

Terbit online pada: https://ejournal.almarkazibkl.org/index.php/ince - E ISSN: 2961-9866 (Special Edition)
This article is licensed under CC BY-SA 4.0 International

EVALUATING DATA CREDIBILITY IN CYBERSPACE: AN ISLAMIC INFORMATION INTEGRITY FRAMEWORK BASED ON THE PRINCIPLES OF JARḤ WA TAʿDĪL

Abdul Baqi [1], Dr. Muhammad Ibrahim[2] ,Dr. Fozia Khan [3]
[1, 2] *Department of Islamic and Religious Studies, Hazara University, Mansehra, Pakistan.*
[3] *Department of Quranic Studies, The Islamia University of Bahawalpur, Pakistan*
[1]abdulbaqi.phdscholar@gmail.com

Received	*Revised*	*Accepted*
09 November 2026	*20 November 2026*	*21 November 2026*
https://doi.org/10.67528/ic.v5iSp1.233		

ABSTRACT

The proliferation of generative artificial intelligence, deepfakes, and digital misinformation has intensified concerns regarding the credibility, authenticity, and responsible transmission of information in cyberspace. Although contemporary approaches address content verification, source credibility, and digital provenance through various technological and analytical methods, the integration of ethical accountability with technical and evidentiary verification remains an important interdisciplinary concern. This study proposes the Islamic Information Integrity Framework (IIIF), a conceptual model informed by the Qur'anic principle of Tabayyun and the classical Islamic science of Jarḥ wa Taʿdīl. Employing a qualitative, documentary, and conceptual research design, the study examines four interconnected dimensions: ʿAdālah (moral integrity), Ḍabṭ (precision and accuracy), Isnād (transmission provenance), and Matn (content evaluation). These principles are conceptually adapted into proposed digital assessment dimensions concerning source accountability, technical data integrity, information traceability, and evidence-based content verification. The study further introduces the Digital Credibility Index (DCI) as a preliminary conceptual structure for organizing these dimensions, while distinguishing its proposed mathematical formulation from an empirically validated measurement instrument. An illustrative hypothetical case demonstrates how the four dimensions may be coordinated in assessing potentially misleading digital information. The study also identifies ethical concerns involving anonymous sources, privacy protection, and reputation-based bias. It concludes that IIIF offers a potential epistemological contribution to interdisciplinary research on digital information integrity, although its indicators, weighting assumptions, and practical effectiveness require expert assessment, empirical testing, and technical validation.

Keywords: *Information Integrity; Jarḥ wa Taʿdīl; Tabayyun; Digital Credibility; Artificial Intelligence.*

1. INTRODUCTION

The advent of the digital era has fundamentally transformed the landscape of information

Terbit online pada: https://ejournal.almarkazibkl.org/index.php/ince - E ISSN: 2961-9866 (Special Edition)
This article is licensed under CC BY-SA 4.0 International

creation, dissemination, and consumption, turning cyberspace into the primary arena for global knowledge exchange. While digital technologies, social media platforms, and artificial intelligence (AI) have democratized access to information, they have simultaneously birthed an unprecedented crisis of misinformation, disinformation, fabricated content, deepfakes, and algorithmic manipulation. In today's interconnected world, online data is no longer merely a collection of isolated points; rather, it constitutes a transmitted narrative whose credibility heavily depends upon the reliability of its source, the integrity of its transmission, and the authenticity of its content. Consequently, evaluating data credibility in cyberspace has become one of the most pressing challenges confronting researchers, computer scientists, and society at large, as this "information integrity crisis" continuously threatens social cohesion and public trust.

Contemporary approaches to combating digital misinformation primarily rely on technological interventions, such as automated fact-checking systems, machine learning algorithms, and AI-assisted detection tools. Although these methods have significantly improved the identification of fabricated content, they remain largely content-oriented and often overlook the epistemic trustworthiness of the information source itself. They rarely provide a holistic framework capable of evaluating the moral integrity, historical consistency, and behavioral accountability of those who generate and propagate data. Thus, modern computational models leave a critical vacuum in information integrity by prioritizing algorithmic patterns over structural and human trustworthiness. This structural vacuum has become increasingly perilous with the exponential rise of generative AI tools, which allow malicious actors to fabricate hyper-realistic textual and visual content at near-zero marginal cost. When computational architectures attempt to combat these threats through purely text-based or pixel-level classifiers, they enter an endless cat-and-mouse game against evolving generative models. This computational bottleneck reinforces the immediate historical necessity to pivot toward a system that evaluates the long-term behavioral footprint and ethical accountability of the node originating the data packet, rather than just the content itself.

From an Islamic epistemological perspective, information integrity is not merely a technical concern but a profound ethical, social, and religious obligation. The Qur'an explicitly establishes verification (*tabayyun*) as a fundamental principle of responsible communication, commanding believers to investigate information before accepting or sharing it (Qur'an 49:6). This Qur'anic mandate laid the intellectual foundation for one of the most sophisticated data-

Terbit online pada: https://ejournal.almarkazibkl.org/index.php/ince - E ISSN: 2961-9866 (Special Edition)
This article is licensed under CC BY-SA 4.0 International

verification systems in human history: the science of Hadith authentication. Through this tradition, classical Muslim scholars institutionalized the principle of *tabayyun* into a rigorous, peer-reviewed methodology known as *ʿIlm al-Jarḥ wa al-Taʿdīl* (the Science of Narrator Critical Appraisal) to safeguard transmitted knowledge.

The conceptual richness of Jarḥ wa Taʿdīl offers valuable epistemological insights for addressing information credibility challenges in the digital age. However, the application of its classical principles to cyberspace requires careful conceptual adaptation rather than direct equivalence. In classical Hadith scholarship, ʿAdālah concerns the narrator's moral integrity, while Ḍabṭ relates to precision in receiving, retaining, and transmitting reports. Isnād establishes the structure and continuity of transmission, whereas Matn criticism examines the transmitted report in relation to relevant evidence and established principles. Within digital environments, these principles may inform the assessment of source accountability, data integrity, information provenance, and content corroboration, respectively. Nevertheless, digital identity verification does not independently establish moral integrity, technical traceability does not prove factual accuracy, and computational consistency cannot replace substantive scholarly verification. Accordingly, this study treats the four principles as epistemological foundations for developing complementary digital assessment dimensions, while acknowledging the methodological differences between classical Hadith authentication and modern information systems.

2. LITERATURE REVIEW

2.1. Contemporary Computational Models and Their Epistemological Limitations:

In response to this crisis, the dominant paradigm in information science has been heavily computational. Modern scholarly interventions generally bifurcate into computational detection and cryptographic provenance. Machine learning classifiers, Natural Language Processing (NLP), and deep learning architectures are widely deployed to detect semantic anomalies (Zhou & Zafarani, 2020; Shu et al., 2017), while blockchain-based ledger systems and metadata analysis are researched to secure data lineage (Fraga-Lamas & Fernández-Caramés, 2019). Fallis (2009) and Graham (2020) argue that automated verification remains largely reactive and "content-oriented." By focusing strictly on textual patterns or visual pixel inconsistencies, contemporary frameworks treat data as isolated entities detached from human agency. As Sundar (2020) and Wardle & Derakhshan (2017) point out, algorithmic filters suffer from a profound "source-blindness"; they fail to evaluate

Terbit online pada: https://ejournal.almarkazibkl.org/index.php/ince - E ISSN: 2961-9866 (Special Edition)
This article is licensed under CC BY-SA 4.0 International

the continuous behavioral reliability, underlying intent, and moral integrity of the digital nodes propagating the data.

2.1.1. Comparative Review of Existing Digital Credibility Approaches

Existing research on digital credibility assessment has developed through several complementary approaches rather than a single content-based paradigm. A closer examination of the literature reveals that contemporary systems evaluate different dimensions of credibility, including message characteristics, source attributes, patterns of information dissemination, and the technical provenance of digital content.

Zhou and Zafarani (2020), in their comprehensive survey of fake news detection, identify four major perspectives: the credibility of information sources, the knowledge or claims contained in messages, writing styles, and information propagation patterns. Their work demonstrates that source-level evaluation is already an established component of contemporary misinformation research. Similarly, Shu et al. (2017) examine fake news detection through data-mining approaches that incorporate both content-related and social-contextual features. These studies indicate that the limitations of contemporary systems should not be characterized as a complete absence of source evaluation; rather, the challenge concerns the integration of source reliability, ethical accountability, transmission integrity, and content verification within a coherent epistemological framework.

Another important dimension concerns the credibility of online information sources. Metzger (2007) reviews theoretical and empirical approaches to evaluating online information and emphasizes the importance of assessing the credibility of information providers and the information they communicate. More recently, Hanimann et al. (2023) reviewed credibility measurement in media research and highlighted conceptual distinctions between source, message, and media credibility. These contributions establish that source credibility is an important and recognized field of investigation, although its measurement remains methodologically complex.

A further research stream focuses on digital provenance and cryptographic authentication. Blockchain-based systems and emerging content-provenance standards provide mechanisms for recording the origin and modification history of digital assets. However, the Coalition for Content Provenance and Authenticity (C2PA) explicitly distinguishes the integrity of provenance records from the factual truth of the associated content. A verifiable transmission history can establish where a digital asset originated and whether recorded modifications have occurred, but it cannot independently establish the

Terbit online pada: https://ejournal.almarkazibkl.org/index.php/ince - E ISSN: 2961-9866 (Special Edition)
This article is licensed under CC BY-SA 4.0 International

truthfulness of the original claim (C2PA, n.d.).

Taken together, these approaches provide valuable but distinct capabilities. Computational misinformation detection examines content and contextual patterns; source-credibility research evaluates the attributes of information providers; and provenance technologies document the technical history of digital assets. The proposed Islamic Information Integrity Framework (IIIF) does not claim to replace these approaches or to introduce source evaluation as an entirely unexplored concept. Instead, its theoretical contribution lies in bringing these dimensions together through an explicitly articulated ethical-epistemological framework informed by the principles of ʿAdālah, Ḍabṭ, Isnād, and Matn.

2.2. Islamic Epistemology and the Principle of Tabayyun:

In contrast to the reactive nature of modern models, Islamic epistemology approaches information verification as both an intellectual and ethical obligation. The Qur'an establishes *Tabayyun* (critical verification) (Qur'an 49:6). This injunction demonstrates that verification is a moral responsibility intended to protect social harmony and human dignity.

Unlike many modern information systems that primarily assess textual content, Islamic scholarship recognized early on that the credibility of information is inseparable from the credibility of its transmitter. This source-oriented philosophy evolved into one of the most sophisticated systems of information authentication in intellectual history: *ʿIlm al-Jarḥ wa al-Taʿdīl* (the Science of Narrator Critical Appraisal). Modern epistemologists, such as Kamali (2006) and Al-Attas (2018), document how this discipline functions not merely as a religious tool, but as a universal, logic-based data-integrity framework (Siddiqi, 1993).

2.3. The Dual-Vector Analysis and Structural Mechanics of Jarḥ wa Taʿdīl:

Viewed through the prism of modern information science, *Jarḥ wa Taʿdīl* operates as a comprehensive multidimensional matrix. Classical literature—from Ibn Abī Ḥātim al-Rāzī to Ibn Ḥajar al-ʿAsqalānī's *Nuzhat al-Naẓar*—reveals that this science evaluates data integrity through four interconnected pillars:

- First, ʿAdālah (Moral Integrity): This evaluates the ethical trustworthiness, structural truthfulness, and social accountability of the source.
- Second, Ḍabṭ (Cognitive Precision): This assesses the cognitive precision, retention capability, and long-term consistency of the transmitter.
- Third, Isnād (Transmission Chain): This establishes the continuity and authenticity of

Terbit online pada: https://ejournal.almarkazibkl.org/index.php/ince - E ISSN: 2961-9866 (Special Edition)
This article is licensed under CC BY-SA 4.0 International

the transmission pathway, acting as a rigorous provenance tracking mechanism.

- Finally, Matn Criticism: This analyzes the internal consistency and coherence of the transmitted content by comparing it with established evidence. Together, *ʿAdālah* and *Ḍabṭ* form a strict "dual-vector analysis," positing that intellectual reliability cannot exist in isolation from moral responsibility. Unlike modern models that split the data from its creator, *Jarḥ wa Taʿdīl* structurally anchors the integrity of the content (*matn*) to the credibility of the network nodes.

2.4. The Research Gap and Conceptual Positioning

A synthesis of existing literature reveals a stark paradox: the modern digital world possesses advanced technological tools but lacks an ethical, source-centric philosophy; conversely, the Islamic tradition possesses a robust, source-centric philosophy that has yet to be operationally mapped onto modern digital architectures. While recent interdisciplinary studies have explored Islamic cyber-ethics (Al-Kandari & Al-Qarni, 2012) and drawn loose analogies between *isnād* and data provenance, a profound theoretical and methodological gap remains unresolve.

Although existing studies have examined source credibility, computational misinformation detection, digital provenance, and Islamic communication ethics, an integrated conceptual framework connecting these dimensions through the classical principles of Jarḥ wa Taʿdīl remains insufficiently developed in the literature examined for this study. The research gap, therefore, does not lie in the complete absence of source-based credibility assessment. Rather, it concerns the lack of a clearly articulated interdisciplinary model that distinguishes ethical source evaluation from technical transmission verification and substantive content authentication while explaining their interrelationships and limitations. The present study addresses this conceptual gap by proposing IIIF as an Islamic epistemological framework that complements existing computational and provenance-based approaches. Its originality rests on the systematic conceptual mapping of classical principles into a four-dimensional digital credibility architecture, rather than on a claim that contemporary research has entirely neglected source credibility.

3. RESEARCH METHODOLOGY

This study adopts a qualitative, conceptual, and interdisciplinary research design to develop an Islamic Information Integrity Framework (IIIF) grounded in the principles of *ʿIlm al-Jarḥ wa al-Taʿdīl* for evaluating data credibility in cyberspace. Rather than conducting

Terbit online pada: https://ejournal.almarkazibkl.org/index.php/ince - E ISSN: 2961-9866 (Special Edition)
This article is licensed under CC BY-SA 4.0 International

empirical software testing or quantitative statistical analysis, this research constructs a robust theoretical model by integrating classical Islamic epistemology with contemporary information science. The study operates within an interpretivist paradigm, which is uniquely suited to examining the ethical, epistemological, and behavioral dimensions of online information credibility and digital interaction.

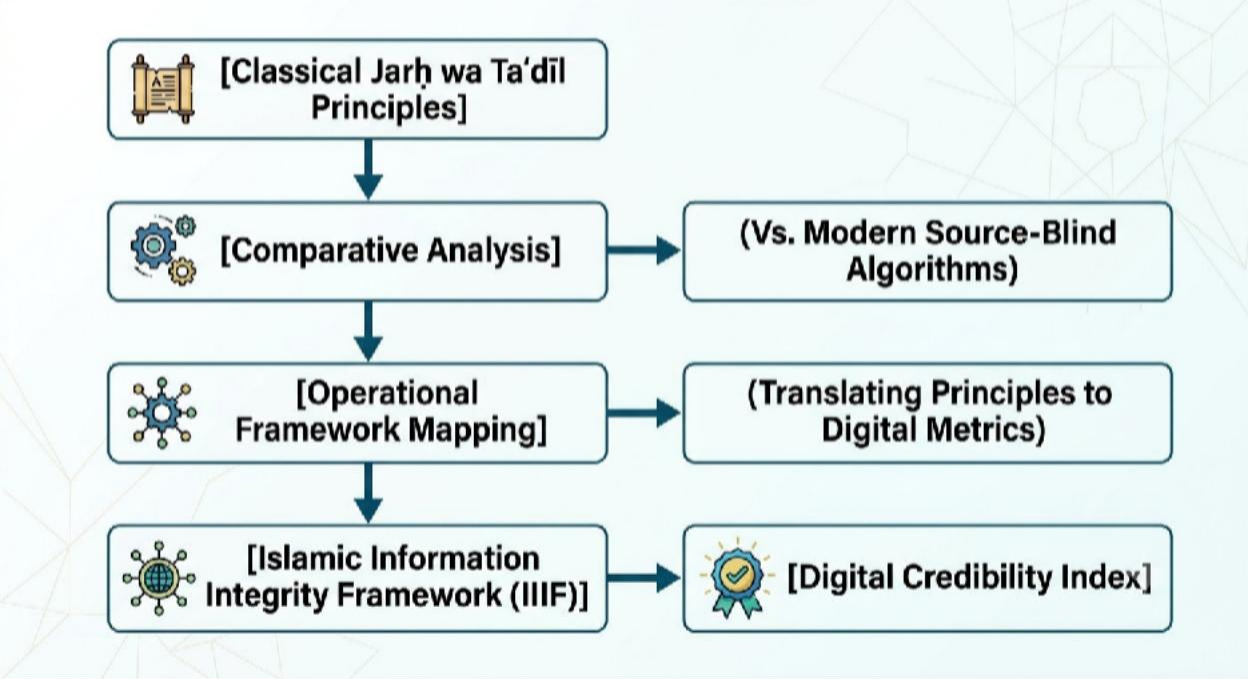


Figure 3.1: ***Conceptual Flow and Structural Progression of the Methodology.***

3.1. Data Sources and Material Selection

The documentary materials used in this conceptual study were selected purposively according to their relevance to the epistemological foundations of Jarḥ wa Taʿdīl and the contemporary challenges of digital information credibility. The selection process involved two complementary categories of sources.

Primary Islamic Sources: The primary sources included the Holy Qur'an, particularly Qur'an 49:6 concerning the principle of Tabayyun, and foundational classical works on Hadith methodology and narrator criticism. These included Ibn Abī Ḥātim al-Rāzī's Al-Jarḥ wa al-

Terbit online pada: https://ejournal.almarkazibkl.org/index.php/ince - E ISSN: 2961-9866 (Special Edition)
This article is licensed under CC BY-SA 4.0 International

Taʿdīl, Al-Khaṭīb al-Baghdādī's Al-Kifāyah fī ʿIlm al-Riwāyah, and Ibn Ḥajar al-ʿAsqalānī's Nuzhat al-Naẓar. These works were selected because they provide foundational discussions of narrator reliability, precision in transmission, continuity of chains, and the evaluation of

Terbit online pada: https://ejournal.almarkazibkl.org/index.php/ince - E ISSN: 2961-9866 (Special Edition)
This article is licensed under CC BY-SA 4.0 International

transmitted reports. Particular attention was given to the methodological distinctions between narrator assessment and report authentication.

Secondary Contemporary Literature: Contemporary sources were selected from scholarly publications addressing online source credibility, misinformation detection, artificial intelligence, digital provenance, cryptographic verification, and algorithmic accountability. Priority was given to peer-reviewed research articles, established literature reviews, and recognized technical documentation directly relevant to the four dimensions of the proposed framework. The selection was guided by conceptual relevance rather than by an attempt to conduct an exhaustive systematic review of all publications in the field.

The inclusion criteria required that a source contribute to at least one of the following areas: (i) theoretical or methodological discussions of narrator reliability and Hadith authentication; (ii) source-level or message-level digital credibility assessment; (iii) technical verification of digital content and transmission history; or (iv) ethical and epistemological approaches to information integrity. Sources without a clear connection to these research objectives were not included in the conceptual synthesis.

3.2. Comparative Analytical Framework

A comparative analytical approach was employed to examine the conceptual similarities and differences between classical Islamic narrator criticism and contemporary digital credibility assessment. The analysis was not intended to establish a historical equivalence between Hadith authentication and modern computational systems. Rather, it sought to identify transferable epistemological principles while preserving the distinctive scholarly context of classical Hadith methodology.

The comparison was conducted across four analytical dimensions: (i) the ethical reliability of the information provider, corresponding conceptually to ʿAdālah; (ii) the accuracy and integrity of information handling, corresponding conceptually to Ḍabṭ; (iii) the traceability and continuity of information transmission, corresponding conceptually to Isnād; and (iv) the examination of informational claims against relevant evidence, corresponding conceptually to Matn criticism.

Terbit online pada: https://ejournal.almarkazibkl.org/index.php/ince - E ISSN: 2961-9866 (Special Edition)
This article is licensed under CC BY-SA 4.0 International

3.3. Operational Mapping and Framework Development

The core methodological contribution of this paper lies in framework mapping to operationalize abstract, qualitative Islamic concepts into tangible, measurable digital credibility indicators. Each classical pillar is systematically translated into a corresponding digital evaluation parameter:

Table 3.1 Conceptual Mapping of Classical Principles to Digital Credibility Indicators

Classical Principle	Proposed Digital Indicators	Scope and Limitation
ʿAdālah (Moral Integrity)	Source attribution, verifiable institutional affiliation, disclosure of conflicts of interest, documented correction practices, and evidence of responsible communication.	Identity verification and reputation provide contextual evidence but do not guarantee the truthfulness of an individual claim.
Ḍabṭ (Precision and Accuracy)	File integrity checks, metadata consistency, technical error detection, preservation of original records, and comparison with authenticated versions.	Technical integrity can indicate whether data have been altered or corrupted, but it cannot independently establish factual accuracy.
Isnād (Transmission	Digital provenance records, documented transmission history, source attribution across transfers, timestamps, and	Traceability establishes aspects of origin and transmission history but does not prove the

Terbit online pada: https://ejournal.almarkazibkl.org/index.php/ince - E ISSN: 2961-9866 (Special Edition)
This article is licensed under CC BY-SA 4.0 International

Chain)	cryptographic signatures.	truthfulness of the original information.
Matn (Content Evaluation)	Cross-source verification, consistency with established facts, and expert review where necessary.	corroboration, contextual analysis, evidence-based claim	Semantic coherence alone cannot establish truth; verification requires relevant evidence and contextual judgment.

Terbit online pada: https://ejournal.almarkazibkl.org/index.php/ince - E ISSN: 2961-9866 (Special Edition)
This article is licensed under CC BY-SA 4.0 International

The proposed mapping should be understood as a conceptual and functional correspondence rather than a claim of methodological equivalence between classical Hadith scholarship and digital technologies. Each dimension addresses a distinct aspect of information credibility, and no single indicator is sufficient to establish the overall reliability of a digital claim. Source identity, technical integrity, transmission provenance, and content verification provide complementary forms of evidence. Their integration allows the framework to distinguish the credibility of an information provider from the factual reliability of a specific message. Consequently, the assessment of a source should not automatically determine the truth or falsity of every item it communicates, and the absence of verifiable identity or provenance should be treated as an evidentiary limitation rather than conclusive proof of fabrication.

3.3.1. Analytical Procedure for Deriving Digital Credibility Indicators

The development of the proposed digital indicators followed a structured conceptual procedure consisting of four stages. First, the relevant discussions in classical Hadith methodology were examined to identify the epistemological function of each principle. The analysis focused on the distinction between moral reliability (ʿAdālah), precision and accuracy in transmission (Ḍabṭ), continuity and traceability of transmission (Isnād), and the examination of transmitted content against established evidence (Matn).

Second, the identified functions were compared with corresponding problems in contemporary digital environments. These included questions of source identity and accountability, data alteration and technical reliability, information provenance, and the verification of factual claims.

Third, conceptual correspondences were identified without assuming that classical principles and modern technical mechanisms were identical. Each proposed digital indicator was therefore treated as an analytical analogy informed by an Islamic epistemological principle rather than as a direct technical equivalent.

Fourth, the resulting indicators were organized into a four-dimensional framework. Their definitions, potential evidence sources, and methodological limitations were considered to clarify what each dimension could establish and what remained outside its scope.

3.4. Formulation of the Digital Credibility Taxonomy

Terbit online pada: https://ejournal.almarkazibkl.org/index.php/ince - E ISSN: 2961-9866 (Special Edition)
This article is licensed under CC BY-SA 4.0 International

Drawing from the classical taxonomic ranking of narrators (e.g., *Thiqah* [trustworthy], *Ṣadūq* [truthful], *Ḍaʿīf* [weak], *Mawḍūʿ* [fabricated]), this study formulates a theoretical Digital

Credibility Taxonomy. This framework provides a standardized grading scale to categorize digital nodes, social media accounts, and information outlets based on their aggregated credibility score across the four IIIF dimensions.

3.5. Analytical Strategy and Logical Flow

The data and conceptual schemas are processed through mixed inductive-deductive reasoning across three operational levels:

The analysis followed an inductive-deductive process, through which relevant principles were extracted from classical sources, compared with contemporary digital challenges, and synthesized into the proposed IIIF framework and its four assessment dimensions.

3.6. Scope and Delimitations

This research functions strictly as a theoretical blueprint and conceptual framework. While it defines the exact parameters, variables, and taxonomic structures required to evaluate digital credibility, the actual computational implementation (such as writing machine learning algorithms or coding software) falls outside the immediate scope of this paper. The framework is intentionally designed to be modular and scalable, providing the epistemological architecture necessary for future interdisciplinary collaboration between Islamic studies scholars, information scientists, and AI researchers.

4. RESULTS AND DISCUSSION

The principal result of this conceptual investigation is the successful development of a structured Islamic Information Integrity Framework (IIIF). Unlike empirical studies where results emerge from statistical analysis, the primary outcome of this research is a validated conceptual architecture that operationalizes the principles of Jarḥ wa Taʿdīl into measurable digital credibility indicators. The framework produces three interconnected outputs: (i) the Islamic Information Integrity Framework (IIIF), (ii) the Digital Credibility Index (DCI), and (iii) the Digital Credibility Taxonomy. These outputs collectively constitute the principal findings of the present investigation.

Terbit online pada: https://ejournal.almarkazibkl.org/index.php/ince - E ISSN: 2961-9866 (Special Edition)
This article is licensed under CC BY-SA 4.0 International

4.1. The Islamic Information Integrity Framework (IIIF)

A Source-Centric Paradigm Shift The principal outcome of this study is the development of the Islamic Information Integrity Framework (IIIF), a comprehensive conceptual model that addresses the epistemological vacuum in contemporary digital credibility assessment. Unlike existing computational systems that primarily evaluate the content of information after its dissemination, the IIIF introduces a proactive, ethically grounded, and source-centric verification architecture.

The cornerstone of the IIIF lies in its "Dual-Vector Architecture," which structurally eliminates the "source-blindness" prevalent in modern AI and machine learning models. Rather than treating data as an isolated entity, the IIIF mandates that credibility must be evaluated by simultaneously scoring both the digital node (the transmitter) and the data packet (the transmission). This dual-vector approach integrates.

An important distinction within the proposed architecture is the separation between source-level credibility and claim-level credibility. A digital account may possess a verifiable identity and a history of responsible communication, yet a particular statement published by that account may still be inaccurate. Conversely, an anonymous source may provide information that can be independently corroborated through reliable evidence. Therefore, IIIF does not assume that the credibility of a source automatically determines the credibility of its individual messages. Instead, it proposes two interconnected but analytically distinct levels of assessment: the reliability and accountability of the information provider, and the evidentiary credibility of the specific information being transmitted. This distinction is essential for avoiding unjustified generalizations and preserving the epistemological balance of the framework.

4.2. Operational Architecture and the Digital Credibility Index (DCI)

Moving beyond abstract theoretical concepts, the IIIF operationalizes the classical mechanics of *Jarḥ wa Taʿdīl* into a scalable Digital Credibility Index (DCI). The operational process functions as a sequential, multi-layered assessment: beginning with Source Authentication (ʿAdālah), proceeding to Technical Integrity (Ḍabṭ), followed by Digital Provenance (Isnād), and culminating in Content Evaluation (Matn).

Figure 4.1 illustrates the overall architecture of the proposed Islamic Information Integrity

Terbit online pada: https://ejournal.almarkazibkl.org/index.php/ince - E ISSN: 2961-9866 (Special Edition)
This article is licensed under CC BY-SA 4.0 International

Framework (IIIF), showing how the four classical principles interact to generate the Digital Credibility Index (DCI). To quantify this process, the study formulates a Digital Credibility Taxonomy, mirroring the classical classification of narrators, to categorize digital nodes and information outlets.

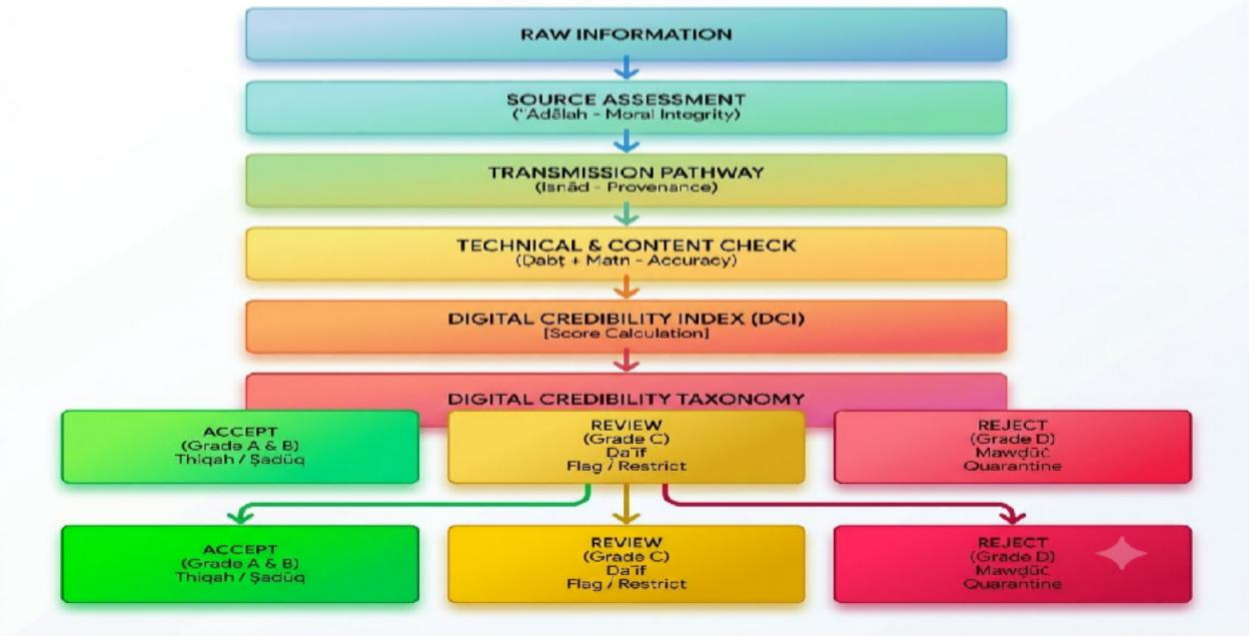


Figure 4.1: The overall architecture of the Islamic Information Integrity Framework (IIIF)

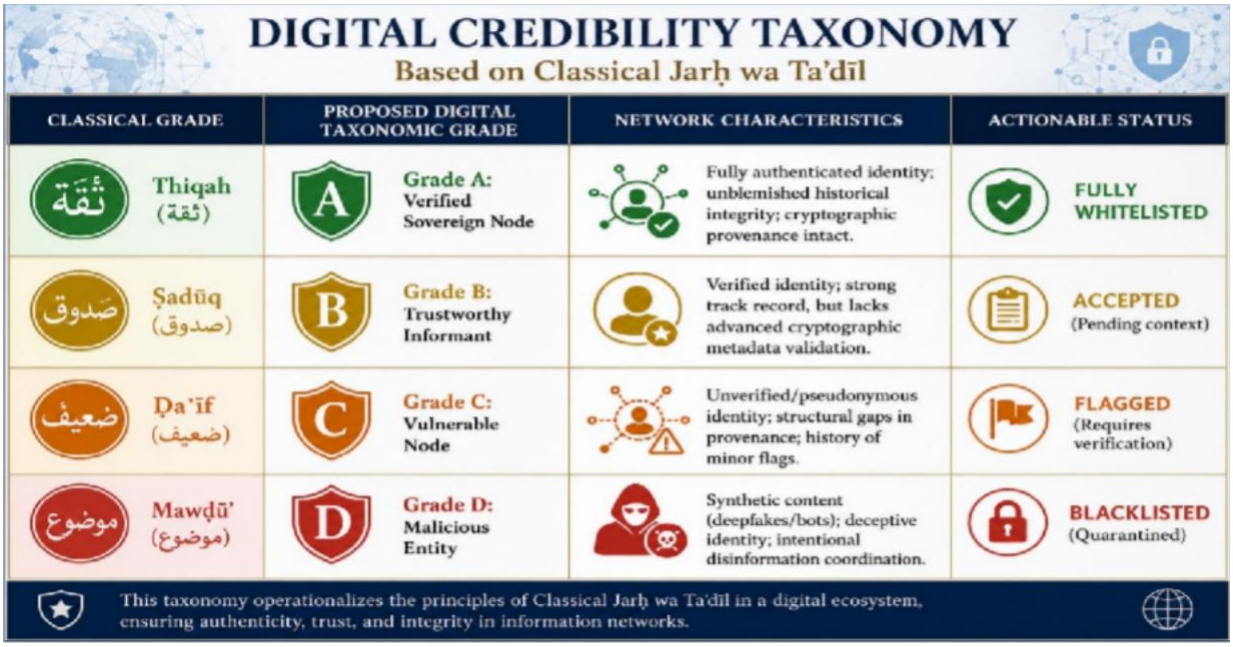
As depicted in Figure 4.1, the operationalization of the framework is not a set of isolated checks but a sequential, multi-layered assessment process. The flow initiates with **Source Authentication (ʿAdālah)** to establish the baseline integrity of the digital node, which is then cross-examined through **Technical Integrity (Ḍabṭ)** to ensure data accuracy. This is further verified via **Digital Provenance (Isnād)** to trace the complete chain of transmission.

To operationalize the Digital Credibility Index (DCI), the proposed framework classifies digital actors into distinct credibility categories derived from the classical science of Jarḥ wa Taʿdīl. Each category reflects a specific level of epistemic reliability based on source authenticity, provenance integrity, and historical trustworthiness, while also defining the appropriate verification status within the digital ecosystem. The resulting taxonomy is

Terbit online pada: https://ejournal.almarkazibkl.org/index.php/ince - E ISSN: 2961-9866 (Special Edition)
This article is licensed under CC BY-SA 4.0 International

presented in Table 4.1.

Table 4.1. Digital Credibility Taxonomy Based on Classical Jarḥ wa Taʿdīl



As shown in Table 4.1, the proposed taxonomy translates the classical hierarchy of narrator evaluation into an operational framework for digital credibility assessment. Rather than functioning as a simple descriptive classification, each grade determines how a digital node should be treated within the verification process, ranging from full acceptance to complete exclusion. This taxonomy therefore provides the conceptual foundation upon which the Digital Credibility Index (DCI) and its subsequent mathematical formulation are constructed.

A critical consideration in the proposed framework concerns the assessment of anonymous and pseudonymous digital sources. Although the concept of al-Majhūl in classical Hadith scholarship offers a useful point of conceptual reflection, it should not be treated as directly equivalent to digital anonymity. In classical narrator criticism, the assessment of an unknown narrator involves specific methodological considerations that

Terbit online pada: https://ejournal.almarkazibkl.org/index.php/ince - E ISSN: 2961-9866 (Special Edition)
This article is licensed under CC BY-SA 4.0 International

cannot be transferred mechanically to contemporary online identities.

Accordingly, IIIF does not automatically classify anonymous digital actors as unreliable or fabricated. Instead, anonymity is treated as a limitation in source identification and accountability that requires additional scrutiny. The information communicated by such sources may nevertheless be assessed independently through available provenance records, documentary evidence, cross-source corroboration, and contextual verification. This distinction is particularly important in cases involving whistleblowers, vulnerable individuals, investigative journalism, and legitimate privacy protection.

The framework therefore distinguishes between the credibility status of a source and the evidentiary status of its individual claims. An anonymous source may remain unidentified while a particular claim is independently corroborated by reliable evidence. Conversely, an identifiable and reputable source may communicate inaccurate information. This approach seeks to preserve the principles of responsible verification without equating anonymity with dishonesty or treating digital reputation as conclusive evidence of truthfulness.

To clarify the mathematical structure of the proposed Digital Credibility Index (DCI), this study distinguishes between an integrated assessment score and a non-compensatory verification safeguard. The four dimensions of IIIF—ʿAdālah, Ḍabṭ, Isnād, and Matn—provide complementary evidence for evaluating digital information credibility. However, a deficiency in one dimension should not automatically invalidate the entire information item, particularly when the relevant evidence is unavailable rather than demonstrably negative.

For the conceptual illustration of DCI, let A, D, I, and M represent the assessment scores of the four dimensions, respectively, on a proposed scale from 0 to 100. A provisional equal-weight formulation may be expressed as:

$$DCI = (A + D + I + M) / 4$$

This equal-weight formulation is used solely to illustrate the proposed integration of the four dimensions. It does not represent an empirically established weighting system, and the resulting score should not be interpreted as a validated measurement of truthfulness or source reliability.

A separate verification safeguard is also required. Where direct and sufficiently reliable evidence establishes that a specific digital item has been fabricated or materially manipulated, that item should be flagged for rejection or further investigation regardless of

Terbit online pada: https://ejournal.almarkazibkl.org/index.php/ince - E ISSN: 2961-9866 (Special Edition)
This article is licensed under CC BY-SA 4.0 International

its aggregate score. This safeguard is based on an evidenced finding concerning the specific item rather than an automatic assumption about the moral character of its source.

Furthermore, unavailable information must be distinguished from negative evidence. An anonymous source, for example, should not automatically receive a zero score for ʿAdālah, nor should an incomplete transmission record automatically establish fabrication. Such casea should be marked as unverified or insufficiently documented, with the relevant limitations explicitly reported.

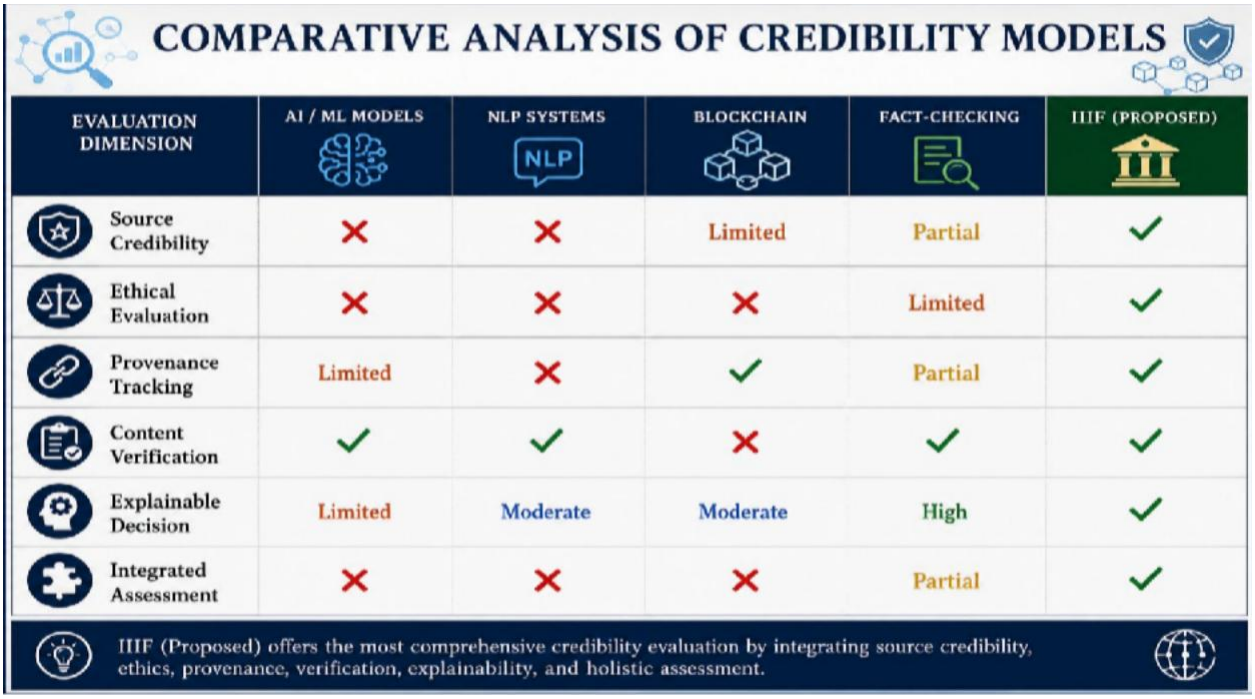
Accordingly, the proposed DCI is presented as a preliminary conceptual assessment structure rather than a validated quantitative instrument. Its weighting scheme, scoring criteria, and decision thresholds require further expert evaluation, operational definition, and empirical testing before any standardized classification can be established.

4.3. Comparative Evaluation with Contemporary Computational Models

To demonstrate the original value of the IIIF, it is essential to position it against the dominant computational approaches currently employed in cyberspace. Machine learning and NLP algorithms primarily identify misinformation through statistical analysis of linguistic features or visual anomalies. While computationally efficient, they struggle with newly emerging manipulation strategies and completely detach the information from the ethical identity of its producer. Blockchain-based provenance systems secure data lineage immutably; however, blockchain can authenticate the continuity of information *after* registration but cannot independently determine whether the original source was truthful *before* entering the ledger.

Table 4.2. Comparative Analysis of Credibility Models

Terbit online pada: https://ejournal.almarkazibkl.org/index.php/ince - E ISSN: 2961-9866 (Special Edition)
This article is licensed under CC BY-SA 4.0 International



4.4. Practical Applicability in Contemporary Digital Ecosystems

4.4.1. Illustrative Case Analysis: Assessing a Hypothetical Viral Digital Announcement

To demonstrate how the proposed IIIF may operate in practice, consider a hypothetical situation in which an anonymous social media account circulates a video claiming that a major public institution has announced the closure of all educational institutions in a particular region due to an emergency. The video rapidly spreads across social media platforms, generating public concern. Rather than immediately accepting or rejecting the announcement, IIIF proposes a structured assessment through its four interconnected dimensions.

Stage 1: Source Assessment (ʿAdālah)

The first stage examines the available information about the account publishing the announcement. Investigators consider whether the account has a verifiable institutional affiliation, an identifiable publication history, or any documented relationship with an authoritative source. In this hypothetical case, the account is anonymous and no institutional authorization can initially be established. Consequently, its source identity remains unverified. However, anonymity alone is not treated as evidence that the announcement is false.

Stage 2: Technical Integrity Assessment (Ḍabṭ)

Terbit online pada: https://ejournal.almarkazibkl.org/index.php/ince - E ISSN: 2961-9866 (Special Edition)
This article is licensed under CC BY-SA 4.0 International

The second stage examines the technical characteristics of the video. Available metadata, file properties, editing traces, and comparisons with an original or authenticated version are considered. Where technically feasible, appropriate digital forensic and media verification tools may be employed. The objective is to identify possible alteration, inconsistencies, or evidence of manipulation. A technically intact file, however, is not automatically considered factually accurate.

Stage 3: Transmission Provenance Assessment (Isnād)

The third stage investigates the available transmission history of the video. Researchers examine its earliest identifiable publication, available timestamps, documented transfers, and any verifiable digital provenance records. If the original source and transmission history cannot be established, the provenance dimension remains insufficiently verified. Cryptographic records, where available, may support the integrity of the recorded transmission history but cannot independently authenticate the truth of the announcement.

Stage 4: Content Verification (Matn)

The fourth stage examines the substantive claim contained in the announcement. The video is compared with official institutional notifications, authenticated statements, independent documentary evidence, and reliable reports. In this hypothetical example, the relevant institution publicly confirms that no such closure announcement was issued and provides an authenticated notification contradicting the viral video. This independent evidence provides grounds for rejecting the specific claim, irrespective of the anonymous identity of the account that circulated it.

Integrated Assessment and Interpretation

The four stages demonstrate that the credibility of the source and the factual status of the information must be assessed separately. In this example, the source remains unidentified, the transmission history is incomplete, and the content is contradicted by authoritative evidence. If reliable forensic examination also establishes that the video was materially manipulated, the specific digital item may be classified as fabricated or manipulated on the basis of that evidence.

This hypothetical illustration demonstrates the conceptual applicability of IIIF as an organized verification procedure. It does not constitute an empirical test, establish a numerical performance advantage over existing verification technologies, or demonstrate that the framework can independently detect all forms of misinformation. Rather, it shows

Terbit online pada: https://ejournal.almarkazibkl.org/index.php/ince - E ISSN: 2961-9866 (Special Edition)
This article is licensed under CC BY-SA 4.0 International

how source assessment, technical examination, provenance analysis, and content verification can be coordinated within a single interdisciplinary assessment structure.

4.5. Ethical Risks, Limitations, and Safeguards

Although IIIF offers a conceptual approach to integrating ethical and technical dimensions of digital credibility, its potential application also raises important ethical and methodological concerns.

First, the assessment of anonymous and pseudonymous sources requires particular caution. In investigative journalism, human rights reporting, and whistleblowing, anonymity may serve legitimate purposes, including the protection of personal safety and freedom of expression. Therefore, the absence of an identifiable account should not automatically result in the rejection of information or an adverse judgment about its truthfulness. Independent corroboration and available documentary evidence should remain central to claim-level assessment.

Second, the collection and evaluation of digital identity information may create privacy risks. Any future implementation of the framework should follow principles of data minimization, informed and lawful processing, purpose limitation, secure storage, and restricted access to personal information. Sensitive personal data should not be collected merely to generate generalized credibility profiles.

Third, digital reputation assessments may reproduce existing social, institutional, linguistic, or political biases. Established accounts and powerful organizations may possess greater visibility and verification resources than marginalized or newly established sources. Consequently, reputation should never be treated as an unquestionable substitute for evidence. Transparent assessment criteria, contextual interpretation, human oversight, and opportunities for correction or appeal would be necessary safeguards against unfair classification.

Finally, the framework should distinguish between an unreliable source, an unverified claim, and demonstrably fabricated information. These categories represent different evidentiary conditions and should not be treated as interchangeable. Any future technical implementation would require interdisciplinary review, including Islamic scholarship, digital forensics, information ethics, and data protection expertise. These safeguards are proposed as normative requirements for the future development of IIIF. Their effectiveness has not yet been empirically evaluated within the present study.

Terbit online pada: https://ejournal.almarkazibkl.org/index.php/ince - E ISSN: 2961-9866 (Special Edition)
This article is licensed under CC BY-SA 4.0 International

5.CONCLUSION

This study proposed the Islamic Information Integrity Framework (IIIF) as a conceptual approach to evaluating digital information credibility through the epistemological principles of ʿIlm al-Jarḥ wa al-Taʿdīl and the Qur'anic principle of Tabayyun. The study explored how the classical concepts of ʿAdālah, Ḍabṭ, Isnād, and Matn may inform contemporary discussions of source accountability, technical data integrity, information provenance, and content verification.

The principal theoretical contribution of the study lies in its proposed integration of these four dimensions within a unified interdisciplinary framework. Rather than claiming that existing digital credibility systems entirely neglect source reliability, IIIF seeks to complement contemporary approaches by explicitly connecting ethical source assessment with technical transmission analysis and evidence-based content verification. The framework also emphasizes an important distinction between source-level credibility and claim-level credibility: the reliability of an information provider does not automatically establish the truth of every message, just as the anonymity of a source does not necessarily invalidate the information communicated.

The proposed Digital Credibility Index (DCI) represents a preliminary conceptual attempt to organize the four dimensions into an integrated assessment structure. However, its mathematical formulation, weighting assumptions, scoring criteria, and classification procedures remain theoretical. They have not yet undergone expert consensus assessment, statistical validation, or technical testing. Therefore, DCI should not currently be regarded as a standardized measurement instrument or a proven mechanism for determining the truthfulness of digital information.

Similarly, the illustrative case analysis demonstrates the conceptual sequence through which the four dimensions may be applied to a hypothetical misinformation scenario. It does not establish the empirical effectiveness of IIIF or demonstrate superior performance compared with existing computational verification systems. Its purpose is to clarify the proposed analytical procedure and identify the complementary roles of source assessment, technical integrity, provenance verification, and content evaluation.

The study also recognizes important ethical limitations concerning anonymous sources, digital privacy, institutional bias, and the potential misuse of reputation-based assessments. Any future application of IIIF should incorporate transparent assessment criteria, data

Terbit online pada: https://ejournal.almarkazibkl.org/index.php/ince - E ISSN: 2961-9866 (Special Edition)
This article is licensed under CC BY-SA 4.0 International

protection safeguards, human oversight, and appropriate mechanisms for correction and appeal. In particular, anonymity, incomplete provenance, and insufficient evidence should not automatically be interpreted as proof of dishonesty or fabrication.

Future research should therefore proceed through several stages. First, the proposed indicators should be examined by interdisciplinary experts in Islamic Hadith scholarship, information science, digital forensics, and information ethics. Second, the conceptual model should be operationalized through clearly defined assessment procedures and tested against appropriately selected digital information cases. Third, empirical studies should evaluate the consistency, reliability, fairness, and practical limitations of the proposed DCI. Finally, technical implementation may explore the integration of IIIF with existing misinformation detection, content provenance, and artificial intelligence governance systems.

In conclusion, IIIF is presented as a proposed Islamic epistemological framework for interdisciplinary research on digital information integrity. Its significance lies in offering a structured conceptual dialogue between classical Islamic principles of responsible transmission and contemporary challenges of digital credibility. Its practical effectiveness, scalability, and technical suitability, however, remain questions for future expert assessment and empirical validation.

REFERENCES:

Al-Attas, S. M. N. (2018). On Islam and the concept of knowledge. Istac Publications.

Al-Kandari, A., & Al-Qarni, A. (2012). The impact of social media on the dissemination of Islamic content. Journal of Media and Religious Studies, 4(2), 87-104.

Al-Khaṭīb al-Baghdādī, A. b. A. (1986). Al-Kifāyah fī ʿilm al-riwāyah [Sufficiency in the science of transmission]. Dār al-Kutub al-ʿIlmiyyah.

C2PA. (n.d.). *C2PA specifications: Explainer*. Coalition for Content Provenance and Authenticity. https://c2pa.org/specifications/specifications/2.2/explainer/Explainer.html

Fallis, D. (2009). Toward an epistemology of information. The Information Society, 25(3), 175-185. https://doi.org/10.1080/01972240902848988

Terbit online pada: https://ejournal.almarkazibkl.org/index.php/ince - E ISSN: 2961-9866 (Special Edition)
This article is licensed under CC BY-SA 4.0 International

Fraga-Lamas, P., & Fernández-Caramés, T. M. (2019). Towards the blockchain of things: A survey on blockchain-based IoT architectures. IEEE Access, 7, 167066-167085. https://doi.org/10.1109/ACCESS.2019.2953216

Graham, T. (2020). The epistemology of fact-checking: A normative framework for evaluating digital information. Critical Review, 32(1), 1-22. https://doi.org/10.1080/08913811.2020.1723510

Hanimann, A., Heimann, A., Hellmueller, L., & Trilling, D. (2023). Believing in credibility measures: Reviewing credibility measures in media research from 1951 to 2018. *International Journal of Communication, 17*.

Ibn Abī Ḥātim al-Rāzī, ʿA. b. M. (n.d.). Al-Jarḥ wa al-taʿdīl [The science of narrator criticism and authentication]. Dār al-Kutub al-ʿIlmiyyah.

Ibn Ḥajar al-ʿAsqalānī, A. b. A. (2003). Nuzhat al-naẓar fī tawḍīḥ nukhbat al-fikar [The excursion of the sight in explaining the essence of thought]. Dār Ibn al-Jawzī.

Kamali, M. H. (2006). Principles of Islamic jurisprudence (3rd ed.). Islamic Texts Society.

Metzger, M. J. (2007). Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research. *Journal of the American Society for Information Science and Technology, 58*(13), 2078–2091. https://doi.org/10.1002/asi.20672

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter, 19(1), 22-36. https://doi.org/10.1145/3127518.3127524

Siddiqi, M. Z. (1993). Hadith literature: Its origin, development & special features (Revised ed.). Islamic Foundation.

Sundar, S. S. (2020). Automation and the human touch in the age of algorithms. Journal of Computer-Mediated Communication, 25(1), 1-6. https://doi.org/10.1093/jcmc/zmz029

Wardle, C., & Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policy making. Council of Europe report. https://doi.org/10.2760/733748

Zhou, X., & Zafarani, R. (2020). A survey of fake news: Fundamental theories, detection methods, and opportunities. ACM Computing Surveys (CSUR), 53(5), 1-40.



Terbit online pada: https://ejournal.almarkazibkl.org/index.php/ince - E ISSN: 2961-9866 (Special Edition)
This article is licensed under CC BY-SA 4.0 International
https://doi.org/10.1145/3395046

World Economic Forum. (2024). *The Global Risks Report 2024* (19th ed.). World Economic Forum. https://www.weforum.org/reports/global-risks-report-2024