany, 2023) and to governance grounded in *tawhid*, *ihsan* and *'adl* (Kannike & Fahm, 2025). These contributions establish that Islamic ethics has something distinctive to say about AI. Their limitation is that they remain general: none offers a sustained analysis of generative AI's specific problem of manufactured falsehood.

The sixth body of literature is Islamic information ethics. Its foundation is the Qur'anic command to verify reports brought by an unreliable source (Qur'an 49:6), supported by the prohibition on pursuing what one has no knowledge of (Qur'an 17:36) and on mixing truth with falsehood (Qur'an 2:42). Kamali (1997) shows that the classical tradition treats speech as a moral act bounded by duties against slander, false accusation and the spreading of unverified reports. Mohamad and Mat Isa (2022) apply *tabayyun* to social media and conclude that authenticity must be established before information is accepted or shared, while Uriawan et al. (2025) extend the discussion to deepfakes and recommend governance based on *'adl*, *amanah* and *tabayyun*-centred digital literacy. Read against the empirical evidence above, this literature faces a serious challenge. If ordinary people cannot detect synthetic content even when they try, a framework that depends mainly on receiver-side verification is incomplete. The literature has not yet developed the producer-side duties of truthfulness and non-deception, nor a clear Islamic account of who bears responsibility for machine-generated falsehood. That is the gap this paper addresses

Methodology

This study is a concept paper. It does not collect primary data; it develops an argument and a framework from existing knowledge. Jaakkola (2020) identifies four designs for conceptual articles: theory synthesis, theory adaptation, typology and model. This paper combines the first two. It synthesises two literatures that have developed separately, the empirical and philosophical study of generative AI and Islamic ethical thought, and it adapts a problem defined in one domain (digital integrity) by viewing it through the lens of the other (Islamic ethics of speech, trade and trust).

The material consists of three groups of sources. The first is the primary Islamic sources: Qur'anic verses on verification, truthful speech and trust, and prophetic traditions on lying, deception and accountability, read with reference to established juristic concepts. The second is contemporary scholarship on Islamic ethics and AI published between 2019 and 2025. The third is empirical and philosophical research on synthetic media, language models and responsibility. Sources were selected purposively for their relevance to the three focal concepts of truth, deception and responsibility, with preference given to peer-reviewed journal articles, scholarly books and documents from international bodies.

The analysis proceeded in three steps:

1. **Problem mapping.** The technical and social literature was reviewed to specify what the crisis of digital integrity consists of and to separate it into its truth, deception and responsibility dimensions.
2. **Conceptual analysis.** Each dimension was matched with the relevant Islamic concepts (*sidq*, *tabayyun*, *kadhib*, *ghish*, *tadlis*, *amanah*, *mas'uliyah*) and examined for fit, tension and limits, including cases where the classical concept assumes a human speaker.
3. **Framework construction.** The results were integrated into a framework that assigns duties to each actor in the generative AI chain and evaluates them against the *maqasid al-shariah*.

The design has limitations. A concept paper yields arguments and propositions, not tested findings, so the proposed framework requires empirical validation. The selection of sources is purposive and not

exhaustive, and the reading of Islamic sources reflects a mainstream Sunni framing that other schools may weigh differently.

Discussion

Truth: from *tabayyun* to *sidq*

Islamic ethics treats truth as a duty on both sides of communication. *Tabayyun* (Qur'an 49:6) governs the receiver, but the speaker is bound first by *sidq*: believers are commanded to speak words that are upright and accurate (Qur'an 33:70) and warned that every utterance is recorded (Qur'an 50:18). The Prophet's statement that it is enough falsehood for a person to relate everything he hears (Sahih Muslim, Introduction) extends the duty to anyone who passes content on. This matters for the objection that a machine cannot lie because it has no intention. A widely held classical definition treats *kadhib* as a report that does not match reality, whether or not the speaker meant to mislead, with sin attaching only where there is knowledge or negligence. On that definition, a fabricated citation or invented quotation is false speech in the descriptive sense, and the moral question shifts to the humans who release or repeat it. The analysis by Hicks et al. (2024) that language models are indifferent to truth therefore has a direct Islamic counterpart: a system that produces claims without regard to their truth is doing what Qur'an 17:36 forbids people to do, and the person who publishes its output unchecked adopts that fault as their own.

Deception: *ghish*, *tadlis* and the question of benefit

Islamic commercial and social ethics contains a detailed vocabulary of deception. *Ghish* is fraud or adulteration, *tadlis* is concealing a defect, and *taghrir* is inducing someone to act through a false impression. The Prophet's ruling that whoever deceives is not of the community (Sahih Muslim) was given in a marketplace, over a pile of grain whose wet portion was hidden beneath the dry. The parallel with synthetic media is close. A deepfake of a real person is fabrication and, where it harms reputation, falls under the Qur'anic condemnation of slander in the incident of the *ifk* (Qur'an 24:11-19). Undisclosed AI-generated content presented as human work is a form of *tadlis*, because a material fact about the product is concealed. A chatbot designed to make users believe they are speaking with a person creates a false impression by design. At the same time, Islamic ethics does not prohibit the technology as such. Following the *maslahah* reasoning discussed by Elmahjub (2023), synthetic media used with disclosure for education, accessibility or creative work is permissible. The dividing line is concealment and harm, not artificiality, and the maxim of blocking the means (*sadd al-dhara'i'*) justifies restricting uses whose predictable result is deception.

Responsibility: *amanah* and the closing of the gap

The responsibility gap arises in secular ethics because responsibility is tied to control and foresight, which learning systems weaken (Matthias, 2004). Islamic ethics begins from a different premise. Moral obligation (*taklif*) applies only to beings with reason and legal capacity, and the trust (*amanah*) that humanity accepted (Qur'an 33:72) cannot be handed to a tool. An AI system is therefore never a bearer of responsibility, and no gap can open in principle: answerability always returns to human agents. The prophetic statement that each person is a shepherd answerable for their flock (al-Bukhari; Muslim) distributes that answerability by role. Two juristic tools help to apportion it. The distinction between the direct actor (*mubashir*) and the indirect cause (*mutasabbib*) separates the user who publishes a falsehood from the developer whose system made it possible, while holding both liable in proportion to knowledge and negligence. The maxim that liability accompanies gain (*al-ghurm bi'l-ghunm*) places a heavier duty on companies that profit from deployment. This supports the call for meaningful human control (Santoni de Sio & Mecacci, 2021) and adds what Mittelstadt (2019) found missing in principle-based ethics: an

inner motivation, since the agent is answerable to God as well as to regulators.

A Digital Integrity Framework

The three analyses combine into one framework. Its foundation is the *maqasid*: generative deception threatens the protection of intellect (*hifz al-'aql*) by corrupting the basis of knowledge, of honour (*hifz al-'ird*) through fabricated likenesses, of religion (*hifz al-din*) through false attribution of religious texts and rulings, and of wealth (*hifz al-mal*) through fraud. The framework assigns duties to each actor in the chain, which corrects the receiver-only emphasis identified in the literature.

Actor	Governing concept	Core duty	Application to generative AI
Developers	<i>Amanah</i> , <i>itqan</i> (trust, diligence)	Build systems that minimise falsehood and reveal their limits	Accuracy testing, source grounding, uncertainty signals, provenance marking
Deployers and platforms	<i>Al-ghurm bi'l-ghunm</i> , avoiding <i>tadlis</i>	Disclose AI involvement and bear liability in proportion to profit	Clear labelling of synthetic content, complaint and correction channels
Content creators and users	<i>Sidq</i> , avoiding <i>kadhib</i> and <i>ghish</i>	Verify before publishing and never present fabrication as fact	Checking AI output, declaring AI assistance, refusing to create harmful deepfakes
Receivers	<i>Tabayyun</i> , <i>husn al-zann</i>	Verify before believing or sharing, without sliding into blanket suspicion	Source checking, withholding judgement on unverified content
Authorities and scholars	<i>Hisbah</i> , <i>sadd al-dhara'i'</i>	Supervise the information market and block predictable harm	Regulation, fatwa guidance, redress for victims of fabricated content

The framework also answers the main finding of the empirical debate. If the central harm of deepfakes is uncertainty and loss of trust (Vaccari & Chadwick, 2020), then verification by receivers cannot restore integrity alone. Trust has to be rebuilt at the source, and the Islamic tradition has a precedent for this in the science of hadith, which protected religious knowledge through chains of transmission (*isnad*) and the assessment of narrators. Content provenance systems that record where a digital item came from and how it was altered are a modern application of the same idea.

Suggestions

1. Policy and regulation. Regulators in Muslim-majority countries should require disclosure of AI-generated content and recognise reputational and psychological harm from deepfakes as grounds for redress, drawing on *hisbah* as an institutional precedent and aligning with the UNESCO (2022) Recommendation.
2. Religious guidance. Fatwa bodies should issue specific guidance on generative AI, covering deepfakes, undisclosed AI authorship and the use of AI to produce religious content, in which fabricated Qur'anic or hadith attributions are a serious risk.
3. Education. Universities should move from prohibition to formation: teach students to verify AI

output, declare AI assistance and treat fabricated references as a breach of *amanah*, in line with wider concerns about academic integrity (Cotton et al., 2024).

4. Design and industry. Developers serving Muslim markets should adopt provenance marking, citation grounding and visible uncertainty signals, and should involve scholars of Islamic ethics during design, not after release.
5. Digital literacy. Community and mosque programmes should teach *tabayyun* as a practical skill suited to synthetic media, paired with the producer-side duty of *sidq*, so that the public sees integrity as a duty of sharing as well as receiving.
6. Future research. The framework should be tested empirically, for example through expert consensus methods to validate its components, survey instruments to measure digital integrity among Muslim users, and case studies of how institutions apply disclosure rules.

Conclusion

Generative AI has made falsehood cheap, fluent and hard to detect, and the resulting loss of trust is the core of the crisis of digital integrity. This paper has argued that Islamic ethics addresses that crisis more completely than the current literature suggests. *Sidq* places the duty of truth on everyone who produces or passes on content, the prohibitions on *ghish* and *tadlis* set limits on how generative systems may be designed and presented, and *amanah* ensures that responsibility never rests with the machine. The proposed Digital Integrity Framework moves the discussion beyond receiver-side *tabayyun* to a chain of duties shared by developers, deployers, users, receivers and authorities. The framework is conceptual and needs empirical testing, but it shows that the Islamic tradition can contribute a coherent and motivating account of integrity to the global conversation on AI ethics.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

Acknowledgement

We would like to thank all those who contributed directly and indirectly to this study.

References

- Auda, J. (2008). *Maqasid al-shariah as philosophy of Islamic law: A systems approach*. International Institute of Islamic Thought.
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). Association for Computing Machinery. <https://doi.org/10.1145/3442188.3445922>
- Chesney, R., & Citron, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review*, 107(6), 1753–1820. <https://doi.org/10.15779/Z38RV0D15J>
- Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2), 228–239. <https://doi.org/10.1080/14703297.2023.2190148>
- Elmahjub, E. (2023). Artificial intelligence (AI) in Islamic ethics: Towards pluralist ethical benchmarking for AI. *Philosophy & Technology*, 36(4), Article 73. <https://doi.org/10.1007/s13347-023-00668-x>
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical

- framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- Frankfurt, H. G. (2005). *On bullshit*. Princeton University Press.
- Ghaly, M. (2024). What makes work "good" in the age of artificial intelligence (AI)? Islamic perspectives on AI-mediated work ethics. *The Journal of Ethics*, 28(3), 429–453. <https://doi.org/10.1007/s10892-023-09456-3>
- Hicks, M. T., Humphries, J., & Slater, J. (2024). ChatGPT is bullshit. *Ethics and Information Technology*, 26(2), Article 38. <https://doi.org/10.1007/s10676-024-09775-5>
- Jaakkola, E. (2020). Designing conceptual articles: Four approaches. *AMS Review*, 10(1–2), 18–26. <https://doi.org/10.1007/s13162-020-00161-0>
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12), Article 248. <https://doi.org/10.1145/3571730>
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399. <https://doi.org/10.1038/s42256-019-0088-2>
- Kamali, M. H. (1997). *Freedom of expression in Islam* (Rev. ed.). Islamic Texts Society.
- Kannike, U. M. M., & Fahm, A. O. (2025). Exploring the ethical governance of artificial intelligence from an Islamic ethical perspective. *Jurnal Fiqh*, 22(1), 134–161. <https://doi.org/10.22452/fiqh.vol22no1.5>
- Kreps, S., McCain, R. M., & Brundage, M. (2022). All the news that's fit to fabricate: AI-generated text as a tool of media misinformation. *Journal of Experimental Political Science*, 9(1), 104–117. <https://doi.org/10.1017/XPS.2020.37>
- Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology*, 6(3), 175–183. <https://doi.org/10.1007/s10676-004-3422-1>
- Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, 1(11), 501–507. <https://doi.org/10.1038/s42256-019-0114-4>
- Mohadi, M., & Tarshany, Y. (2023). Maqasid al-shari'ah and the ethics of artificial intelligence: Contemporary challenges. *Journal of Contemporary Maqasid Studies*, 2(2). <https://doi.org/10.52100/jcms.v2i2.107>
- Mohamad, N. M., & Mat Isa, N. H. (2022). The principles of tabayyun according to the Qur'an: Solutions to the challenges and strife of the social-media. *Al-Burhan: Journal of Qur'an and Sunnah Studies*, 6(Special Issue 1), 1–14. <https://journals.iium.edu.my/al-burhan/index.php/al-burhan/article/view/260>
- Nightingale, S. J., & Farid, H. (2022). AI-synthesized faces are indistinguishable from real faces and more trustworthy. *Proceedings of the National Academy of Sciences*, 119(8), Article e2120481119. <https://doi.org/10.1073/pnas.2120481119>
- Raquist, A., Channa, B., Zubair, T., & Qadir, J. (2022). Islamic virtue-based ethics for artificial intelligence. *Discover Artificial Intelligence*, 2(1), Article 11. <https://doi.org/10.1007/s44163-022-00028-2>
- Santoni de Sio, F., & Mecacci, G. (2021). Four responsibility gaps with artificial intelligence: Why they matter and how to address them. *Philosophy & Technology*, 34(4), 1057–1084. <https://doi.org/10.1007/s13347-021-00450-x>
- Simon, F. M., Altay, S., & Mercier, H. (2023). Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown. *Harvard Kennedy School Misinformation Review*, 4(5). <https://doi.org/10.37016/mr-2020-127>
- Spitale, G., Biller-Andorno, N., & Germani, F. (2023). AI model GPT-3 (dis)informs us better than humans. *Science Advances*, 9(26), Article eadh1850. <https://doi.org/10.1126/sciadv.adh1850>
- UNESCO. (2022). *Recommendation on the ethics of artificial intelligence*. United Nations Educational, Scientific and Cultural Organization. <https://unesdoc.unesco.org/ark:/48223/pf0000381137>

- Uriawan, W., Rahman, I. F., Zidan, M., Rohmatillah, I., Raihan, M. A., & Dwiyantri, I. (2025). *Preventing AI deepfake abuse: An Islamic ethics framework*. arXiv. <https://doi.org/10.48550/arXiv.2512.17218>
- Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society*, 6(1). <https://doi.org/10.1177/2056305120903408>
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151. <https://doi.org/10.1126/science.aap9559>
- Walters, W. H., & Wilder, E. I. (2023). Fabrication and errors in the bibliographic citations generated by ChatGPT. *Scientific Reports*, 13, Article 14045. <https://doi.org/10.1038/s41598-023-41032-5>
- Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policy making* (Council of Europe Report DGI(2017)09). Council of Europe.

Publisher: CLM Publishing Resources Malaysia



Open Access: This article is licensed under a [Creative Commons Attribution 4.0 International License](#), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third-party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Data Availability Statement: All relevant data are within the manuscript and its [Supporting Information](#) files.