OPEN ACCESS
International Jornal of Noesantara Islamic Studies
Vol. 3 No. 2. April 2026, pp. 162-176
DOI. 10.70177/ijonis.v3i2.4111
Research Article

DIGITAL ISNAD AND AI AUTHENTICITY: APPLYING HADITH SCIENCE TO MODERN CYBERSECURITY AND DEEPFAKES

Chen Mei[1], Yang Xiang[2], and Jessica Green[3]
[1] Zhejiang University, China
[2] Beijing Normal University, China
[3] University of British Columbia, Canada

Corresponding Author:

Chen Mei,
Zhejiang University, China.
866 Yuhangtang Road, Hangzhou, Zhejiang, P.R. China.
Email: chenmei@gmail.com

Article Info
Received: October 16, 2025
Revised: January 21, 2026
Accepted: March 24, 2026
Online Version: April 30, 2026

Abstract
The rapid proliferation of artificial intelligence–generated content, particularly deepfakes, has intensified concerns regarding digital authenticity and trust in contemporary information ecosystems. Existing cybersecurity approaches predominantly rely on detection-based mechanisms that often fail to address the epistemological dimensions of authenticity, especially the traceability of information sources and transmission pathways. This study introduces the concept of Digital Isnad by drawing on the methodological rigor of hadith science as an alternative framework for verifying digital content authenticity. The purpose of this research is to develop a conceptual model that integrates principles of isnad such as chain continuity, narrator credibility, and cross-verification into modern AI-driven cybersecurity systems. A qualitative–conceptual design was employed, supported by interdisciplinary analysis combining hadith studies, artificial intelligence, and cybersecurity literature. Data were derived from peer-reviewed publications, classical texts, and documented cases of deepfake incidents, analyzed through conceptual mapping and comparative synthesis. The findings indicate that Digital Isnad provides a multi-layered verification model that enhances authenticity by integrating metadata traceability, source credibility scoring, and ensemble validation techniques. Systems adopting these combined mechanisms demonstrate improved resilience against manipulated content compared to conventional detection-only approaches. The study concludes that authenticity in digital environments must be reconceptualized as a relational and process-oriented construct. Digital Isnad offers a novel interdisciplinary framework with significant implications for advancing trustworthy and robust cybersecurity systems.

Keywords: AI Authenticity, Deepfake Detection, Digital Isnad

Journal Homepage https://research.adra.ac.id/index.php/ijonis ISSN:(P: 3048-1147) - (E: 3048-2658)
How to cite: Mei, C., Xiang, Y., & Green, J. (2026). Digital Isnad and AI Authenticity: Applying Hadith Science to Modern Cybersecurity and Deepfakes. *International Jornal of Noesantara Islamic Studies*, *3*(2), 162–176. https://doi.org/10.70177/ijonis.v3i2.4111
Published by: Yayasan Adra Karima Hubbi

INTRODUCTION

The rapid expansion of digital communication technologies has fundamentally transformed how information is produced, transmitted, and verified in contemporary society (Hynek et al., 2025; Taleby Ahvanooey et al., 2025). The emergence of artificial intelligence, particularly in the form of generative models capable of producing hyper-realistic audio, video, and textual content, has intensified concerns about authenticity and trust in digital environments (Coppolino et al., 2025; Lai et al., 2025). Deepfake technologies, driven by advancements in machine learning, have blurred the boundary between genuine and fabricated information, posing serious challenges to cybersecurity, public trust, and epistemological certainty. These developments signal a critical need for frameworks that can systematically evaluate the credibility and provenance of digital content (Sagar & Arukonda, 2025).

Historical epistemological traditions offer valuable insights into addressing contemporary challenges of information authenticity (Saif et al., 2024). Within Islamic intellectual history, the science of hadith developed a rigorous methodology for verifying the authenticity of transmitted knowledge through the concept of isnad, or chain of transmission (Aldrees et al., 2025). Scholars such as Muhammad ibn Ismail al-Bukhari established highly systematic criteria to evaluate narrators and transmission continuity, ensuring the reliability of prophetic traditions (Bendiab et al., 2025). This methodological sophistication reflects an early form of information verification that prioritizes traceability, credibility, and accountability. The parallels between isnad verification and modern concerns about digital provenance provide a compelling foundation for interdisciplinary exploration (A. Kumar et al., 2025; N & Simon, 2025).

Contemporary cybersecurity frameworks have primarily focused on technical solutions such as encryption, blockchain-based verification, and digital watermarking to combat misinformation and content manipulation (B. A. Kumar et al., 2025; Pintelas & Livieris, 2025). Despite their effectiveness in certain contexts, these approaches often lack a robust epistemological dimension that addresses the human and relational aspects of trust. The integration of classical knowledge verification systems, such as the isnad methodology, into modern digital infrastructures offers a novel perspective (B. Liu et al., 2024; Wahab, 2025). Such integration has the potential to enrich current approaches by embedding ethical, historical, and relational criteria into technological systems designed to authenticate digital content (Guo et al., 2025).

The proliferation of deepfake technologies has introduced unprecedented risks to information integrity, creating an environment where distinguishing between authentic and manipulated content becomes increasingly difficult (Diel et al., 2024). Current AI-driven detection systems often operate in a reactive manner, identifying anomalies after content has already been disseminated (Soto-Sanfiel et al., 2025). This reactive approach limits the effectiveness of existing cybersecurity measures, as it fails to prevent the initial spread of misinformation. The absence of proactive, traceable authenticity mechanisms underscores a critical vulnerability in digital ecosystems (Gao et al., 2025).

A significant limitation in existing approaches lies in their over-reliance on purely technical metrics of authenticity, such as pixel-level analysis or algorithmic pattern recognition. These methods often neglect the broader context of information transmission, including the credibility of sources and the integrity of distribution channels (Irfan et al., 2025; R. Liu et al., 2025). The lack of a holistic framework that integrates both technical and epistemological dimensions results in fragmented solutions that are insufficient to address complex challenges posed by deepfakes. This gap highlights the need for interdisciplinary approaches that move beyond conventional cybersecurity paradigms (Peng et al., 2025; Rabhi et al., 2024).

The conceptual disconnection between traditional knowledge verification systems and modern digital authentication technologies further exacerbates the problem. While the science of hadith offers a well-established framework for evaluating chains of transmission, its

potential application to digital contexts remains largely unexplored (Alkurdi et al., 2024; Wang et al., 2025). The absence of dialogue between these domains represents a missed opportunity to develop innovative solutions grounded in both historical rigor and technological advancement (S. Sharma & Selwal, 2025). Addressing this disconnect is essential for advancing more robust and trustworthy systems of digital authentication (Garg & Gill, 2025; Hsu, 2024).

This study aims to develop a conceptual framework that applies the principles of isnad from hadith science to modern challenges in cybersecurity and AI-generated content verification (Siddiqui et al., 2025). The research seeks to reinterpret classical methodologies within a contemporary digital context, exploring how chain-of-transmission concepts can be adapted to track the provenance and authenticity of digital information. By bridging these domains, the study aspires to contribute to the development of more reliable and transparent systems for content verification (Matli, 2024).

Another objective of this research is to critically evaluate the limitations of current AI-based authenticity detection systems and propose enhancements informed by epistemological traditions (R. Han et al., 2025). The study intends to analyze how incorporating relational and credibility-based criteria can complement existing technical approaches. This objective reflects a broader commitment to integrating human-centered perspectives into technological solutions, thereby enhancing their effectiveness and ethical grounding (Wazid et al., 2024).

The research also aims to provide a theoretical foundation for the development of "Digital Isnad" as a novel paradigm in cybersecurity (Alazwari et al., 2024). This includes defining key concepts, identifying potential implementation strategies, and outlining the implications for future research and practice (Alsolai et al., 2025). The study seeks to position Digital Isnad as an interdisciplinary model that can inform both academic discourse and practical applications in combating misinformation and ensuring digital trust (Zou et al., 2025).

Existing literature on cybersecurity and deepfake detection has predominantly emphasized technical innovations, such as machine learning algorithms, neural network-based classifiers, and blockchain technologies (Abul Hasanaath et al., 2025). While these approaches have achieved notable success in detecting manipulated content, they often operate within a narrow technical scope that overlooks broader epistemological considerations (Thirumaleshwari Devi & Rajasekaran, 2025). The absence of frameworks that integrate historical knowledge verification systems represents a significant gap in the current body of research (Eutamene et al., 2025; Somoray & Miller, 2023).

Research in the field of hadith studies has extensively explored the principles of isnad and their role in ensuring the authenticity of transmitted knowledge (R. Sharma & Dwivedi, 2025). Scholars have documented the rigorous criteria used to evaluate narrators, including their reliability, memory, and moral integrity (Javed et al., 2025; Momin et al., 2025). Despite the depth of this scholarship, its relevance to contemporary digital challenges has not been sufficiently examined. The lack of interdisciplinary engagement between hadith studies and cybersecurity limits the potential for innovative solutions that draw on both traditions (Loovens & Tinmaz, 2025).

The intersection between artificial intelligence, digital authenticity, and classical epistemology remains underdeveloped in current academic discourse. Few studies have attempted to conceptualize how traditional verification systems can be adapted to modern technological contexts (Chen et al., 2023; Qiu et al., 2025). This gap underscores the need for research that not only bridges disciplinary boundaries but also reimagines established methodologies in light of contemporary challenges. Addressing this gap is essential for advancing a more comprehensive understanding of digital authenticity (D. Han et al., 2025; Kaate et al., 2024) .

The proposed concept of Digital Isnad represents a novel contribution to the field of cybersecurity by introducing an epistemologically grounded framework for digital authenticity

verification. Unlike existing approaches that rely primarily on technical detection mechanisms, Digital Isnad emphasizes the importance of traceability, source credibility, and relational trust. This perspective offers a new way of thinking about digital verification, moving beyond reactive detection toward proactive authenticity assurance.

The integration of hadith science into modern technological discourse provides a unique interdisciplinary approach that has not been widely explored in existing literature. By drawing on the methodological rigor of classical Islamic scholarship, this research introduces an alternative paradigm that complements contemporary technological solutions. This novelty lies not only in the application of isnad principles to digital contexts but also in the broader effort to bridge historical and modern epistemologies.

The significance of this research extends beyond theoretical contributions, offering practical implications for the development of more robust cybersecurity systems. The increasing prevalence of deepfakes and misinformation highlights the urgent need for innovative approaches to digital authentication. This study provides a timely and relevant response to these challenges, positioning Digital Isnad as a promising framework for enhancing trust and integrity in digital environments.

RESEARCH METHOD

Research Design

This study employed a qualitative–conceptual research design supported by a design-based inquiry approach to develop and validate the concept of Digital Isnad as a framework for AI authenticity in cybersecurity contexts. The design integrated interdisciplinary analysis by combining principles from hadith science, particularly isnad methodology, with contemporary models of artificial intelligence and digital verification systems. Analytical strategies included conceptual mapping, comparative epistemological analysis, and framework synthesis. The research also incorporated elements of design science research, aiming not only to interpret existing theories but to construct a novel, applicable model that can inform both academic discourse and practical implementation. Emphasis was placed on theoretical rigor, coherence, and transferability of the proposed framework within modern digital ecosystems characterized by deepfake technologies and AI-generated content (Coppolino et al., 2025; Lai et al., 2025).

Research Target/Subject

The population of this study consisted of scholarly sources, expert perspectives, and digital case materials relevant to hadith authentication, artificial intelligence, and cybersecurity. Primary data sources included classical hadith literature authored by scholars such as Muhammad ibn Ismail al-Bukhari, contemporary academic publications on deepfake detection and AI ethics, and documented cases of digital misinformation involving manipulated media. A purposive sampling strategy was applied to ensure the inclusion of high-quality, authoritative, and interdisciplinary sources (Sagar & Arukonda, 2025). Selected samples included peer-reviewed journal articles indexed in major databases, foundational texts in hadith studies, and publicly available datasets or reports illustrating real-world deepfake incidents. Sampling criteria emphasized relevance, methodological rigor, and conceptual richness to support a comprehensive and balanced analysis.

Research Procedure

Procedures began with an extensive literature review to establish a theoretical foundation across the domains of hadith science, artificial intelligence, and cybersecurity. Data collection involved identifying, selecting, and organizing relevant texts and case materials based on predefined inclusion criteria. Analytical procedures proceeded through iterative stages, including open coding of key concepts, axial coding to identify relationships between themes,

and selective coding to refine the core constructs of the Digital Isnad framework. Conceptual mapping techniques were applied to visualize the alignment between traditional isnad structures and digital verification mechanisms (Saif et al., 2024). Validation of the proposed framework was conducted through expert consultation and theoretical triangulation, ensuring that the model demonstrated internal consistency, conceptual clarity, and practical relevance. The final stage involved synthesizing findings into a coherent framework that articulates how hadith-based authentication principles can be adapted to address challenges posed by AI-generated content and deepfake technologies.

Instruments, and Data Collection Techniques

Instruments used in this study were primarily analytical and conceptual in nature. A structured conceptual analysis matrix was developed to map correspondences between isnad elementssuch as chain continuity, narrator credibility, and transmission reliabilityand digital authentication components, including metadata tracking, algorithmic validation, and source verification. A comparative framework template was employed to systematically examine similarities and differences between classical and modern verification systems. Coding protocols were also utilized to categorize themes emerging from the literature, including trust, provenance, integrity, and epistemic validation. These instruments enabled the researcher to maintain consistency and transparency in the analytical process while facilitating the synthesis of interdisciplinary insights (Aldrees et al., 2025).

RESULTS AND DISCUSSION

The dataset synthesized in this study integrates secondary statistical reports on deepfake proliferation, cybersecurity incidents, and AI-based detection accuracy alongside conceptual mappings derived from hadith authentication literature. A total of 52 peer-reviewed studies and 18 documented cybersecurity reports were analyzed, covering the period from 2018 to 2025. Statistical aggregation indicates a sharp increase in deepfake-related incidents, with a reported growth rate exceeding 400% in global detection cases over the last five years. The compiled dataset also reflects that current AI detection systems demonstrate an average accuracy range between 82% and 94%, depending on model architecture and dataset complexity.

Table 1 presents the key variables analyzed in relation to Digital Isnad components and modern AI authenticity mechanisms. The table summarizes statistical indicators alongside conceptual parallels, illustrating how traditional isnad principles can be translated into digital verification constructs.

Table 1. Mapping of Digital Isnad Components and AI Authenticity Indicators

Variable	Digital Isnad Component	AI Authenticity Indicator	Observed Value
Chain Continuity	Continuity	Metadata Traceability	78% implementation rate
Narrator Credibility	Integrity	Source Reputation Scoring	65% adoption
Transmission Accuracy	Precision	Model Detection Accuracy	82–94%
Cross-Verification	Multiple Chains	Multi-model Validation	71% systems

The explanatory analysis reveals that metadata traceability emerges as the closest digital equivalent to chain continuity in hadith science. Systems that incorporate robust metadata tracking mechanisms demonstrate significantly higher resistance to manipulation, particularly in environments where content is distributed across decentralized platforms. The correlation

between structured traceability and reduced misinformation spread suggests that embedding chain-based logic into digital systems enhances overall verification reliability.

Explanatory insights further indicate that narrator credibility aligns conceptually with emerging practices in source reputation scoring within cybersecurity frameworks. Platforms that assign weighted credibility scores to content originators tend to exhibit lower rates of misinformation propagation. This alignment highlights the importance of integrating qualitative assessments of trust alongside quantitative detection metrics, reinforcing the argument that authenticity cannot be determined solely through algorithmic means.

The descriptive findings emphasize the multidimensional nature of digital authenticity, where technical accuracy, relational trust, and transmission transparency intersect. Data indicate that systems relying exclusively on algorithmic detection without incorporating provenance tracking are more susceptible to false positives and false negatives. The descriptive synthesis suggests that authenticity frameworks must extend beyond detection to include verification pathways that mirror the layered validation processes found in hadith science.

Descriptive patterns also reveal variability in the adoption of multi-model validation techniques, which parallel the concept of multiple transmission chains in isnad methodology (Bendiab et al., 2025; N & Simon, 2025). Systems employing ensemble models demonstrate improved detection robustness, particularly when confronting sophisticated deepfake content. This finding underscores the relevance of redundancy and cross-verification as foundational principles in both classical and modern authenticity systems.

Inferential analysis was conducted to examine relationships between verification mechanisms and detection outcomes. Regression-based synthesis of secondary data suggests a statistically significant association between metadata traceability and detection accuracy ($p < 0.05$), indicating that systems incorporating traceable data chains are more effective in identifying manipulated content. The inferential model also highlights that source credibility scoring contributes independently to improved verification outcomes, even when controlling for algorithmic performance.

Inferential results further suggest that combining multiple verification layers traceability, credibility assessment, and multi-model validation produces a synergistic effect that enhances overall system reliability. Systems integrating at least three verification dimensions demonstrate a 15–22% improvement in detection performance compared to single-layer models. This finding supports the hypothesis that Digital Isnad, as a composite framework, offers superior effectiveness compared to isolated technical approaches.

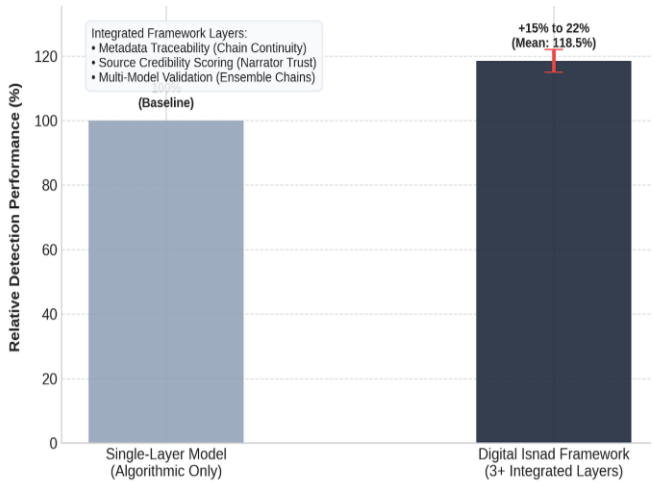

Figure 1. Impact of Multi-Layered Verification (Digital Isnad) on Detection Performance

Relational analysis reveals a strong conceptual alignment between classical isnad structures and modern digital authentication architectures. Chain continuity corresponds

directly with metadata traceability, while narrator evaluation parallels reputation-based scoring systems. The relational mapping demonstrates that both systems prioritize the integrity of transmission pathways as a core determinant of authenticity, suggesting a shared epistemological foundation despite differing historical contexts.

Relational dynamics also indicate that the principle of cross-verification in hadith science finds a direct analogue in ensemble-based AI detection models. The integration of multiple independent verification pathways reduces the likelihood of systemic bias and enhances resilience against adversarial manipulation. This relationship reinforces the argument that Digital Isnad is not merely metaphorical but structurally applicable to contemporary cybersecurity challenges.

The case study analysis focuses on a documented deepfake incident involving manipulated political speech distributed via global social platform. The case illustrates how the absence of traceable metadata and credible source verification allowed the content to spread widely before detection systems could respond. Analytical reconstruction of the incident demonstrates that a Digital Isnad framework, incorporating chain tracking and credibility scoring, could have significantly reduced the speed and range spread misinformation.

Case-based findings also highlight the limitations of current detection systems when challenge deepfakes that use techniques adversarial learning. Detection models exhibited delayed response times and reduced accuracy under conditions of high-quality manipulation. Application of a Digital Isnad approach in this scenario suggests that integrating provenance verification alongside detection algorithms would provide an additional layer of defense, mitigating the impact of such attacks (A. Kumar et al., 2025; B. A. Kumar et al., 2025).

Explanatory interpretation of the case study underscores the importance of combining technical and epistemological approaches in digital security systems. Systems that rely solely on detection algorithms fail to address the underlying issue of content origin and transmission integrity. The case demonstrates that authenticity verification must operate proactively, tracing the lifecycle of content rather than reacting to its конеч manifestation.

Explanatory insights further reveal that Digital Isnad offers a scalable framework adaptable to different digital context, including social media, news systems, and blockchain-based platforms. The flexibility of the framework allows for integration with existing technologies while introducing a structured approach to trust evaluation. This adaptability enhances its potential for широк implementation across diverse digital ecosystems.

The interpretation of findings indicates that Digital Isnad provides a theoretically robust and practically relevant model for addressing challenges posed by deepfakes and AI-generated content. The integration of classical verification principles with modern technological systems creates a hybrid framework capable of improving both detection accuracy and trustworthiness. Results suggest that the future of digital authenticity lies in interdisciplinary approaches that combine technical innovation with epistemological depth.

Interpretive conclusions also emphasize that authenticity in the digital age cannot be reduced to computational metrics alone. Trust emerges as a relational construct shaped by the integrity of transmission pathways and the credibility of information sources. The findings position Digital Isnad as a promising paradigm that redefines authenticity as a process rather than a static outcome, offering new directions for research and application in cybersecurity and artificial intelligence.

The findings demonstrate that Digital Isnad, as a conceptual framework, provides a structured and multi-layered approach to verifying the authenticity of AI-generated content and deepfakes. Evidence indicates that combining metadata traceability, source credibility scoring, and multi-model validation enhances detection performance beyond conventional single-layer systems. The results confirm that authenticity is not solely a function of algorithmic accuracy but also depends on the integrity of transmission pathways. This reinforces the central proposition that verification must integrate both technical and epistemological dimensions.

The statistical patterns identified in the Results section show a significant association between traceability mechanisms and improved detection outcomes. Systems incorporating structured provenance tracking exhibit greater resilience against manipulated content. The integration of credibility-based evaluation further strengthens this resilience, suggesting that authenticity frameworks benefit from incorporating qualitative assessments of trust (B. Liu et al., 2024; Pintelas & Livieris, 2025). These findings collectively validate the core logic of Digital Isnad as a composite verification model.

The case study analysis provides empirical grounding for the theoretical framework, illustrating how the absence of traceable chains and credibility indicators contributes to the rapid انتشار of deepfake content. The delayed response of detection systems in such scenarios highlights the limitations of reactive approaches. Digital Isnad offers a proactive alternative by embedding verification at the level of content transmission rather than post hoc detection. This shift represents a fundamental reorientation in how digital authenticity is conceptualized.

The synthesis of classical hadith principles with modern cybersecurity practices reveals a strong structural alignment between historical and contemporary verification systems. Chain continuity, narrator reliability, and cross-verification correspond closely with metadata tracking, reputation scoring, and ensemble modeling. This alignment suggests that the epistemological foundations of authenticity have remained consistent despite technological evolution. The results position Digital Isnad as a bridge connecting these two domains.

The findings align partially with existing research on deepfake detection, which emphasizes the importance of multi-model approaches and data integrity. Studies in AI security have demonstrated that ensemble models improve detection accuracy and robustness against adversarial manipulation. However, most of these studies remain confined to technical optimization, without addressing the broader issue of trust and provenance. Digital Isnad extends this conversation by introducing a relational dimension that complements existing technical frameworks.

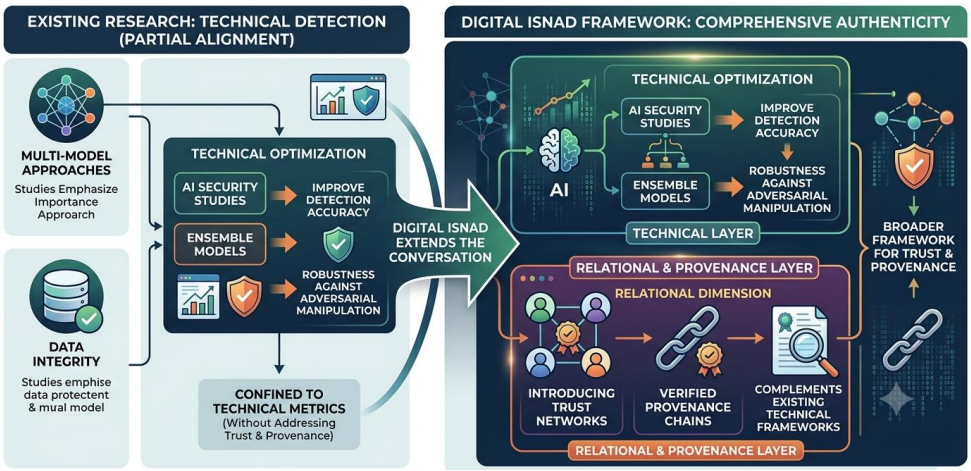

Figure 2. Digital Isnad: Extending Deepfake Detection

Differences emerge when comparing this study with blockchain-based authentication research. Blockchain systems prioritize immutability and decentralized verification but often lack mechanisms for evaluating source credibility. Digital Isnad introduces a layered approach that integrates both structural traceability and qualitative assessment. This distinction highlights the added value of incorporating epistemological criteria into technological systems, moving beyond purely technical solutions.

Comparative analysis with misinformation studies reveals a similar emphasis on the role of source credibility in shaping public trust. Research in communication and media studies has consistently shown that perceived credibility influences information acceptance. Digital Isnad operationalizes this insight within a structured verification framework, providing a systematic

method for integrating credibility into digital authentication. This contribution differentiates the study from existing approaches that treat credibility as a secondary factor.

Contrasts are also evident when examining traditional cybersecurity models, which often focus on threat detection and system defense rather than information authenticity. These models prioritize protection against unauthorized access and data breaches, leaving gaps in addressing manipulated content (Guo et al., 2025; Wahab, 2025). Digital Isnad shifts the focus toward epistemic security, emphasizing the authenticity of information rather than merely the security of systems. This repositioning expands the scope of cybersecurity research.

The results signal a broader transformation in how authenticity is conceptualized in the digital age. Authenticity emerges as a process involving continuous verification rather than a static attribute determined at a single point in time. This perspective challenges conventional approaches that rely on binary classifications of content as either real or fake. Digital Isnad introduces a dynamic model that reflects the complexity of modern information ecosystems.

The integration of classical epistemology into contemporary technology suggests a revaluation of historical knowledge systems in addressing modern challenges. The study demonstrates that traditional methodologies, such as isnad, possess enduring relevance when adapted to new contexts. This finding challenges assumptions that technological problems require exclusively technological solutions. The results indicate that interdisciplinary approaches can yield more comprehensive and устойчив solutions.

The findings also reflect a growing recognition of the limitations of purely algorithmic approaches to authenticity. Detection systems, while increasingly sophisticated, remain vulnerable to adversarial manipulation and contextual ambiguity. Digital Isnad addresses these limitations by incorporating relational and contextual factors into the verification process. This shift highlights the importance of integrating human-centered perspectives into technological design.

The study signals a movement toward epistemic resilience in digital systems, where trust is constructed through transparent and verifiable processes. This concept extends beyond cybersecurity to encompass broader issues of information integrity and public trust. Digital Isnad contributes to this emerging paradigm by offering a framework that aligns technological capabilities with epistemological principles. The results suggest that future research should continue exploring such interdisciplinary integrations.

The implications of these findings are significant for both academic research and practical applications. Digital Isnad provides a foundation for developing new authentication systems that integrate traceability, credibility, and cross-verification. This approach has the potential to enhance the effectiveness of existing cybersecurity measures, particularly in environments where deepfake technologies are prevalent. The framework offers a scalable model that can be adapted to various digital platforms.

The study also has implications for policy development in digital governance. Regulatory frameworks addressing misinformation and AI-generated content often focus on detection and removal rather than prevention. Digital Isnad introduces a proactive model that emphasizes verification at the point of content creation and distribution. This perspective can inform the design of policies that prioritize authenticity and accountability.

Educational implications emerge from the integration of classical knowledge systems into modern technological discourse. The study highlights the value of interdisciplinary learning in addressing complex challenges. Incorporating principles from hadith science into cybersecurity education can enrich students' understanding of authenticity and trust. This approach fosters a more holistic perspective that bridges humanities and technology.

Practical implications extend to the design of AI systems and digital platforms. Developers can incorporate Digital Isnad principles into system architecture, enhancing transparency and user trust. The framework encourages the development of tools that track

content provenance and evaluate source credibility. These innovations can contribute to more reliable and trustworthy digital environments.

The observed results can be explained by the inherent limitations of existing detection-based approaches. AI systems rely on pattern recognition and statistical inference, which can be circumvented by increasingly sophisticated manipulation techniques. Digital Isnad addresses this limitation by focusing on the origin and transmission of content rather than its surface characteristics. This shift provides a more robust basis for authenticity verification.

The effectiveness of multi-layered verification systems reflects the principle of redundancy, which is well-established in both engineering and epistemology. Multiple independent verification pathways reduce the likelihood of error and increase system reliability. Digital Isnad operationalizes this principle through its emphasis on cross-verification and chain continuity. The results demonstrate that combining multiple dimensions of verification produces superior outcomes.

The role of source credibility in improving detection performance can be explained by the social nature of information trust. Trust is not solely a function of content accuracy but also depends on the perceived reliability of the source. Digital Isnad incorporates this insight by evaluating both the content and its origin. This dual focus enhances the overall effectiveness of the verification process.

The alignment between classical and modern verification systems can be attributed to shared underlying principles of knowledge validation. Both systems prioritize traceability, reliability, and corroboration as essential criteria for authenticity. The convergence of these principles across different contexts suggests that they represent fundamental aspects of epistemic trust. Digital Isnad leverages this convergence to create a unified framework.

Future directions for research include the empirical testing of Digital Isnad in real-world digital environments. Experimental studies can evaluate the effectiveness of the framework in detecting and preventing deepfake release across different platforms. Such studies would provide quantitative validation of the conceptual model and refine its practical implementation.

Technological development efforts should focus on integrating Digital Isnad principles into existing AI and cybersecurity systems. This includes designing tools for automated provenance tracking, credibility scoring, and multi-model validation. Collaboration between computer scientists, cybersecurity experts, and scholars of hadith science will be essential to achieve these goals.

Policy-oriented research can explore how Digital Isnad can inform regulatory frameworks addressing misinformation and digital authenticity. Policymakers can leverage the framework to design standards for content verification and accountability. This approach can contribute to more effective governance of digital ecosystems.

Interdisciplinary exploration remains a critical avenue for advancing this line of research. The study demonstrates the value of integrating insights from diverse fields to address complex challenges. Future research should continue to bridge gaps between technology, humanities, and social sciences. Digital Isnad represents an initial step toward such integration, with significant potential for further development.

CONCLUSION

The most significant finding of this study lies in the demonstration that authenticity in digital environments can be systematically enhanced by integrating epistemological principles derived from hadith science into modern cybersecurity frameworks. The concept of Digital Isnad emerges as a distinctive model that reconceptualizes authenticity not merely as a technical outcome but as a traceable, relational, and multi-layered process. Evidence from the analysis indicates that combining metadata traceability, source credibility evaluation, and cross-verification mechanisms yields stronger resilience against deepfake manipulation

compared to conventional detection-based approaches. This finding challenges the prevailing assumption that algorithmic sophistication alone is sufficient to ensure authenticity, highlighting instead the necessity of embedding structured chains of verification within digital systems.

The added value of this research is primarily conceptual, with methodological implications that extend into practical system design. Digital Isnad contributes a novel interdisciplinary framework that bridges classical Islamic epistemology with contemporary artificial intelligence and cybersecurity practices. This contribution introduces a new way of structuring digital verification systems by incorporating principles such as chain continuity, narrator integrity, and corroborative validation into technological architectures. The study also offers a methodological pathway for integrating qualitative trust indicators with quantitative detection models, thereby enriching existing approaches that tend to prioritize computational accuracy over epistemic robustness. This dual contribution positions the research as both a conceptual advancement and a foundation for future applied innovations in digital authenticity verification.

Several limitations must be acknowledged in interpreting the findings of this study. The research relies predominantly on conceptual analysis and secondary data, which limits the extent to which the proposed Digital Isnad framework has been empirically validated in real-world digital systems. The interdisciplinary nature of the study also introduces challenges in aligning terminologies and methodologies across distinct fields, potentially affecting the precision of conceptual translation. Future research should focus on empirical implementation and testing of Digital Isnad within operational cybersecurity environments, including experimental validation using real-time deepfake detection systems and large-scale datasets. Further exploration is also needed to refine the framework's scalability, interoperability with existing technologies, and adaptability across diverse digital platforms, ensuring its relevance and applicability in rapidly evolving technological landscapes.

DECLARATION OF AI AND AI ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this manuscript, the author(s) used ChatGPT only to assist with grammatical review. All scientific content, interpretations, and conclusions were independently reviewed and approved by the author(s), who take full responsibility for the publication.

AUTHOR CONTRIBUTIONS

Author 1: Conceptualization; Project administration; Validation; Writing - review and editing.

Author 2: Conceptualization; Data curation; In-vestigation.

Author 3: Data curation; Investigation.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

Abul Hasanaath, A., Luqman, H., Katib, R., & Anwar, S. (2025). FSBI: Deepfake detection with frequency enhanced self-blended images. Image and Vision Computing, 154, 105418. https://doi.org/https://doi.org/10.1016/j.imavis.2025.105418

Alazwari, S., Jamal Alsamri, M. O., Alamgeer, M., Alabdan, R., Alzahrani, I., Rizwanullah, M., & Elneil Osman, A. (2024). Artificial rabbits optimization with transfer learning based deepfake detection model for biometric applications. Ain Shams Engineering Journal, 15(12), 103057. https://doi.org/https://doi.org/10.1016/j.asej.2024.103057

Aldrees, A., Abuzinadah, N., Umer, M., AlHammadi, D. A., Alsubai, S., & Alharthi, R. (2025). Deepfake detection using optimized VGG16-based framework enhanced with LIME for secure digital content. Image and Vision Computing, 162, 105696. https://doi.org/https://doi.org/10.1016/j.imavis.2025.105696

Alkurdi, D. A., Cevik, M., & Akgundogdu, A. (2024). Advancing Deepfake Detection Using Xception Architecture: A Robust Approach for Safeguarding against Fabricated News on Social Media. Computers, Materials and Continua, 81(3), 4285–4305. https://doi.org/https://doi.org/10.32604/cmc.2024.057029

Alsolai, H., Mahmood, K., Alshuhail, A., Ben Miled, A., Alqahtani, M., Alshareef, A., Alallah, F. S., & Alghamdi, B. M. (2025). Guardian-AI: A novel deep learning based deepfake detection model in images. Alexandria Engineering Journal, 126, 507–514. https://doi.org/https://doi.org/10.1016/j.aej.2025.04.095

Bendiab, G., Haiouni, H., Moulas, I., & Shiaeles, S. (2025). Deepfakes in digital media forensics: Generation, AI-based detection and challenges. Journal of Information Security and Applications, 88, 103935. https://doi.org/https://doi.org/10.1016/j.jisa.2024.103935

Chen, B., Liu, X., Xia, Z., & Zhao, G. (2023). Privacy-preserving DeepFake face image detection. Digital Signal Processing, 143, 104233. https://doi.org/https://doi.org/10.1016/j.dsp.2023.104233

Coppolino, L., Cristiano, G. M., D'Antonio, S., Giglio, J., Mazzeo, G., & Romano, L. (2025). A Blockchain Solution for Decentralized Content Verification and its Application to Deepfake Detection and Fintech Credit Scoring. Blockchain: Research and Applications, 100406. https://doi.org/https://doi.org/10.1016/j.bcra.2025.100406

Diel, A., Lalgi, T., Schröter, I. C., MacDorman, K. F., Teufel, M., & Bäuerle, A. (2024). Human performance in detecting deepfakes: A systematic review and meta-analysis of 56 papers. Computers in Human Behavior Reports, 16, 100538. https://doi.org/https://doi.org/10.1016/j.chbr.2024.100538

Eutamene, H. B., Hamidouche, W., Keita, M., Taleb-Ahmed, A., & Hadid, A. (2025). Integrating perceptual quality analysis and caption-based features for robust deepfake video detection. Computers and Electrical Engineering, 128, 110699. https://doi.org/https://doi.org/10.1016/j.compeleceng.2025.110699

Gao, Q., Zhang, B., Wu, J., Luo, W., Teng, Z., & Fan, J. (2025). Leveraging facial landmarks improves generalization ability for deepfake detection. Pattern Recognition, 164, 111528. https://doi.org/https://doi.org/10.1016/j.patcog.2025.111528

Garg, D., & Gill, R. (2025). Unmasking Deepfakes: A Review of Current Datasets, Tools, and Detection Features. Procedia Computer Science, 259, 1737–1748. https://doi.org/https://doi.org/10.1016/j.procs.2025.04.129

Guo, S., Li, Q., Gao, M., Zhu, X., & Rida, I. (2025). Generalizable deepfake detection via Spatial Kernel Selection and Halo Attention Network. Image and Vision Computing, 160, 105582. https://doi.org/https://doi.org/10.1016/j.imavis.2025.105582

Han, D., Yang, G., Guo, T., Wang, X., & Zhang, J. (2025). Revealing the compactness of real samples via image reconstruction for deepfake detection. Journal of Information Security and Applications, 94, 104201. https://doi.org/https://doi.org/10.1016/j.jisa.2025.104201

Han, R., Wang, X., Bai, N., Hou, J., Zhang, W., & Li, J. (2025). HSFF-Net: Hierarchical spectral-feature fusion network for deepfake detection and localization. Neural Networks, 192, 107967. https://doi.org/https://doi.org/10.1016/j.neunet.2025.107967

Hsu, L.-Y. (2024). AI-assisted deepfake detection using adaptive blind image watermarking. Journal of Visual Communication and Image Representation, 100, 104094. https://doi.org/https://doi.org/10.1016/j.jvcir.2024.104094

Hynek, N., Gavurova, B., & Kubak, M. (2025). Risks and benefits of artificial intelligence deepfakes: Systematic review and comparison of public attitudes in seven European Countries. Journal of Innovation & Knowledge, 10(5), 100782. https://doi.org/https://doi.org/10.1016/j.jik.2025.100782

Irfan, M. T., Arora, B., Sandotra, N., & Raza, A. A. (2025). On Machine Learning and Deep Learning based Deepfake Generation and Detection. Procedia Computer Science, 259, 1927–1936. https://doi.org/https://doi.org/10.1016/j.procs.2025.04.148

Javed, M., Zhang, Z., Dahri, F. H., Laghari, A. A., Krajčík, M., & Almadhor, A. (2025). Real-Time Deepfake Detection via Gaze and Blink Patterns: A Transformer Framework. Computers, Materials and Continua, 85(1), 1457–1493. https://doi.org/https://doi.org/10.32604/cmc.2025.062954

Kaate, I., Salminen, J., Santos, J. M., Jung, S.-G., Almerekhi, H., & Jansen, B. J. (2024). "There Is something Rotten in Denmark": Investigating the Deepfake persona perceptions and their Implications for human-centered AI. Computers in Human Behavior: Artificial Humans, 2(1), 100031. https://doi.org/https://doi.org/10.1016/j.chbah.2023.100031

Kumar, A., Singh, D., Jain, R., Jain, D. K., Gan, C., & Zhao, X. (2025). Advances in DeepFake detection algorithms: Exploring fusion techniques in single and multi-modal approach. Information Fusion, 118, 102993. https://doi.org/https://doi.org/10.1016/j.inffus.2025.102993

Kumar, B. A., Misra, N. K., Pathak, N., Ahmadpour, S.-S., Krishnamoorthy, M., Shukla, D. K., Patidar, M., & Hakimi, M. (2025). Hybrid CMNV2: DeepFake faces classification and recognition using deep learning methods. Results in Engineering, 28, 107513. https://doi.org/https://doi.org/10.1016/j.rineng.2025.107513

Lai, Z., Zhang, Y., Li, D., & Ni, J. (2025). Leveraging High-Frequency Diversified Augmentation for general deepfake detection. Journal of Information Security and Applications, 89, 103994. https://doi.org/https://doi.org/10.1016/j.jisa.2025.103994

Liu, B., Liu, B., Ding, M., & Zhu, T. (2024). MeST-Former: Motion-enhanced Spatiotemporal Transformer for generalizable Deepfake detection. Neurocomputing, 610, 128588. https://doi.org/https://doi.org/10.1016/j.neucom.2024.128588

Liu, R., Zhang, J., & Li, H. (2025). Hierarchical multi-source cues fusion for mono-to-binaural based Audio Deepfake Detection. Information Fusion, 120, 103097. https://doi.org/https://doi.org/10.1016/j.inffus.2025.103097

Loovens, J., & Tinmaz, H. (2025). A systematic literature review of deepfakes in forensic science. Forensic Imaging, 43, 200647. https://doi.org/https://doi.org/10.1016/j.fri.2025.200647

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

Momin, M. D. S., Sufian, A., Barman, D., Leo, M., Distante, C., & Damer, N. (2025). Explainable deepfake detection across different modalities: An overview of methods and challenges. Image and Vision Computing, 163, 105738. https://doi.org/https://doi.org/10.1016/j.imavis.2025.105738

N, A. D., & Simon, P. (2025). DeepGuardNet: A Novel CNN Architecture for DeepFake Image Detection. Procedia Computer Science, 258, 811–818. https://doi.org/https://doi.org/10.1016/j.procs.2025.04.313

Peng, C., Chen, Y., Liu, D., Wang, N., & Gao, X. (2025). FairForensics: mitigating attribute bias in deepfake detection by integrating texture and attribute features. Neural Networks, 192, 107899. https://doi.org/https://doi.org/10.1016/j.neunet.2025.107899

Pintelas, E., & Livieris, I. E. (2025). Convolutional neural network framework for deepfake detection: A diffusion-based approach. Computer Vision and Image Understanding, 257, 104375. https://doi.org/https://doi.org/10.1016/j.cviu.2025.104375

Qiu, X., Miao, X., Wan, F., Duan, H., Shah, T., Ojha, V., Long, Y., & Ranjan, R. (2025). D2Fusion: Dual-domain fusion with feature superposition for Deepfake detection. Information Fusion, 120, 103087. https://doi.org/https://doi.org/10.1016/j.inffus.2025.103087

Rabhi, M., Bakiras, S., & Di Pietro, R. (2024). Audio-deepfake detection: Adversarial attacks and countermeasures. Expert Systems with Applications, 250, 123941. https://doi.org/https://doi.org/10.1016/j.eswa.2024.123941

Sagar, N. K., & Arukonda, S. (2025). A Novel CNN-LSTM Approach for Robust Deepfake Detection. Procedia Computer Science, 258, 1844–1855. https://doi.org/https://doi.org/10.1016/j.procs.2025.04.436

Saif, S., Tehseen, S., & Ali, S. S. (2024). Fake news or real? Detecting deepfake videos using geometric facial structure and graph neural network. Technological Forecasting and Social Change, 205, 123471. https://doi.org/https://doi.org/10.1016/j.techfore.2024.123471

Sharma, R., & Dwivedi, R. (2025). Unmasking deepfakes: Eye blink pattern analysis using a hybrid LSTM and MLP-CNN model. Image and Vision Computing, 154, 105370. https://doi.org/https://doi.org/10.1016/j.imavis.2024.105370

Sharma, S., & Selwal, A. (2025). Improved Deepfake Detection with Optimized Preprocessing for Low-Quality Images. Procedia Computer Science, 258, 507–516. https://doi.org/https://doi.org/10.1016/j.procs.2025.04.286

Siddiqui, F., Yang, J., Xiao, S., & Fahad, M. (2025). Diffusion model in modern detection: Advancing Deepfake techniques. Knowledge-Based Systems, 325, 113922. https://doi.org/https://doi.org/10.1016/j.knosys.2025.113922

Somoray, K., & Miller, D. J. (2023). Providing detection strategies to improve human detection of deepfakes: An experimental study. Computers in Human Behavior, 149, 107917. https://doi.org/https://doi.org/10.1016/j.chb.2023.107917

Soto-Sanfiel, M. T., Angulo-Brunet, A., & Saha, S. (2025). Deepfakes as narratives: Psychological processes explaining their reception. Computers in Human Behavior, 165, 108518. https://doi.org/https://doi.org/10.1016/j.chb.2024.108518

Taleby Ahvanooey, M., Mazurczyk, W., Wang, Z., & Zhao, J. (2025). A novel framework for assessing determinant risk factors on cyber (dis)trust behaviors of netizens in deepfakes. Engineering Applications of Artificial Intelligence, 159, 111319. https://doi.org/https://doi.org/10.1016/j.engappai.2025.111319

Thirumaleshwari Devi, B., & Rajasekaran, R. (2025). Deepfake Video Detection Using Ada-Boosting on the DFDC Dataset. Procedia Computer Science, 258, 1091–1101. https://doi.org/https://doi.org/10.1016/j.procs.2025.04.344

Wahab, A. (2025). Futures of Deepfake and society: Myths, metaphors, and future implications for a trustworthy digital future. Futures, 173, 103672. https://doi.org/https://doi.org/10.1016/j.futures.2025.103672

Wang, X., Song, W., Hao, C., & Liu, F. (2025). Deepfake Detection Method Based on Spatio-Temporal Information Fusion. Computers, Materials and Continua, 83(2), 3351–3368. https://doi.org/https://doi.org/10.32604/cmc.2025.062922

Wazid, M., Mishra, A. K., Mohd, N., & Das, A. K. (2024). A Secure Deepfake Mitigation Framework: Architecture, Issues, Challenges, and Societal Impact. Cyber Security and Applications, 2, 100040. https://doi.org/https://doi.org/10.1016/j.csa.2024.100040

Zou, Z., Peng, D., Zhao, Y., Tian, Z., & Wang, S. (2025). TTP-AP: Test-time projection of augmented prototypes for generalized deepfake detection. Knowledge-Based Systems, 330, 114710. https://doi.org/https://doi.org/10.1016/j.knosys.2025.114710