Jurnal El-umdah
Jurnal Kajian Ilmu Al-Qur'an dan Tafsir
doi: 10.26811/xxxx.xxxx.xxxx

Reconstructing and Developing the Paradigm of Qur'anic Studies in the Digital Era

Lalu Abdurrahman
Universitas islam Negeri Mataram

[1]*Correspondence Email:* rahmanlaluabdur593@gmail.com

Received: 2026-10-29	**Accepted:** 2026-11-05	**Published:** 2026-11-16

Abstract

Digital technologies now mediate how the Qur'an is edited, analysed, taught and interpreted, yet scholarship on these developments remains divided between manuscript philology, corpus linguistics, digital pedagogy and artificial intelligence, with no shared account of what the discipline itself should become. This article aims to reconstruct the paradigm of Qur'anic studies for the digital era and to propose principles for its development. It combines a conceptual-hermeneutic analysis, drawing on Kuhn's account of paradigms and Gadamer's philosophical hermeneutics, with a documentary analysis of four digital infrastructures: a philological database, an annotated linguistic corpus, social-media and online learning practices, and generative language models used for tafsir. The findings show that each infrastructure reshapes the object, method, authority, community and ethics of Qur'anic studies in a different way, and that the sharpest anomaly lies in source attribution, where language models reproduce interpretations fluently while failing to preserve chains of transmission. The discussion proposes a provenance-first paradigm that treats the classical discipline of isnad as the epistemic ancestor of data provenance, and draws implications for curricula, peer review and open annotation. The study is limited by its reliance on published documentation and reported evaluations.

Keywords *Qur'anic studies; paradigm reconstruction; digital hermeneutics; isnād; source attribution*
:

A. Introduction

Qur'anic studies have always been shaped by the technologies through which the text is transmitted, from oral recitation and manuscript copying to print. The digital turn is the latest such transformation, but it differs in one respect: it changes not only how the text circulates but also who can analyse it, at what scale, and under what claims of authority. Research on new media in the Muslim world has long examined how communication technologies reconfigure religious authority and the public sphere,[1] and a volume on the digital humanities and Islamic and Middle East studies has placed the same questions on the wider agenda of the field.[2] Within Qur'anic studies in particular, it has been argued that digital tools alter access to, and comprehension of, early manuscripts, and that a digital paradigm can open a new epistemological approach by hosting a plurality of hypotheses on a shared platform.[3] In Indonesia, the Corpus Coranicum project has been proposed as a methodological model for manuscript studies,[4] while annotated corpora are described as adequate for monolingual corpus-based research even though parallel Qur'anic corpora are still scarce.[5]

[1] Dale F. Eickelman and Jon W. Anderson, eds., *New Media in the Muslim World: The Emerging Public Sphere*, 2nd ed. (Bloomington: Indiana University Press, 2003).

[2] Elias Muhanna, ed., *The Digital Humanities and Islamic & Middle East Studies* (Berlin: De Gruyter, 2016).

[3] Alba Fedeli, "Digital Humanities and Qur'ānic Manuscript Studies: New Perspectives and Challenges for Collaborative Spaces and Plural Views," *Journal of College of Sharia and Islamic Studies* 38, no. 1 (2020): 147–58, https://doi.org/10.29117/jcsis.2020.0260.

[4] Syahrullah et al., "Corpus Coranicum and Digital Philology: A Methodological Model for Advancing Qur'anic Manuscript Studies in Indonesia," *Khazanah Theologia* 6, no. 3 (2024): 201–22.

[5] Nur Hizbullah and Abdul Mutaali, "Quranic Corpus Models for Corpus-Based Studies," in *Proceedings of the Eleventh Conference on Applied Linguistics (CONAPLIN 2018)* (Atlantis Press, 2019), https://doi.org/10.2991/conaplin-18.2019.270.

A second line of work looks at lived digital practice. A study of Qur'anic commentary on Indonesian Facebook speaks of a new face of tafsir,[6] and research on digital Qur'an learning points to a shift in the isnād system and the consolidation of new forms of authority.[7] Conceptual proposals for a digital hermeneutics of the Qur'an have sought a holistic integration of faith and technology, though largely through literature review rather than through the analysis of concrete infrastructures.[8]

A third line of work concerns generative artificial intelligence, which is now consulted for religious guidance. An evaluation of seven mainstream language models on the interpretation of the phrase fī sabīlillāh in Q 9:60 found that the models generally identified the relevant exegetes but performed poorly in preserving isnād, narration structure, citation accuracy and attribution.[9] A benchmark in Bengali reports that most of the open-weight models tested confidently fabricate Islamic sources,[10] and shared tasks such as IslamicEval 2025 now treat the detection and correction of misquoted verses and hadith as a research problem in its own right.[11]

---

[6] F. Lukman, "Digital Hermeneutics and a New Face of the Qur'an Commentary: The Qur'an in Indonesian's Facebook," *Al-Jāmiʿah: Journal of Islamic Studies* 56, no. 1 (2018): 95–120, https://doi.org/10.14421/ajis.2018.561.95-120.

[7] A. Majit and M. Miski, "Pembelajaran al-Qur'an secara digital: Pergeseran sistem isnad dan peneguhan otoritas baru," *Jurnal SMART (Studi Masyarakat, Religi, dan Tradisi)* 9, no. 1 (2023): 133–46, https://doi.org/10.18784/smart.v9i1.1795.

[8] Alfian Dhany Misbakhuddin, "Digital Hermeneutics of the Qur'an: Towards a Holistic Model of Faith-Technology Integration," ICOERESS, n.d., https://icoeress.pascasarjana.uinjambi.ac.id/ojs/index.php/icoeress/article/view/82.

[9] Saifullah, Soleh Hasan Wahid, and Muhammad Ihza Fazrian, "Between Machines and the Tafsīr Tradition: Reorienting Tafsīr bi al-Ma'ṡūr in the Age of Computational Interpretation," *Jurnal Studi Ilmu-ilmu Al-Qur'an dan Hadis* 27, no. 2 (2026): 458–93, https://doi.org/10.14421/qh.v27i1.7499.

[10] "Do Open Language Models Invent Scripture? BILAL-BENCH: Benchmarking Islamic-Source Fabrication in Bengali," NeurIPS 2026, https://neurips.cc/virtual/2026/166644.

[11] Hamdy Mubarak et al., "IslamicEval 2025: The First Shared Task of Capturing LLMs Hallucination in Islamic Content," in *Proceedings of the Third Arabic Natural Language Processing Conference: Shared Tasks* (Suzhou, China: Association for Computational Linguistics, 2025), 480–93.

These strands are rarely read together. Each answers a local question, namely how to edit manuscripts, how to annotate grammar, how to describe social-media commentary, or how to evaluate a chatbot, and none asks what the discipline of Qur'anic studies becomes when all of these infrastructures coexist. This is a question about paradigm. In Kuhn's sense, a paradigm is a set of shared exemplars, methods and commitments that defines normal research, and it is revised when anomalies accumulate that the existing framework cannot absorb.[12] Gadamer's hermeneutics adds that understanding always takes place within an effective history, so that the instruments through which a text is mediated belong to the interpreter's own horizon and cannot be treated as neutral.[13] Read together, the two perspectives suggest that digital infrastructures are not merely new tools for an unchanged discipline but new conditions of interpretation that the discipline must theorise.

This article therefore asks how the paradigm of Qur'anic studies can be reconstructed so that it accommodates digital infrastructures without surrendering the accountability that the interpretive tradition developed, and which principles should guide its further development. Its argument, and the novelty it claims, is that the decisive anomaly is not technological capacity but attribution: the classical concern with isnād and with traceable sources offers a ready resource for formulating what digital Qur'anic scholarship must preserve. The contribution is a five-component model of the reconstructed paradigm and a provenance-first principle, derived from an integrated reading of four infrastructures that are seldom analysed together. The implications concern curricula, peer review and open annotation, and

[12] Thomas S. Kuhn, *The Structure of Scientific Revolutions* (Chicago: University of Chicago Press, 1962).

[13] Hans-Georg Gadamer, *Truth and Method*, 2nd rev. ed., trans. Joel Weinsheimer and Donald G. Marshall (London: Continuum, 2004).

they extend beyond the Muslim context to any scriptural tradition confronting machine-mediated interpretation.

B. Method

This study adopts a qualitative, interpretive design combining two strands of analysis. The first is a conceptual-hermeneutic analysis, which clarifies what a paradigm of Qur'anic studies comprises and what criteria a reconstruction must meet. The second is a documentary analysis of four digital infrastructures, which supplies the empirical material against which the concept is tested. This combination is appropriate because the research question concerns the configuration of a whole field, which cannot be reduced to a single measurable variable, and because it matches the multidisciplinary orientation of the journal by joining Islamic studies, digital humanities, computational linguistics and the study of artificial intelligence.

For the conceptual strand, the study operationalises the Kuhnian notion of a paradigm as a disciplinary matrix with five analytic components: the object of study, the method, the basis of authority and transmission, the community and its access, and the ethics and verification of claims. These components are the author's heuristic and are justified by Kuhn's emphasis on shared exemplars, methods and commitments, and by Gadamer's insistence that the mediation of a text is part of the interpretive situation.

For the documentary strand, four cases were selected purposively because they occupy different points on the continuum from textual edition to interpretation: the Corpus Coranicum database as a philological infrastructure, the Quranic Arabic Corpus as an annotated linguistic resource, social-media and online learning practices as sites of lay and pedagogical commentary, and generative language models as automated interpreters. The material consists of project documentation and of peer-reviewed or preprint studies published up to October 2026, located through searches in scholarly databases and journal repositories using terms such as digital Qur'anic studies, Corpus Coranicum, Quranic corpus, hallucination

and language-model tafsir. Sources were retained when they bore directly on one of the four cases, reported their methods transparently, and offered verifiable bibliographic details. Preprints and conference pages are treated as provisional evidence and are identified as such in the notes.

The analysis proceeded in four steps. First, descriptive claims and reported figures were extracted for each case and dated, because digital infrastructures change quickly. Second, each case was coded against the five components, noting whether the classical-print paradigm is preserved, transformed or strained. Third, practices that the earlier paradigm could not accommodate were identified as anomalies. Fourth, principles for a reconstructed paradigm were derived abductively and tested for coherence against Gadamer's account of understanding. Trustworthiness was pursued by triangulating project documentation with independent critique, as in the case of Corpus Coranicum, and by reporting figures with their dates. The study did not test any system independently, and the coding was done by the author, which is acknowledged as a limitation. This study did not involve human or animal participants and therefore did not require ethical approval. However, all sources were handled in accordance with institutional ethical standards for scholarly citation.

C. Results and Discussion

This section presents the documentary findings for the four infrastructures in turn and then summarises their coding against the five components. It reports what the sources document, while the interpretation of these findings in terms of paradigm reconstruction is reserved for the discussion that follows, so that the two parts can be read as one argument.

1. Results

1.1 A Philological Infrastructure: Corpus Coranicum

Corpus Coranicum began in 2007 at the Berlin-Brandenburg Academy of Sciences and Humanities with a planned duration of twelve years, aiming at a holistic documentation of the Qur'anic text and an

extensive commentary on each sūrah.[14] Its digital environment is organised into separate databases, namely Manuscripta Coranica, Lectiones Coranicae, Bibliographia Coranica, Glossarium Coranicum, a collection of intertexts, and the Qur'anic text itself, and the connecting key between these databases is the sūrah and the verse.[15] The intertext module collects heterogeneous texts attached to each verse so that users can trace the traditions to which the text refers.

Two assessments of the project bear on the question of paradigm. Syahrullah and colleagues present it as a methodological model that could allow Islamic studies to compete epistemologically with the textual studies of the Bible and Midrash, with implications for digital training and access to manuscript data.[16] Rashwani, by contrast, evaluates part of its published data critically and describes the project as a kind of digital treasury of Western contributions to Qur'anic studies since the nineteenth century, combining classical philology, material history and modern literary analysis, and he examines its method and epistemological premises.[17] The same infrastructure is therefore read both as an enabling model and as a carrier of particular scholarly premises.

1.2 An Annotated Linguistic Resource: The Quranic Arabic Corpus

The Quranic Arabic Corpus, developed at the University of Leeds, presents itself as offering three levels of analysis: morphological annotation,

[14] "Corpus Coranicum," project description, Berlin-Brandenburgische Akademie der Wissenschaften, DH Abstracts, University of Virginia Library, https://dh-abstracts.library.virginia.edu/works/1433.

[15] "Corpus Coranicum," project description.

[16] Syahrullah et al., "Corpus Coranicum and Digital Philology."

[17] Samer Rashwani, "al-Mawsūʿah al-Qurʾāniyyah (Corpus Coranicum) min al-Fīlūlūjiyā ilā al-Taḥlīl al-Adabī: Qirāʾah Naqdiyyah fī Qāʿidat al-Bayānāt" [Corpus Coranicum from philology to literary analysis: A critical reading of the database], *Journal of College of Sharia and Islamic Studies* 38, no. 1 (2020): 72–90.

a syntactic treebank and a semantic ontology.[18] It adopts a tagset that closely follows traditional Arabic grammar rather than a framework imported from general linguistics,[19] and its syntax is annotated as a dependency treebank developed through online collaboration.[20] As reported in 2010, all 77,430 words of the Qur'an had a manually verified morphological analysis, while 11,000 words had been syntactically annotated as a gold-standard treebank.[21] The treebank is open to online collaborative correction, with suggestions reviewed by expert Arabic linguists and compared with published books on Qur'anic syntax, and annotation guidelines are used to keep contributions consistent among participants of differing expertise.[22]

The project states that its scope is the linguistic annotation of morphology and syntax, that English translation is offered only as a guide to meaning, and that its data are freely available for viewing and download under a public licence.[23] A survey of Qur'anic corpus models for corpus-based research concludes that existing monolingual corpora are adequate for such studies, whereas parallel corpora are lacking.[24]

1.3 Social Media and Online Learning

The studies of everyday digital practice describe a different kind of infrastructure, one in which the unit of interpretation is the post or the online lesson. The analysis of Qur'anic commentary on Indonesian Facebook frames

[18] "Frequently Asked Questions," The Quranic Arabic Corpus, University of Leeds, https://corpus.quran.com/faq.jsp.

[19] Kais Dukes and Nizar Habash, "Morphological Annotation of Quranic Arabic," in *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)* (2010).

[20] Kais Dukes, Eric Atwell, and Abdul-Baquee M. Sharaf, "Syntactic Annotation Guidelines for the Quranic Arabic Dependency Treebank," in *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)* (2010).

[21] Dukes, Atwell, and Sharaf, "Syntactic Annotation Guidelines."

[22] Dukes, Atwell, and Sharaf, "Syntactic Annotation Guidelines."

[23] "Frequently Asked Questions," Quranic Arabic Corpus.

[24] Hizbullah and Mutaali, "Quranic Corpus Models."

this activity as a form of digital hermeneutics and as a new face of the commentary tradition.[25] Research on digital Qur'an learning identifies a shift in the isnād system and the confirmation of new authority.[26] These sources document the phenomenon at the level of practice and do not offer a technical specification of the platforms involved, so that the coding for this case rests on the authors' own framing.

1.4 Generative AI as Interpreter

The evidence on generative AI is the most heterogeneous. In the evaluation of seven mainstream models on Q 9:60, two raters independently assessed 28 responses and produced 248 audited error findings.[27] The models identified the three reference exegetes, al-Ṭabarī, al-Qurṭubī and Ibn Kaṡīr, and their works, but they preserved isnād, narration structure, citation accuracy and attribution poorly, with unsupported additions and shifts of meaning dominating the errors, and higher-performing models produced more polished answers without eliminating source-level failures.[28]

Other evaluations point in the same direction while differing in language and task. The Bengali benchmark examined seven open-weight models ranging from 1 billion to 120 billion parameters on 1,680 double-annotated responses; six of the seven confidently fabricated in 79.3 to 95.3 per cent of source-production questions, performance did not improve monotonically with scale, and hadith provenance was the weakest category.[29] A study of Arabic Islamic question answering reports that some models, particularly open-source ones, fabricated Qur'anic verses when

[25] Lukman, "Digital Hermeneutics."
[26] Majit and Miski, "Pembelajaran al-Qur'an secara digital."
[27] Saifullah, Wahid, and Fazrian, "Between Machines and the Tafsīr Tradition."
[28] Saifullah, Wahid, and Fazrian, "Between Machines and the Tafsīr Tradition."
[29] "Do Open Language Models Invent Scripture?"

justifying answers on inheritance law.[30] A preprint evaluating generative AI across Qur'anic, hadith and fiqh knowledge finds that the most frequent problems were incomplete citation and insufficient attribution rather than entirely fabricated content.[31] Evaluation of religious reliability has also been extended to abstention: one study built 960 questions across the four Sunni schools in English and Arabic in order to examine whether models know when not to answer.[32] IslamicEval 2025 formalised two subtasks, the detection and correction of quoted verses and hadith, and question answering over the Qur'an and hadith.[33] Retrieval-augmented systems have been proposed as a response, but the evaluation of seven models notes that such work has concentrated on accuracy, fluency and usefulness rather than on the interpretive hierarchy of tafsir.[34]

1.5 Cross-Case Coding

Table 1 summarises the coding of the four cases against the five components. The coding indicates that every case transforms at least one component, and that the cases differ most markedly in the basis of authority and in the mechanisms of verification.

[30] "Detecting Hallucinations and Recovering Verified Answers in Arabic Islamic Question Answering," arXiv preprint 2608.03720, https://arxiv.org/pdf/2608.03720.

[31] "AI and Authenticity in Islamic Research: A Critical Evaluation of Generative AI Reliability, Hallucination, and Source Fidelity in Quranic, Hadith, and Fiqh Knowledge," arXiv preprint 2607.28237, https://arxiv.org/pdf/2607.28237.

[32] Farah Atif et al., "Sacred or Synthetic? Evaluating LLM Reliability and Abstention for Religious Questions," in *Proceedings of the Eighth AAAI/ACM Conference on AI, Ethics, and Society* (2025).

[33] Mubarak et al., "IslamicEval 2025."

[34] Saifullah, Wahid, and Fazrian, "Between Machines and the Tafsīr Tradition."

Table 1. Coding of four digital infrastructures against the five paradigm components

Component	Philological database (Corpus Coranicum)	Annotated corpus (Quranic Arabic Corpus)	Social media and online learning	Generative AI
Object	Manuscripts, variants and intertexts linked verse by verse	Word-level morphological, syntactic and conceptual annotation	Qur'an and commentary as circulating posts and lessons	Text as statistical pattern; verses and exegetes reproduced as output
Method	Historical-critical philology in linked databases	Computational linguistics using categories of traditional Arabic grammar	Informal, dialogical commentary; digital learning	Prompted generation; retrieval-augmented variants proposed
Authority and transmission	Institutional editorial control; premises questioned in critique	Collaborative corrections reviewed by expert linguists	Isnād system shifting; new authority confirmed	Fluent output with weak isnād, attribution and fabricated sources
Community and access	Online modules for researchers	Open data and volunteer collaboration	Mass participation	Mass consumer use for guidance
Ethics and verification	Documented sources and bibliographic module	Manual verification and annotation guidelines	Verification question raised by changing authority	Benchmarks, shared tasks and abstention tests emerging

2. Discussion

The most specific finding of this study is that the four infrastructures do not form a single technological wave but expose different strata of the same problem. Corpus Coranicum and the Quranic Arabic Corpus extend the print-era paradigm: they enlarge the object of study and the methods available while keeping an identifiable editorial or expert responsibility, and the Quranic Arabic Corpus does so explicitly by retaining the categories of traditional Arabic grammar. Social-media and online-learning practices strain the paradigm in a different place, because they redistribute interpretive authority across participants who stand outside the guild. Generative models strain it most acutely, because they produce the surface form of tafsir,

including the names of exegetes and the grammar of citation, without the underlying chain that makes a citation answerable. In Kuhnian terms, the accumulating anomaly is the separation of fluency from attribution, and this is something the earlier paradigm never had to name because, in the print era, an interpretation was assumed to be traceable to an identifiable interpreter and text.

This reading refines the existing literature in a specific way. Work on manuscripts and corpora has stressed access and plurality,[35] and work on digital practice has stressed shifts in authority,[36] whereas evaluations of language models have stressed accuracy and abstention.[37] Placed side by side, these emphases show that the common denominator is provenance, that is, whether a claim can be traced to its source with the intermediate steps intact. The evidence is not uniform: the Bengali benchmark documents confident fabrication at high rates,[38] while another evaluation finds incomplete citation more common than invention.[39] The difference is itself informative, since it suggests that the failure is a spectrum from missing attribution to fabricated attribution, and that any verification regime must address both ends.

On this basis the reconstructed paradigm can be stated through the five components, summarised in Table 2. As to the object, Qur'anic studies need to work with a layered text in which the received text, scholarly annotation and machine-generated output are kept distinct and labelled, so that a generated paraphrase is never mistaken for a witness. As to the method, philology, computational linguistics and hermeneutics should be used in triangulation, with computational output treated as a hypothesis

[35] Fedeli, "Digital Humanities and Qur'ānic Manuscript Studies."
[36] Majit and Miski, "Pembelajaran al-Qur'an secara digital."
[37] Atif et al., "Sacred or Synthetic?"
[38] "Do Open Language Models Invent Scripture?"
[39] "AI and Authenticity in Islamic Research."

awaiting interpretive validation rather than as a result. As to authority, the principle proposed here is provenance-first: every interpretive claim, whether human or machine, should expose the path by which it reaches its sources, and uncertainty or abstention should be displayed rather than concealed. As to community, the open but expert-moderated model of the Quranic Arabic Corpus, where wide participation is filtered through specialist review, offers a workable pattern for plural contribution. As to ethics, machine-mediated guidance must not fabricate scripture, must disclose its limits, and should be positioned as an aid to retrieval and study rather than a source of ruling.

Table 2. Reconstructed paradigm of Qur'anic studies in the digital era

Component	Classical-print paradigm	Digital-era anomaly	Reconstructed principle
Object	Mushaf and its commentaries	Text, annotation and generated output blur together	Layered text with labelled strata
Method	Disciplinary reading within tafsir and philology	Computational output presented as finding	Triangulation; output as hypothesis
Authority	Personal chains of transmission	Fluency without attribution; diffuse lay authority	Provenance-first claims; visible uncertainty
Community	Guild of recognised scholars	Mass participation and consumption	Open contribution with expert review
Ethics	Scholarly integrity and ijāzah	Fabricated verses and sources	No invented scripture; disclosure of limits

The provenance-first principle is offered as a conceptual proposal, not as an empirical result, and its justification draws on both of the study's theoretical resources. The classical discipline of isnād and the evaluative hermeneutic of tafsīr bi al-ma'ṡūr, which centres on traceable attribution, tarjīḥ and fidelity to sources, supply the criteria that the AI evaluation used

to detect failure.[40] It is in this sense that isnād can be called the epistemic ancestor of data provenance: both are practices of making the path of a claim inspectable. Gadamer's hermeneutics explains why this matters beyond technique. If understanding is always mediated by an effective history, then a database, a corpus annotation scheme or a language model is part of the interpreter's horizon, and its prejudices, in Gadamer's technical sense, include the selection of data, the categories of annotation and the statistical regularities of training.[41] A fusion of horizons is possible only when this mediation is visible and open to correction. Hidden mediation produces the illusion of immediacy, in which the user believes that the text speaks directly while an unexamined apparatus is speaking for it.

The same argument clarifies the critical reading of Corpus Coranicum. Rashwani's concern with its epistemological premises is not a rejection of digital philology but an instance of making mediation visible.[42] A reconstructed paradigm would welcome such critique as a standing feature of digital infrastructures, in the way that the Quranic Arabic Corpus invites correction and subjects it to expert review.[43] It would also reject the opposite temptation to treat any digital tool as inherently suspect, since the evidence shows that infrastructures built on traditional categories can strengthen rather than displace disciplinary accountability.

Several developments follow for the further growth of the field. In pedagogy, curricula in Qur'anic studies need to teach three literacies together: philological literacy, computational literacy, and attributional literacy, the last meaning the habit of asking where a claim comes from and whether the chain is intact. The finding that models can name the right exegetes while losing the chain makes this third literacy a practical necessity

[40] Saifullah, Wahid, and Fazrian, "Between Machines and the Tafsīr Tradition."
[41] Gadamer, *Truth and Method*.
[42] Rashwani, "al-Mawsūʿah al-Qurʾāniyyah."
[43] Dukes, Atwell, and Sharaf, "Syntactic Annotation Guidelines."

for students who will encounter machine-generated tafsir.[44] In peer review, journals can ask authors to declare the digital tools and generated material used, to verify every quoted verse and citation against a primary source, and to distinguish annotation from interpretation; shared tasks on the detection of misquoted verses show that such verification can be partly automated and benchmarked.[45] In infrastructure, the sūrah and verse key that already connects the databases of Corpus Coranicum could serve as a shared identifier through which manuscript databases, linguistic corpora and retrieval systems are interlinked, so that a generated answer can be traced to specific annotated witnesses.[46] Such an arrangement would not remove the need for scholarly judgement, but it would make judgement auditable.

The contribution of these findings extends beyond the local context. The claim that digital philology allows Islamic studies to stand on equal epistemological footing with the study of other scriptural traditions has been made for Corpus Coranicum,[47] and the difficulties of fabricated scripture are not confined to Arabic, as the Bengali benchmark shows.[48] A provenance-first paradigm is therefore transferable to other languages and scriptural corpora, and it speaks to current debates on the governance of AI in religious domains, where the question of who may speak with authority is now posed to machines. For Indonesian scholarship, where digital Qur'anic practice is already vigorous,[49] the model offers a way to connect pesantren and university traditions of isnād with open digital infrastructures.

The study has limitations that qualify these claims. It rests on published documentation and reported evaluations rather than on its own

[44] Saifullah, Wahid, and Fazrian, "Between Machines and the Tafsīr Tradition."
[45] Mubarak et al., "IslamicEval 2025."
[46] "Corpus Coranicum," project description.
[47] Syahrullah et al., "Corpus Coranicum and Digital Philology."
[48] "Do Open Language Models Invent Scripture?"
[49] Lukman, "Digital Hermeneutics."

tests of the systems, and the coding was performed by a single analyst. The evidence on social-media and online-learning practices is drawn from studies whose own framing the coding follows, so this case is the least independently verified. Part of the evidence on generative AI comes from preprints and conference pages and may be revised, and language models change faster than scholarly publication, so that specific figures should be read as dated observations. Future research could test the five-component model through direct audits of platforms, experimental comparison of retrieval-augmented and unaugmented systems against a tafsīr bi al-ma'ṡūr rubric, and ethnographic study of how teachers, students and religious scholars use these tools in practice.

D. Conclusion

This article set out to reconstruct the paradigm of Qur'anic studies for the digital era and to identify principles for its development. Through a conceptual-hermeneutic analysis and a documentary analysis of four infrastructures, it showed that philological databases and annotated corpora extend the existing paradigm, that social-media and online-learning practices redistribute authority, and that generative language models expose the sharpest anomaly by reproducing the form of tafsir without the chain of attribution that gives it standing. The study's main contribution is a five-component model, covering object, method, authority, community and ethics, together with the provenance-first principle that links the classical discipline of isnād to contemporary data provenance.

The academic position taken is that digital transformation should be met neither by wholesale acceptance nor by wholesale suspicion, but by a paradigm in which the mediation of the text is always visible and every claim remains answerable to its sources. This position has practical consequences for curricula that teach attributional literacy alongside philological and computational skills, for peer review that verifies quotations and declares digital tools, and for infrastructures that interlink open resources through

shared verse identifiers while retaining expert review. These consequences also invite comparison with other scriptural traditions facing the same questions.

Since the study relied on published documentation, its limitations point to the next steps: direct audits of platforms and models, controlled comparisons of verification strategies, and field studies of digital Qur'anic practice among teachers, students and scholars. Such work would test whether the proposed model holds across languages and institutions, and it would refine the principle that a digital Qur'anic scholarship worth the name is one in which fluency never substitutes for provenance.

References

"AI and Authenticity in Islamic Research: A Critical Evaluation of Generative AI Reliability, Hallucination, and Source Fidelity in Quranic, Hadith, and Fiqh Knowledge." arXiv preprint 2607.28237. https://arxiv.org/pdf/2607.28237.

Atif, Farah, Nursultan Askarbekuly, Kareem Darwish, and Monojit Choudhury. "Sacred or Synthetic? Evaluating LLM Reliability and Abstention for Religious Questions." In *Proceedings of the Eighth AAAI/ACM Conference on AI, Ethics, and Society*. 2025.

Berlin-Brandenburgische Akademie der Wissenschaften. "Corpus Coranicum." Project description. DH Abstracts, University of Virginia Library. https://dh-abstracts.library.virginia.edu/works/1433.

"Detecting Hallucinations and Recovering Verified Answers in Arabic Islamic Question Answering." arXiv preprint 2608.03720. https://arxiv.org/pdf/2608.03720.

"Do Open Language Models Invent Scripture? BILAL-BENCH: Benchmarking Islamic-Source Fabrication in Bengali." NeurIPS 2026. https://neurips.cc/virtual/2026/166644.

Dukes, Kais, and Nizar Habash. "Morphological Annotation of Quranic Arabic." In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. 2010.

Dukes, Kais, Eric Atwell, and Abdul-Baquee M. Sharaf. "Syntactic Annotation Guidelines for the Quranic Arabic Dependency Treebank." In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. 2010.

Eickelman, Dale F., and Jon W. Anderson, eds. *New Media in the Muslim World: The Emerging Public Sphere*. 2nd ed. Bloomington: Indiana University Press, 2003.

Fedeli, Alba. "Digital Humanities and Qur'ānic Manuscript Studies: New Perspectives and Challenges for Collaborative Spaces and Plural Views." *Journal of College of Sharia and Islamic Studies* 38, no. 1 (2020): 147–58. https://doi.org/10.29117/jcsis.2020.0260.

Gadamer, Hans-Georg. *Truth and Method*. 2nd rev. ed. Translated by Joel Weinsheimer and Donald G. Marshall. London: Continuum, 2004.

Hizbullah, Nur, and Abdul Mutaali. "Quranic Corpus Models for Corpus-Based Studies." In *Proceedings of the Eleventh Conference on Applied Linguistics (CONAPLIN 2018)*. Atlantis Press, 2019. https://doi.org/10.2991/conaplin-18.2019.270.

Kuhn, Thomas S. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press, 1962.

Lukman, F. "Digital Hermeneutics and a New Face of the Qur'an Commentary: The Qur'an in Indonesian's Facebook." *Al-Jāmiʿah: Journal of Islamic Studies* 56, no. 1 (2018): 95–120. https://doi.org/10.14421/ajis.2018.561.95-120.

Majit, A., and M. Miski. "Pembelajaran al-Qur'an secara digital: Pergeseran sistem isnad dan peneguhan otoritas baru." *Jurnal SMART (Studi Masyarakat, Religi, dan Tradisi)* 9, no. 1 (2023): 133–46. https://doi.org/10.18784/smart.v9i1.1795.

Misbakhuddin, Alfian Dhany. "Digital Hermeneutics of the Qur'an: Towards a Holistic Model of Faith-Technology Integration." ICOERESS, n.d. https://icoeress.pascasarjana.uinjambi.ac.id/ojs/index.php/icoeress/article/view/82.

Mubarak, Hamdy, Rana Malhas, Watheq Mansour, Abubakr Mohamed, Mahmoud Fawzi, Majd Hawasly, Tamer Elsayed, Kareem Mohamed Darwish, and Walid Magdy. "IslamicEval 2025: The First Shared Task of Capturing LLMs Hallucination in Islamic Content." In *Proceedings of the Third Arabic Natural Language Processing Conference: Shared Tasks*,

480–93. Suzhou, China: Association for Computational Linguistics, 2025.

Muhanna, Elias, ed. *The Digital Humanities and Islamic & Middle East Studies*. Berlin: De Gruyter, 2016.

Quranic Arabic Corpus. "Frequently Asked Questions." University of Leeds. https://corpus.quran.com/faq.jsp.

Rashwani, Samer. "al-Mawsūʿah al-Qurʾāniyyah (Corpus Coranicum) min al-Fīlūlūjiyā ilā al-Taḥlīl al-Adabī: Qirāʾah Naqdiyyah fī Qāʿidat al-Bayānāt" [Corpus Coranicum from philology to literary analysis: A critical reading of the database]. *Journal of College of Sharia and Islamic Studies* 38, no. 1 (2020): 72–90.

Saifullah, Soleh Hasan Wahid, and Muhammad Ihza Fazrian. "Between Machines and the Tafsīr Tradition: Reorienting Tafsīr bi al-Ma'ṡūr in the Age of Computational Interpretation." *Jurnal Studi Ilmu-ilmu Al-Qur'an dan Hadis* 27, no. 2 (2026): 458–93. https://doi.org/10.14421/qh.v27i1.7499.

Syahrullah, Wildan Imaduddin Muhammad, Eva Nugraha, and Aulia Raudhatul Jannah. "Corpus Coranicum and Digital Philology: A Methodological Model for Advancing Qur'anic Manuscript Studies in Indonesia." *Khazanah Theologia* 6, no. 3 (2024): 201–22.