PEDAGOGIK | JURNAL PENDIDIKAN
P-ISSN : 2528-5793
E-ISSN : 2354-7960

THEORETICAL FRAMEWORK USE IN ARTIFICIAL INTELLIGENCE RESEARCH IN ISLAMIC EDUCATION: A SYSTEMATIC LITERATURE REVIEW

Ali Muhtarom [1], Mufrodi Mufrodi [1]*
[1] Universitas Islam Negeri Sultan Maulana Hasanuddin, Banten, Indonesia
Correspondence Author: alimuhtarom@uinbanten.ac.id

Received: April 2026	**Accepted:** October 2026	**Published:** November 2026
DOI: https://doi.org/10.33650.pjp.v13i3.14937		

Abstract : *Research on artificial intelligence (AI) in Islamic education has grown rapidly, raising questions about the role of Islamic intellectual traditions in its theoretical foundations. This systematic review examines how scholarship engages with Islamic concepts and develops theoretical frameworks. Following PRISMA 2020 guidelines, 118 articles published between 2020 and 2026 were identified from Scopus and the Directory of Open Access Journals, representing 87 journals. Articles were coded by research design, Islamic conceptual engagement, and theoretical frameworks, including non-Islamic perspectives. Inter-rater reliability was assessed on 20% of the corpus. Findings show that 52.5% of articles discuss at least one Islamic concept, while 32.2% articulate no theoretical framework. This gap is concentrated in empirical studies, which show lower framework engagement than conceptual and review articles. Differences between Scopus-indexed and open-access publications are limited. The review distinguishes conceptual presence from theoretical integration, providing a basis for evaluating Islamic intellectual contributions to AI research in education. Ta'lim and isnad offer potential foundations for context-sensitive frameworks, but their theoretical and methodological utility requires further conceptual development and empirical validation.*

Keywords : *Artificial Intelligence; Islamic Education; Theoretical Frameworks.*

Abstrak : *Research on artificial intelligence (AI) in Islamic education has grown rapidly, raising questions about the role of Islamic intellectual traditions in its theoretical foundations. This systematic review examines how scholarship engages with Islamic concepts and develops theoretical frameworks. Following PRISMA 2020 guidelines, 118 articles published between 2020 and 2026 were identified from Scopus and the Directory of Open Access Journals, representing 87 journals. Articles were coded by research design, Islamic conceptual engagement, and theoretical frameworks, including non-Islamic perspectives. Inter-rater reliability was assessed on 20% of the corpus. Findings show that 52.5% of articles discuss at least one Islamic concept, while 32.2% articulate no theoretical framework. This gap is concentrated in empirical studies, which show lower framework engagement than conceptual and review articles. Differences between Scopus-indexed and open-access publications are limited. The review distinguishes conceptual presence from theoretical integration, providing a basis for evaluating Islamic intellectual contributions to AI research in education. Ta'lim and isnad offer potential foundations for context-sensitive frameworks, but their theoretical and methodological utility requires further conceptual development and empirical validation.*

Kata Kunci : *Kecerdasan Buatan; Pendidikan Islam; Kerangka Teoretis.*

INTRODUCTION

The integration of artificial intelligence (AI) into education has expanded rapidly, encompassing adaptive learning, automated assessment, intelligent tutoring, and conversational systems. Systematic reviews have documented the growing volume and diversity of AI applications, particularly in higher education, while also emphasizing the need for stronger conceptual foundations, ethical scrutiny, and methodological rigour (Bond et al., 2024; Crompton & Burke, 2023; Zawacki-richter et al., 2019). These developments have also reached Islamic education, where AI is increasingly discussed as a means of supporting instruction, learning, assessment, and access to religious knowledge (Zawacki-richter et al., 2019). However, the expansion of AI applications raises a question that cannot be answered by examining technological adoption or educational outcomes alone: how are the pedagogical and epistemological assumptions underlying these applications theoretically justified within Islamic educational contexts?

This question is particularly important because Islamic education involves more than the acquisition of knowledge or the improvement of instructional efficiency. Its intellectual tradition encompasses questions of knowledge, authority, moral formation, and the purposes of education. Concepts such as *tarbiyah*, *ta'lim*, and *ta'dib* articulate complementary dimensions of nurturing human development, transmitting knowledge, and cultivating appropriate conduct (Al-Attas, 1999; Halstead, 2004). Related concepts, including *adab*, *akhlaq*, *fitrah*, and *amanah*, provide additional perspectives on character, moral responsibility, human disposition, and trust. These concepts offer potential resources for examining AI-mediated education, particularly when technological systems influence how knowledge is presented, interpreted, assessed, or transmitted (Fatih et al., 2026; Mazrur et al., 2025; Thoriquttyas & Rohmawati, 2024). Nevertheless, the availability of an intellectual tradition does not establish that its concepts are systematically incorporated into contemporary research. Their scholarly significance depends on how researchers define them, connect them to research questions, and use them to guide interpretation.

The central problem, therefore, is not simply whether AI research in Islamic education mentions Islamic values. It is whether such research articulates theoretical positions that make its assumptions explicit and its conclusions open to critical examination (Munawwaroh & Putri, 2024; Soulisa et al., 2022). A study may invoke Islamic terminology without using it to structure its research design or analysis. Conversely, a study may employ an established educational or technological framework without explicitly engaging Islamic epistemological or pedagogical perspectives (Azizah & Wardefi, 2026; Mukaffan & Siswanto, 2025; Shalih, 2026). These possibilities represent different forms of theoretical

engagement and should not be treated as equivalent. Distinguishing them is necessary to assess whether Islamic concepts function as analytical resources or primarily as contextual descriptions.

Existing research provides a starting point for examining this problem, but its scope requires careful differentiation. Reviews of AI in education have primarily mapped applications, research trends, educational outcomes, and methodological developments across broad educational settings (Crompton & Burke, 2023; Zawacki-richter et al., 2019). Bond et al., (2024), through a meta-systematic review of AI in higher education, further emphasized the importance of conceptual grounding, ethics, collaboration, and research rigour. These contributions establish the broader significance of theoretical and methodological questions in AI research. However, their general educational scope does not, by itself, establish how theoretical frameworks are used in Islamic education or whether Islamic normative concepts inform the analysis of AI-related educational practices.

Reviews specifically addressing AI in Islamic education have begun to map this more focused field. For example, Amrani et al., (2025) examined the implementation of AI in Islamic education, while Arif et al., 2026 reviewed applications, challenges, and future directions in Islamic religious education. Taha & Ajmain@Jima'ain, (2025), examined research trends in Islamic religious education using a corpus of 34 Scopus-indexed articles published between 2021 and 2025. These reviews contribute to understanding the applications, implementation challenges, and development of AI-related research in religious educational contexts. However, their stated research objectives differ from the present study's focus on the explicit articulation of theoretical frameworks and the analytical role of Islamic normative concepts. This distinction identifies a more precise research gap than the general assertion that AI research in Islamic education has not previously been reviewed. The relevant gap concerns the systematic assessment of how theoretical positions are stated, what kinds of frameworks are employed, and how these patterns relate to research design and publication context.

The distinction matters because theoretical frameworks perform functions beyond providing terminology or background context. They establish the assumptions through which a research problem is defined, evidence is interpreted, and conclusions are justified. In AI-related educational research, these functions are especially consequential because technological systems embody assumptions about learning, knowledge, assessment, and the roles of teachers and learners (Holmes et al., 2022; Holmes & Tuomi, 2022; Selwyn, 2019). Without an explicit analytical framework, researchers may leave these assumptions insufficiently examined. In Islamic education, this raises a further question: whether and

how established educational theories and Islamic intellectual concepts are brought into a productive analytical relationship. This concern does not imply that every study must adopt an explicitly Islamic framework. Rather, it calls for distinguishing between studies that state and apply a theoretical position, studies that invoke Islamic concepts without demonstrable analytical integration, and studies that articulate no theoretical framework.

To investigate this problem, the present study conducts a systematic review of 118 articles published between 2020 and 2026, retrieved from Scopus and the Directory of Open Access Journals (DOAJ) and drawn from 87 journals. The review follows PRISMA 2020 reporting guidance (Page et al., 2021). Each article is examined in relation to three dimensions: research design, engagement with Islamic normative concepts, and the use of theoretical frameworks, including frameworks originating outside Islamic intellectual traditions. Inter-rater reliability is assessed on 20% of the corpus by two independent coders. The analysis investigates the prevalence and forms of theoretical engagement, examines whether engagement differs between empirical and conceptual or review studies, and compares patterns across Scopus-indexed and open-access publications. This design enables the study to move beyond general claims about the strengths or weaknesses of AI research and examine the distribution of theoretical practices within a defined body of literature.

The study makes three contributions. First, it provides a systematic empirical account of theoretical framework use in AI research on Islamic education, complementing reviews that primarily examine applications, implementation, and research trends. Second, it distinguishes the presence of Islamic normative concepts from the explicit articulation of a theoretical framework, allowing conceptual visibility and theoretical integration to be assessed as separate dimensions. Third, it identifies patterns of theoretical engagement across research designs and publication contexts, thereby establishing an evidence-based foundation for future methodological and conceptual development. The review also considers how concepts such as *ta'lim* and *isnad* might inform more context-sensitive analytical frameworks. *Ta'lim* directs attention to the purposes, processes, and responsibilities involved in teaching and transmitting knowledge, whereas *isnad* foregrounds chains of transmission and the evaluation of the authority and provenance of transmitted knowledge. Their relevance to AI-mediated education lies in the questions they may generate about instructional responsibility, the provenance of AI-generated information, and the conditions under which knowledge claims are warranted. These concepts are presented as starting points for theoretical development, not as already validated frameworks. Accordingly, the study aims to clarify the current state of theoretical engagement and identify directions for developing more explicit, critically defensible frameworks for research on AI in Islamic education.

METHOD

The review follows the PRISMA 2020 statement for reporting systematic reviews (Page et al., 2021). Its object is the theoretical standing of a body of scholarship rather than the effects of an intervention, so the extracted unit is an article's stated position rather than a finding or an effect size. The protocol, the full search strings, and the complete inclusion and exclusion criteria are given in Appendix A, and the coding manual with its decision rules is available from the authors. The review was not registered, which is noted here as a limitation on its prospective transparency.

Two databases were searched on 13 October 2026. Scopus was queried through the advanced search interface using a TITLE-ABS-KEY expression combining terms for artificial intelligence with terms for Islamic education and its institutional settings, and the Directory of Open Access Journals was queried through its public application programming interface using three separate expressions, since that interface does not accept a single query of equivalent length. Scopus returned 75 records and the Directory of Open Access Journals returned 193 unique records across the three queries. A third database, Garuda, was tested and then excluded on methodological grounds, because its indexing treats transliteration variants inconsistently in a way that makes recall unquantifiable: a search for ta'dib returned no records while tadib returned twenty-one, and ta'lim returned none while talim returned ninety-four. Including a source whose recall cannot be characterised would have compromised the reproducibility of the search rather than improved its coverage.

The combined yield of 268 records reduced to 258 after duplicates were removed by digital object identifier where present and by normalised title otherwise. Ten records were duplicates, eight of them appearing in both databases and two within the open-access directory. Titles, abstracts, and keywords were screened against criteria specifying the topic, the educational level, the publication type, and the language. Articles were retained if they addressed artificial intelligence in an educational context within Islamic education, were published between 2020 and 2026 in English or Indonesian, and took the form of a peer-reviewed journal article. Conference abstracts and editorials were excluded. The review was originally scoped to higher education, but educational level proved too often unstated in abstracts to serve as a reliable exclusion criterion, so it was recorded as a coded variable instead and the distribution is reported below. Screening reduced the set to 209 records, and the full flow of records is shown in Figure 1.

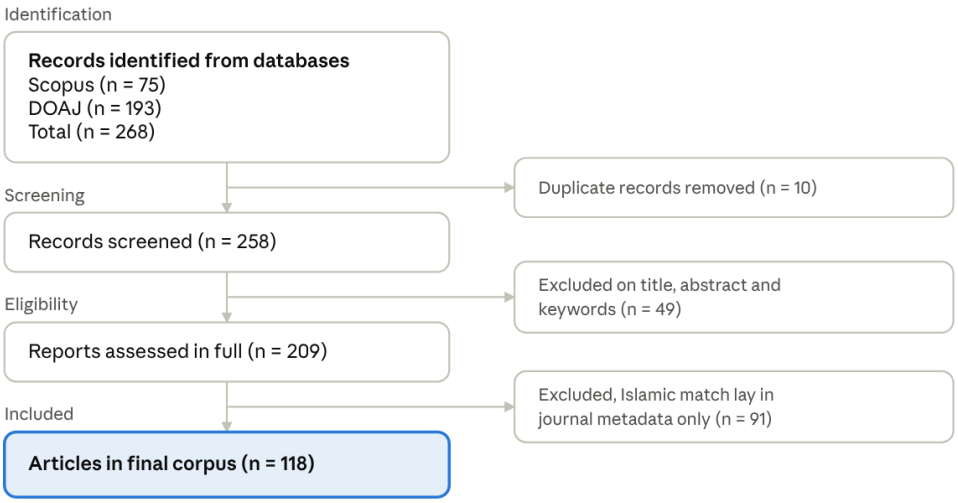


Figure 1: Flow of Records Through Identification, Screening, Eligibility Assessment, and Inclusion, Reported According to PRISMA 2020

A further exclusion was applied after an error was detected during coder training. The Directory of Open Access Journals matches query terms against journal and publisher metadata as well as article content, with the consequence that articles on unrelated subjects published by institutions whose names contain Islamic educational terms were returned by the search. Articles on trigonometry instruction, audit quality, and halal tourism entered the initial set through this mechanism. Each of the 209 records was therefore re-examined for the presence of Islamic educational content in the article itself rather than in the metadata of its outlet, and ninety-one records were excluded on this basis. The final corpus comprises 118 articles drawn from 87 journals, and the list with digital object identifiers is given in Appendix B.

Coding proceeded through a scheme of two layers. The first layer was applied to all 118 articles from their titles, abstracts, and keywords, and recorded bibliographic metadata, research design, the presence of engagement with Islamic normative concepts, the presence of a non-Islamic theoretical framework, and the location within the article at which Islamic concepts appeared. The second layer, which records the analytic function a concept performs within an argument, requires full text and has not been executed, because nineteen of the sixty-two articles that engage Islamic concepts are indexed only in Scopus and have no accessible full text. Coding only the accessible remainder would have produced an open-access-biased subset and so invalidated the comparison between Scopus-indexed and open-access outlets reported below. The second layer is therefore reported as a stated continuation rather than as a partial result.

Two coders participated. The first coded the full corpus, and the second independently coded a random sample of twenty-four articles, representing just over

twenty per cent of the corpus, without sight of the first coder's decisions. The first round of agreement was poor for the two substantive variables, with Cohen's kappa of 0.391 for research design and 0.098 for engagement with Islamic concepts (Cohen, 1960). Inspection of the disagreements identified two ambiguities in the rules rather than inattention by the coders: the manual did not distinguish an Islamic setting from an Islamic concept, and it offered a mixed category for research design without defining it. The rules were sharpened to require that a concept be used as a term of analysis rather than as a description of the institution where data were collected, and the mixed category was removed. A re-test on the same twenty-four articles under the revised rules, conducted after an interval and with the original decisions withheld, yielded kappa of 0.913 for research design and 0.595 for engagement with Islamic concepts, which falls in the moderate range (Landis & Koch, 1977).

Agreement on the variables recording where in an article a concept appears was affected by skewed marginal distributions. The category recording a match arising only from institutional affiliation showed 91.7 per cent observed agreement yet a kappa of 0.000, which is the documented behaviour of the coefficient when one category is rare (Cicchetti & Feinstein, 1990). Gwet's first-order agreement coefficient, which is robust to this condition, was therefore computed alongside kappa and returned 0.909 for that variable (Gwet, 2008; Wongpakaran et al., 2013). Both statistics are reported, and neither replaces the other. Two further location variables, recording appearance in a theoretical framework section and in a discussion section, proved unreliable because they required inference about full text from abstracts alone, and they were excluded from the first layer and deferred to the second.

Disagreements were resolved by discussion to consensus, and where the two coders could not be reconciled from abstracts the article was marked for full-text adjudication. Every cell in the final dataset carries a record of its origin, distinguishing values taken from the initial coding, from the revised re-test, from the verification exercise, and from consensus, and the resulting provenance file accompanies the dataset. Frequencies were computed for all coded variables, and the association between research design and framework engagement was tested with a chi-square test of independence, with the conventional expected-count requirement checked before the test was applied (Krippendorff, 2019).

RESULT AND DISCUSSION

1. Results

a) Characteristics and Methodological Profiles of the Included Studies

The final corpus comprised 118 articles published across 87 journals. Although the search period covered 2020–2026, no article published before 2023 met the

inclusion criteria. Publications were concentrated in 2025 and 2026, which together accounted for 81.4% of the corpus. English-language publications predominated (89.0%), and Indonesia represented 66 of the 77 articles with usable affiliation data. Higher education was the most frequently identified educational level, although 28.0% of articles did not specify an educational level. The distribution of publication years, languages, databases, and educational levels is presented in Table 1.

Table 1: Characteristics of the Included Articles (n = 118)

Characteristic	Category	n	%
Year of publication	2023	6	5.1
	2024	16	13.6
	2025	52	44.1
	2026	44	37.3
Language	English	105	89.0
	Indonesian	12	10.2
	Arabic	1	0.8
Source database	Directory of Open Access Journals	69	58.5
	Scopus	41	34.7
	Both	8	6.8
Educational level	Higher education	44	37.3
	Not stated	33	28.0
	Secondary	16	13.6
	Not level-specific	12	10.2
	Primary	7	5.9
	Pesantren	6	5.1

Regarding research characteristics, empirical studies constituted the largest category, comprising 84 articles (71.2%), followed by literature reviews (16.1%) and conceptual studies (12.7%). Qualitative approaches were the most frequently reported design, followed by mixed-methods and quantitative approaches. Primary data were collected in 68.6% of the corpus. However, methodological transparency was uneven: 77.1% of articles explicitly stated their methods, while others reported them only partially or did not provide an identifiable methodological statement. These characteristics are summarised in Table 2.

Table 2: Research Design, Method, and Data Source (n = 118)

Variable	Category	n	%
Type of study	Empirical	84	71.2
	Literature review	19	16.1
	Conceptual	15	12.7
Design	Qualitative	41	34.7

	Mixed methods	22	18.6
	Quantitative	20	16.9
	Conceptual analysis	12	10.2
	Systematic review	11	9.3
	Not stated	9	7.6
	Bibliometric	3	2.5
Collects primary data	Yes	81	68.6
	No	37	31.4
Method stated explicitly	Yes	91	77.1
	Partially	20	16.9
	No	7	5.9

Taken together, the corpus reflects a rapidly expanding body of research dominated by empirical investigations and publications associated with Indonesia. The combination of substantial empirical activity and uneven methodological reporting provides important context for examining how theoretical positions are articulated. However, the geographical and publication distributions should be interpreted as characteristics of the included corpus rather than as representative estimates of the entire field of AI research in Islamic education.

b) Theoretical Frameworks and Engagement with Islamic Normative Concepts

Engagement with Islamic normative concepts were more prevalent than anticipated in the preliminary assessment. Of the 118 articles, 62 (52.5%) engaged at least one Islamic concept as an analytical term. The preliminary scan had suggested a proportion below 5%, but it focused on only three concepts, whereas the final coding included a broader vocabulary encompassing *akhlaq, maqasid, amanah, fitrah*, Islamic values, and Islamic epistemology. This difference demonstrates the importance of a sufficiently broad coding framework when assessing conceptual engagement in a diverse body of literature.

The configuration of theoretical frameworks provides a more differentiated picture. As shown in Table 3, 47 articles (39.8%) engaged Islamic concepts without a named non-Islamic framework, 18 (15.3%) used a non-Islamic framework without an identified Islamic concept, and 15 (12.7%) incorporated both. The remaining 38 articles (32.2%) stated neither. Among these 38 articles, 31 were empirical studies, and 29 collected primary data. Thus, a substantial proportion of the corpus consists of empirical research in which no theoretical framework was explicitly identified.

Table 3: Configuration of Stated Theoretical Frameworks by Type of Study (n = 118)

Configuration	Empirical	Literature review	Conceptual	Total	%
Islamic concepts only	29	11	7	47	39.8
No framework stated	31	4	3	38	32.2
Non-Islamic framework only	16	1	1	18	15.3
Both	8	3	4	15	12.7

The distribution of non-Islamic frameworks further indicates that explicit framework use was limited. Table 4 shows that 85 articles (72.0%) did not name a non-Islamic framework. The Technology Acceptance Model and SWOT analysis were the most frequently identified approaches, with five articles each. Other frameworks, including Technological Pedagogical Content Knowledge, constructivism, technology ethics, the Unified Theory of Acceptance and Use of Technology, and Bloom's taxonomy, appeared less frequently.

Table 4: Non-Islamic Theoretical Frameworks Named (n = 118)

Framework	n	%
None named	85	72.0
Other, each appearing once	16	13.6
Technology Acceptance Model	5	4.2
SWOT analysis	5	4.2
Technological Pedagogical Content Knowledge	2	1.7
Constructivism	2	1.7
Technology ethics	1	0.8
Unified theory of acceptance and use of technology	1	0.8
Bloom's taxonomy	1	0.8

Engagement with Islamic concepts varies systematically with research design. Thirty-seven of the eighty-four empirical articles engage such a concept, a rate of 44.0 per cent, against twenty-five of the thirty-four conceptual and review articles, a rate of 73.5 per cent. A chi-square test of independence on this two-by-two classification returns a value of 8.44 with one degree of freedom and a probability below 0.01, with the smallest expected count standing at 16.1. The same pattern appears when data collection is used as the classifier, since articles collecting primary data engage Islamic concepts in 44.4 per cent of cases against 70.3 per cent among those that do not. Table 5 reports these comparisons together with those for source database and year.

c) Patterns of Theoretical Engagement, Concept Location, and Coding Reliability

Engagement with Islamic concepts varied by study type. As reported in Table 5, 44.0% of empirical studies engaged at least one Islamic concept, compared with 73.7% of literature reviews and 73.3% of conceptual studies. When empirical studies were compared with the combined group of conceptual and review articles, the difference was statistically significant, $\chi^2(1) = 8.44$ $p < .01$. This result indicates an association between study type and engagement with Islamic concepts, but it does not establish that research design causes the observed difference.

A similar descriptive pattern emerged when articles were classified by primary data collection: 44.4% of articles collecting primary data engaged Islamic concepts, compared with 70.3% of articles that did not. By contrast, the differences across database sources were comparatively limited: engagement was identified in 46.3% of Scopus-indexed articles and 52.2% of DOAJ articles. The proportion for articles indexed in both sources was higher, but this subgroup contained only eight articles. Publication-year differences were also variable, and no formal statistical tests were conducted for database or year because the relevant cell counts were considered insufficient for the planned chi-square analysis.

Table 5. Engagement with Islamic Concepts by Study Characteristic

Characteristic	Category	Engaging / total	%
Type of study	Empirical	37 / 84	44.0
	Literature review	14 / 19	73.7
	Conceptual	11 / 15	73.3
Collects primary data	Yes	36 / 81	44.4
	No	26 / 37	70.3
Source database	Scopus	19 / 41	46.3
	Directory of Open Access Journals	36 / 69	52.2
	Both	7 / 8	87.5
Year of publication	2023	3 / 6	50.0
	2024	6 / 16	37.5
	2025	28 / 52	53.8
	2026	25 / 44	56.8

The comparison by source database does not support the expectation that international indexing promotes more explicit theoretical engagement. Articles indexed in Scopus engage Islamic concepts at 46.3 per cent, compared with 52.2 per cent in the open-access directory. Although this difference runs counter to the initial

expectation, it is too small to support a meaningful interpretation given the sample sizes involved. The trend across publication years is similarly inconclusive. Engagement rises from 37.5 per cent in 2024 to 56.8 per cent in 2026, but the pattern is not monotonic, while the limited number of articles published in 2023 prevents reliable inference. Neither comparison was formally tested because the cell counts fall below the threshold required for an appropriate chi-square test. These results therefore indicate descriptive variation rather than evidence of statistically meaningful differences between databases or publication years.

Among the sixty-two articles engaging with at least one Islamic concept, sixty-one mention the concept in the abstract, twenty-four include it in the keywords, and twenty-one refer to it in the title, as shown in Table 6. This distribution suggests that Islamic concepts are more consistently visible in abstracts than in titles or keywords, although their presence alone does not establish theoretical integration. Across the full corpus, eight articles were retrieved solely through institutional affiliations containing Islamic educational terminology, without corresponding conceptual content in the article itself. These articles were retained but classified as affiliation-only matches to maintain transparency in the screening process. Two additional variables, recording whether concepts appeared in theoretical framework or discussion sections, were initially coded but later withdrawn. During reconciliation, both coders acknowledged that their assessments relied on inferred section contents that were not directly accessible. Removing these variables reduced the risk of unsupported classification and strengthened the methodological consistency of the analysis.

Table 6: Location of Islamic Concepts Within the Articles that Engage Them (n = 62)

Location	n	% of 62
Abstract	61	98.4
Keywords	24	38.7
Title	21	33.9

Agreement between the two coders under the revised rules is reported in Table 7. Type of study reached a kappa of 0.913, which falls in the highest conventional band (Landis & Koch, 1977). Engagement with Islamic concepts reached 0.595, which is moderate and is reported without adjustment. Three of the location variables behaved in the manner documented for rare categories, returning high observed agreement alongside a depressed or null kappa (Cicchetti & Feinstein, 1990), so Gwet's first-order coefficient, which is robust to that condition, is reported beside kappa for every variable (Gwet, 2008). The weakest result concerns the appearance of a concept in the

title, where observed agreement of 70.8 per cent yields a kappa of 0.045 and an agreement coefficient of 0.581, and the corresponding figure in Table 6 should therefore be read as indicative rather than as a stable estimate.

Table 7: Inter-rater Agreement Under the Revised Coding Rules (n = 24)

Variable	Observed agreement	Cohen's kappa	Gwet's AC1
Type of study	95.8%	0.913	0.945
Engagement with Islamic concepts	79.2%	0.595	0.590
Concept in abstract	79.2%	0.595	0.590
Concept in keywords	87.5%	0.647	0.807
Concept in title	70.8%	0.045	0.581
Affiliation-only match	91.7%	0.000	0.909

Overall, the results identify study type as the clearest observed correlate of engagement with Islamic concepts. Conceptual and review articles were more likely than empirical studies to engage these concepts, while the database comparisons did not establish a similarly clear pattern. The reliability results also indicate that findings concerning conceptual engagement should be interpreted alongside the variation in coding agreement. Together, these findings provide an empirical basis for examining the distinction between the visibility of Islamic concepts and their explicit integration into theoretical frameworks.

2. Discussion

The first result corrects a premise that circulates widely in critical commentary and that the present authors themselves held before coding began. Islamic normative concepts are not largely absent from this literature; more than half the corpus engages at least one. What is absent is something different and more serious, since nearly a third of the articles state no theoretical position of any kind, and that group is composed overwhelmingly of empirical studies working with primary data. The pattern therefore does not describe a field that has forgotten its tradition, but one that sets the tradition aside at the precise point where a framework would have to do analytic work.

The association with research design supports this reading. Conceptual papers and literature reviews engage Islamic concepts at roughly 73 per cent, empirical studies at 44 per cent, and the difference is statistically reliable. Conceptual work can invoke tarbiyah or adab and leave the term at the level at which the tradition states it, because nothing in the genre requires the concept to discriminate between observations.

Empirical work cannot, and the evidence here suggests that when authors reach that point they more often drop the concept than specify it. The constraint that the data reveal is therefore operationalisability rather than neglect, which matters because the two diagnoses imply entirely different remedies.

The remedy that follows from an operationalisability diagnosis is conceptual specification rather than exhortation. Ta'lim, taken as the accountable transmission of knowledge from a source whose provenance can be established, already contains the elements of a measurable criterion (Al-Attas, 1999; Fajriyah et al., 2026; W. M. N. W. Daud, 1998). Rendered as an expectation, it holds that an instructional system satisfies ta'lim only to the degree that a learner can trace a transmitted claim to an identifiable source and assess the standing of that source. This is directly testable against generative systems, which characteristically produce fluent assertions whose provenance is unavailable to the learner, and it yields a design criterion as well as a research question. The tradition's own apparatus for assessing transmission, the isnad with its graded evaluation of each narrator in a chain, supplies a vocabulary for that assessment that is considerably more precise than the general appeals to reliability found in the technology literature (Brown, 2009).

A second reformulation follows the same pattern. Ibn Khaldun treats the substance of instruction as the teacher's accumulated judgement, acquired through habituation rather than through the possession of information (Khaldun, 1967), and adab specifies the conduct through which a learner stands in relation to a source of knowledge (Memon & Zaman, 2016). Stated as an expectation, this holds that a system which supplies conclusions without requiring the learner to exercise judgement does not perform the function the tradition assigns to teaching, however accurate its outputs. Both reformulations are offered as illustrations of method rather than as settled instruments, and neither has been validated here. Their point is to show that the gap between a normative concept and a measurement can be closed by argument, which is work that the field has so far largely left undone (Sahin, 2018).

The findings extend an established critique to a literature that had not been examined for it. Reviews of artificial intelligence in education have reported weak theoretical grounding for some years (Bond et al., 2024; Crompton & Burke, 2023; Zawacki-richter et al., 2019), and critical scholarship has explained the pattern by the way educational technologies carry their own implicit models of learning into the studies that evaluate them (Bearman et al., 2023; Selwyn, 2019). The present corpus shows that the availability of an elaborate alternative vocabulary does not by itself prevent this. It also shows that the frameworks that are named, when any are, skew heavily toward

acceptance and adoption models rather than toward accounts of learning, which reproduces at the level of theory the tendency Biesta identifies at the level of measurement (Biesta, 2010; Mundiri et al., 2025). A field studying instruction while theorising adoption will describe uptake and call it education.

The comparison by indexing status carries a further implication for editorial practice. Articles in Scopus-indexed outlets engage Islamic concepts no more often than those in open-access journals, and if anything slightly less. International indexing therefore does not function as a mechanism of theoretical discipline in this literature, which places the responsibility on reviewers and editors at the level of the individual submission. Requiring authors to state a theoretical position, and to say what observations would be inconsistent with it, is a low-cost editorial intervention that the present results suggest would bind on roughly a third of submissions. The ethics literature on artificial intelligence in education makes a parallel argument about normative commitments, holding that they cannot be deferred because they are already operative in the systems under study (Holmes & Tuomi, 2022; Nguyen et al., 2023).

Several limitations qualify these conclusions. The first coding layer worked from titles, abstracts, and keywords, so an article that develops a framework in its body without signalling it in the abstract is recorded here as stating none, and the true rate of framework use is therefore a lower bound rather than a point estimate. Agreement on the principal substantive variable was moderate at 0.595, and agreement on whether a concept appears in the title was poor, so the corresponding figure should be treated as indicative. The second coding layer, which would record the analytic function a concept performs, was not executed, because full text is unavailable for a subset of the relevant articles and coding only the accessible remainder would have biased the comparison between indexing sources. The review covers two databases and omits regional indexes whose transliteration handling could not be characterised, and it was not prospectively registered (Page et al., 2021). Educational level is heterogeneous within the corpus and was treated as a descriptive variable rather than as a boundary, which limits what can be said about higher education specifically.

CONCLUSION

This systematic review examined the theoretical positioning of AI research in Islamic education through an analysis of 118 articles published between 2020 and 2026. The findings reveal that engagement with Islamic normative concepts is more prevalent than initially anticipated, yet their presence does not necessarily correspond to the explicit articulation of

theoretical frameworks. Nearly one-third of the reviewed articles stated no framework, with empirical studies accounting for most of this group. Furthermore, engagement with Islamic concepts was significantly associated with study type, being more frequent in conceptual and review articles than in empirical studies. These findings distinguish the presence of Islamic concepts from their explicit theoretical articulation and provide an empirical basis for examining how Islamic intellectual traditions are incorporated into AI research. However, the results do not establish why this difference occurs or whether it reflects difficulties in operationalising normative concepts.

The study contributes a framework for distinguishing conceptual presence from explicit theoretical positioning and highlights the need for greater transparency in connecting theoretical assumptions to research questions, analytical procedures, and empirical evidence. Concepts such as *ta'lim*, *isnad*, and *adab* offer potential starting points for developing context-sensitive approaches to evaluating AI-mediated teaching, knowledge transmission, and learner judgement. Their analytical and methodological value, however, requires further conceptual clarification and empirical validation. Future research should examine the full texts of relevant studies, assess how Islamic concepts function within theoretical and analytical frameworks, and test their applicability across different educational settings. By clarifying how theoretical positions are articulated and applied, subsequent research can strengthen the conceptual and methodological foundations of AI research in Islamic education without assuming that a single theoretical tradition is appropriate for every research question.

ACKNOWLEDGMENT

The authors would like to express their sincere appreciation to Universitas Islam Negeri Sultan Maulana Hasanuddin Banten, Indonesia, for providing an academic environment that supported the completion of this study. The authors also acknowledge the valuable contributions of scholars whose publications served as the foundation for this systematic literature review. Their work has contributed significantly to the development of knowledge on the use of theoretical frameworks in artificial intelligence research within Islamic education.

BIBLIOGRAPHY

Al-Attas, S. M. N. (1999). The Concept of Education in Islam: A Framework for an Islamic Philosophy of Education. In *The First World Conference on Muslim Education*. International Institute of Islamic Thought and Civilization (ISTAC).

Amrani, M. Y. E., Vakayil, A., Mohammed, F., & Amri, F. A. (2025). Foundations of domain-specific large language models for Islamic studies: A comprehensive review. *Journal of ICT Research and Applications*, *19*(1), 69-85. https://doi.org/https://doi.org/10.5614/itbj.ict.res.appl.2025.19.1.4

Arif, M., Abd Aziz, M. K. N., Hassan, H., Fikri, Y., Zekkari, M., & Rismawati, B. V. (2026). Artificial intelligence in Islamic religious education: A PRISMA-based review and the SEI-CHAT framework. *Online Learning in Educational Research*, *6*(1), 65-85. https://doi.org/10.58524/oler.v6i1.993

Artha, B., Santoso, F., Bahri, B., Asri, C. P., Sari, U. T., Sari, N. P., & Ascasaput. (2026). Artificial intelligence and Islamic education: A theoretical review. *Anjasmoro*, *3*(1), 74-84. https://doi.org/https://doi.org/10.69965/anjasmoro.v3i1.214

Awaludin, A. (2024). Documenting the half-century evolution of Islamic education research: A probabilistic topic modeling study. *Studia Islamika*, *31*(3), 439-475. https://doi.org/https://doi.org/10.36712/sdi.v31i3.41513

Azizah, R., & Wardefi, R. (2026). Artificial intelligence and academic integrity in Islamic higher education: Exploring ethical challenges among Islamic education students. *Al-Adabiyah*, *7*(4), 625-646. https://doi.org/https://doi.org/10.35719/adabiyah.v7i4.1449

Bearman, M., Ryan, J., & Ajjawi, R. (2023). Discourses of artificial intelligence in higher education: A critical literature review. *Higher Education*, *86*(2), 369–385. https://doi.org/https://doi.org/10.1007/s10734-022-00937-2

Biesta, G. (2010). *Good education in an age of measurement: Ethics, politics, democracy*. Paradigm Publishers.

Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S. W., & Siemens, G. (2024). A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. *International Journal of Educational Technology in Higher Education*, *21*(4), 41-52. https://doi.org/https://doi.org/10.1186/s41239-023-00436-z

Brown, J. A. C. (2009). *Hadith: Muhammad's legacy in the medieval and modern world*. Oneworld Publications.

Cicchetti, D. V, & Feinstein, A. R. (1990). High agreement but low kappa: II. Resolving the paradoxes. *Journal of Clinical Epidemiology*, *43*(6), 551–558. https://doi.org/https://doi.org/10.1016/0895-4356(90)90159-M

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/https://doi.org/10.1177/001316446002000104

Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, *20*(22), 1132-1145. https://doi.org/https://doi.org/10.1186/s41239-023-00392-8

Daud, M. N. W. (1998). *The Educational Philosophy and Practice of Syed Muhammad Naguib Al-Attas: An Exposition of the Original Concept of Islamization*. International Institute of Islamic Thought and Civilization. https://books.google.co.id/books?id=0nM2PwAACAAJ

Daud, W. M. N. W. (1998). *The educational philosophy and practice of Syed Muhammad Naquib al-Attas*. International Institute of Islamic Thought and Civilization.

Fajriyah, N., Munawwaroh, I., & Sain, S. H. (2026). Beyond Inclusion: Uncovering The Role

Of Inclusive Learning In Shaping Students'perspective-Taking. *PEDAGOGIK: Jurnal Pendidikan*, *13*(1), 81-98. https://doi.org/10.33650/pjp.v13i1.13871

Fatih, N. I., Mubarokah, N., Risqianto, F., Hanifah, L., & Hanif, M. (2026). Perceptions and impacts of the use of artificial intelligence (AI) in deep learning among PAI students. *Tarbiyah Islamiyah*, *16*(1), 75-85. https://doi.org/https://doi.org/10.18592/jtipai.v16i1.20574

Gough, D., Oliver, S., & Thomas, J. (2017). *An introduction to systematic reviews* (2nd ed.). Sage.

Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal*, *26*(2), 91–108. https://doi.org/https://doi.org/10.1111/j.1471-1842.2009.00848.x

Gwet, K. L. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology*, *61*(1), 29–48. https://doi.org/https://doi.org/10.1348/000711006X126600

Halstead, J. M. (2004). An Islamic Concept of Education. *Comparative Education*, *40*(4), 517–529. https://doi.org/10.1080/0305006042000284510

Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., Santos, O. C., Rodé, M. T., Cukurova, M., Bittencourt, I. I., & Koedinger, K. R. (2022). Ethics of AI in education: Towards a community-wide framework. *International Journal of Artificial Intelligence in Education*, *32*(3), 504–526. https://doi.org/https://doi.org/10.1007/s40593-021-00239-1

Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European Journal of Education*, *57*(4), 542–570. https://doi.org/https://doi.org/10.1111/ejed.12533

Hwang, G.-J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, *1*, 100001. https://doi.org/https://doi.org/10.1016/j.caeai.2020.100001

Khaldun, I. (1967). *The Muqaddimah: An Introduction to History The Classic Islamic Histroy of The World* (F. Rosenthal (ed.); 2nd ed. NV). Princeton University Press.

Krippendorff, K. (2019). *Content analysis: An introduction to its methodology* (4th ed.). Sage.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. https://doi.org/https://doi.org/10.2307/2529310

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson.

Mazrur, M., Jennah, R., Norhidayah, S., & Surawan, S. (2025). Integrating technology acceptance and pedagogical deep learning in Islamic education: A TAM-based study. *Ta'dib*, *28*(2), 501-514. https://doi.org/https://doi.org/10.31958/jt.v28i2.15981

Memon, N. A., & Zaman, M. (2016). *Philosophies of Islamic education: Historical perspectives and emerging discourses* (N. A. Memon & M. Zaman (eds.); Editors). Routledge.

Mukaffan, & Siswanto, A. H. (2025). Ethical integration of generative AI in Islamic education: Toward inclusive and sustainable human capital development. *Al-Adabiyah*, *6*(3), 303-325. https://doi.org/https://doi.org/10.35719/adabiyah.v6i3.1178

Munawwaroh, I., & Putri, D. F. (2024). Enhancing Critical Thinking Through the Integration of Self-Directed Learning in Sustainable Education in Madrasah. *AFKARINA: Jurnal Pendidikan Agama Islam*, *9*(1), 1-10. https://doi.org/10.33650/afkarina.v9i1.9352

Mundiri, A., Munawwaroh, I., Hadi, M. I., Baharun, H., Shudiq, W. J. F., & Maulidy, A.

(2025, July). Artificial Intelligence (AI) Innovation in Education: From Data-Driven Learning to Automated Teaching. In *2025 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)* (pp. 173-180). IEEE. 10.1109/IAICT65714.2025.11100623

Nguyen, A., Ngo, H. N., Hong, Y., Dang, B., & Nguyen, B.-P. T. (2023). Ethical principles for artificial intelligence in education. *Education and Information Technologies*, *28*(4), 4221–4241. https://doi.org/10.1007/s10639-022-11316-w

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ Journal*, *372*, 71. https://doi.org/https://doi.org/10.1136/bmj.n71

Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Blackwell Publishing.

Sahin, A. (2018). Critical issues in Islamic education studies: Rethinking Islamic and Western liberal secular values of education. *Religions*, *9*(11), 335-341. https://doi.org/https://doi.org/10.3390/rel9110335

Selwyn, N. (2019). *Should robots replace teachers? AI and the future of education*. Polity Press.

Shalih, B. (2026). From kitab kuning to ChatGPT: Authority, authenticity, and pesantren pedagogy in the age of generative artificial intelligence. *Hunafa: Jurnal Studia Islamika*, *23*(1), 23-41. https://doi.org/https://doi.org/10.24239/jsi.v23i1.952

Soulisa, I., Supratman, M., Rosfiani, O., Renaldi, R., Sopiah, Utomo, W. T., Hermawan, C. M., Ariati, C., Riyanti, A., Tauran, S. F., Irwanto, Astiswijaya, N., & Yenni, A. S. (2022). Evaluasi Pembelajaran. In E. Damayanti (Ed.), *Widina Bhakti Persada Bandung*. Widina Bhakti Persada Bandung.

Suparto, S., Kusaeri, K., & Putri, R. K. A. (2025). Deep learning in automated essay scoring for Islamic education: A systematic review. *Online Learning in Educational Research*, *5*(2), 319-338. https://doi.org/https://doi.org/10.58524/oler.v5i2.753

Taha, J. Z. A., & Ajmain@Jima'ain, M. T. (2025). The role of ChatGPT in enhancing learning among Islamic education undergraduate students: A systematic literature review. *Journal of Research, Innovation, and Strategies for Education*, *2*(2), 52-68. https://doi.org/https://doi.org/10.70148/rise.v2i2.24

Thoriquttyas, T., & Rohmawati, N. (2024). How far artificial intelligence influenced mu'allim, murabbi, and mudarris? Transhumanism and diffusion of innovation theory. *Suhuf*, *36*(2), 139-154. https://doi.org/https://doi.org/10.23917/suhuf.v36i2.6302

Waghid, Y. (2011). *Conceptions of Islamic education: Pedagogical framings*. Peter Lang.

Wongpakaran, N., Wongpakaran, T., Wedding, D., & Gwet, K. L. (2013). A comparison of Cohen's kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: A study conducted with personality disorder samples. *BMC Medical Research Methodology*, *13*, 61. https://doi.org/https://doi.org/10.1186/1471-2288-13-61

Zawacki-richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators ? *International Journal of Educational Technology in Higher Education*, *16*(39), 1–27. https://doi.org/https://doi.org/10.1186/s41239-019-0171-0

Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., Istenic, A., Spector, M., Liu, J.-B., Yuan, J., & Li, Y. (2021). A review of artificial intelligence (AI) in education from 2010 to 2020. *Complexity*, *2021*, 8812542. https://doi.org/10.1155/2021/8812542