

A Decade in Computational Quranic Studies (2016–2025): A Systematic Survey of the Evolution from Statistical Methods to Large Language Models

Wadee A. Nashir ^(1,*)

Belal Al-Fuhaidi ^(1,*)

Naseebah Maqtary ^(1,*)

Ahmed A. F. Mhoub ^(2,*)

© 2026 University of Science and Technology, Sana'a, Yemen. This article can be distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

© 2026 جامعة العلوم والتكنولوجيا، اليمن. يمكن إعادة استخدام المادة المنشورة حسب رخصة مؤسسة المشاع الإبداعي شريطة الاستشهاد بالمؤلف والمجلة

¹ Department of Computer Science, University of Science and Technology, Sana'a, Yemen

² Small Project Management Department, Sana'a Community College; Department of Administrative Sciences, University of Science and Technology, Sana'a, Yemen

* Corresponding authors: w.alqubaty@ust.edu.ye, belalarh@ust.edu.ye, n.maqtary@ust.edu.ye, dr.ahmedfarhan55@gmail.com

A Decade in Computational Quranic Studies (2016–2025): A Systematic Survey of the Evolution from Statistical Methods to Large Language Models

Abstract:

The intersection of artificial intelligence and sacred scripture represents a critical yet underexplored frontier in the digital humanities. This paper presents the first decade-spanning, systematic survey of computational Qur'anic studies (2016–2025), tracing the field's evolution from fragmented, ontology-based prototypes to a coherent, methodologically mature research discipline. Through a rigorous, multi-phase systematic review of 112 peer-reviewed studies, we map a transformative trajectory across three eras: a Foundational Era (2016–2019) focused on knowledge representation; a Transformer Turn (2020–2023) driven by community benchmarks and Arabic transformer models; and an Acceleration Era (2024–2025) marked by the integration of Large Language Models (LLMs) and Retrieval-Augmented Generation (RAG). Our longitudinal analysis uncovers a clear paradigm shift—from rule-based and statistical methods toward deep architectures explicitly engineered for ethical and theological fidelity. While remarkable advances have been achieved in Qur'anic Question Answering, Verse Authentication, and Recitation Analysis, the field still faces major gaps in multilingual applications, computational modeling of canonical recitation traditions (Qira'ā't), and the deep integration of classical tafsīr scholarship. By introducing a structured taxonomy, a diachronic synthesis of research trends, and a strategic roadmap for the next decade, this survey reframes computational Qur'anic studies as both a case study and a catalyst for innovation in responsible, explainable, and theologically grounded AI.

Keywords: Computational Qur'anic studies, Arabic Natural Language Processing (NLP), Transformer Models; Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), Systematic review, digital humanities.

الدراسات الحاسوبية للقرآن الكريم خلال عقد (2016-2025): مراجعة منهجية لمسار التطور من النماذج الإحصائية إلى النماذج اللغوية الضخمة

الملخص:

يمثل التقاطع بين الذكاء الاصطناعي والنصوص المقدسة مجالاً نقدياً ومهماً، لكنه لا يزال غير مستكشف بالقدر الكافي ضمن الدراسات الإنسانية الرقمية. تقدم هذه الورقة أول مسح منهجي يمتد على عقد كامل للدراسات الحاسوبية للقرآن الكريم (2016-2025)، حيث ترصد تطور هذا المجال من نماذج أولية متفرقة قائمة على الأنطولوجيا إلى حقل بحثي متماسك ناضج منهجياً. ومن خلال مراجعة منهجية صارمة متعددة المراحل شملت 112 دراسة محكمة، نرسم مسارا تحولياً عبر ثلاث مراحل زمنية رئيسية: المرحلة التأسيسية (2016-2019) التي ركزت على تمثيل المعرفة، ومرحلة التحول نحو نماذج المحولات (2020-2023) المدفوعة بالمعايير المجتمعية ونماذج المحولات العربية، ومرحلة التسارع (2024-2025) التي تميزت بدمج النماذج اللغوية الضخمة (LLMs) وتقنيات التوليد المعزز بالاسترجاع (RAG). يكشف التحليل الطولي عن تحول نموذجي واضح من الأساليب القائمة على القواعد والإحصاء إلى معماريات عميقة صُممت صراحة مع مراعاة الضوابط الأخلاقية والدقة اللاهوتية. وعلى الرغم من التقدم الملحوظ في مجالات مثل الإجابة الآلية عن الأسئلة القرآنية، وتوثيق الآيات، وتحليل التلاوة، لا يزال المجال يواجه فجوات جوهرية، لاسيما في التطبيقات متعددة اللغات، والنمذجة الحاسوبية للقراءات القرآنية المتواترة، والدمج العميق للتراث التفسيري الكلاسيكي. ومن خلال تقديم تصنيف بنيوي، وتركيب تاريخي للتوجهات البحثية، و خارطة طريق استراتيجية للعقد القادم، تعيد هذه الدراسة تأطير الدراسات الحاسوبية للقرآن بوصفها دراسة حالة ومحفزاً للابتكار في تطوير ذكاء اصطناعي مسؤول، قابل للتفسير، ومتجذر لاهوتياً.

الكلمات المفتاحية: الدراسات الحاسوبية للقرآن الكريم، المعالجة الآلية للغة العربية (NLP)، نماذج المحولات (Transformer Models)، النماذج اللغوية الضخمة (LLMs)، التوليد المعزز بالاسترجاع (RAG)، المراجعة المنهجية، الإنسانيات الرقمية.

1. Introduction

The intersection of artificial intelligence (AI) and religious scripture represents a rapidly expanding and methodologically complex frontier in the digital humanities—one where the computational ‘reading’ of sacred texts is redefining the epistemic foundations of millennia-old traditions of scholarship and interpretation [1]. Within this domain, the Holy Qur’an—the central religious text for approximately 1.8 billion Muslims worldwide [2]—presents a case of extraordinary significance and linguistic sophistication. Its Classical Arabic composition, marked by intricate morphology, layered semantic density, and an unbroken oral transmission, poses formidable challenges that continually stretch the expressive limits of Natural Language Processing (NLP). As Quranic engagement increasingly migrates to digital platforms, the development of robust, transparent, and theologically sensitive computational models has become not only an academic pursuit but also a cultural and ethical necessity—vital for education, accessibility, and the preservation of textual integrity in an era of algorithmic mediation and rampant misinformation [3]. This paper presents the first systematic, decade-long survey of this rapidly evolving field, tracing its evolution from a landscape of fragmented, ontology-driven experiments to a methodologically coherent and empirically grounded research discipline.

The motivation for this survey arises from the unprecedented acceleration and diversification of AI capabilities over the past decade. To establish a clear historical baseline for our analysis, we conducted a preliminary mapping of the preceding period (2003–2016). This initial review revealed a field in its nascent and exploratory phase, overwhelmingly dominated by knowledge engineering, with ontological methods constituting nearly 60% of all research efforts. The primary challenges of that era were a critical absence of theological validation—with only about 8% of studies verifying their findings against classical exegetical sources (*tafsīr*)—and the proliferation of fragmented, non-reusable, and often inaccessible ontological resources. As it stood at the beginning of our survey period in 2016, the field was thus defined by proof-of-concept systems that had yet to achieve an integrative, linguistically rigorous, or semantically robust understanding of the Qur’an. This well-defined historical baseline, established as part of the present study, provides a unique vantage point from which to measure the field’s remarkable transformation over the subsequent decade.

While several recent surveys have provided useful snapshots of computational approaches to Islamic texts [4], [5], the velocity, diversity, and conceptual reorientation of recent developments—particularly the transition from classical machine learning to transformer architectures and, most recently, to generative AI—renders static inventories quickly obsolete. A new, comprehensive and diachronic synthesis is therefore required, not merely to update the corpus of existing works, but to trace the intellectual genealogy of the field, identify the catalysts of its paradigm shifts, and critically assess how the gaps identified in 2016 have been addressed, reinterpreted, or perpetuated. This survey meets that need by offering a longitudinal and interpretive perspective on the decade from 2016 to 2025—a period that encapsulates the field’s entire modern evolutionary arc, from its formative struggles to its current state of methodological maturity.

Existing literature from this period, when viewed chronologically, reveals a clear, multi-stage evolutionary trajectory. The initial phase (c. 2016–2019) extended the trends of the preceding decade, maintaining a focus on ontology development and rule-based systems [6], [7]. However, the subsequent “Transformer Turn” (c. 2020–2023) marked a decisive methodological consolidation, as the adoption of BERT-style models and the emergence of community benchmarks—such as the Qur’an QA shared tasks [8]—galvanized the research community and enabled more rigorous, comparative experimentation. Yet, no single study to date has systematically documented this ten-year intellectual trajectory or analyzed how the field’s methodological and epistemological orientations evolved in relation to its historical origins. Prior surveys have tended to be either too broad—encompassing all religious texts—or too narrow, confined to specific computational tasks, thereby failing to capture the holistic interplay of technology, linguistics, and theology that defines this domain.

This paper fills this critical gap by presenting a systematic and comprehensive review of all peer-reviewed computational and AI-driven Qur’anic studies published between January 1, 2016, and October 24, 2025. Our primary objective is to move beyond a simple inventory of methods to provide a critical, data-driven analysis of the field’s intellectual and technical development. Specifically, this survey aims to: first, document the chronological evolution of research volume and thematic focus, identifying distinct developmental phases; second, analyze the methodological paradigm shifts from ontology-driven models to deep learning and Retrieval-Augmented Generation (RAG)

architectures [9]; third, map the co-evolution of research and infrastructural resources, including annotated corpora and evaluation benchmarks [10]; fourth, evaluate the extent to which the critical gaps identified in 2016 have been addressed in the subsequent decade; and fifth, assess how the unique challenges of a sacred text have catalyzed innovation in areas such as explainability [11] and hallucination mitigation.

The remainder of this paper is organized to systematically unfold this narrative. Section 2 details the rigorous, multi-phase systematic methodology employed to identify and select the primary studies. Section 3 presents a chronological reconstruction of the field’s evolution through three distinct eras. Building on this historical foundation, Section 4 provides a cross-cutting thematic analysis, examining dominant research tasks, methodological shifts, and infrastructural developments. Section 5 offers a critical synthesis, distilling the primary insights and trajectories. Section 6 outlines a strategic roadmap for the next decade, proposing concrete short-, medium-, and long-term research directions. Finally, Section 7 concludes by reflecting on the field’s trajectory and its broader implications for the intersection of AI, religion, and society.

2. Survey Methodology

To ensure a comprehensive, transparent, and reproducible review of the research landscape, this study employed a rigorous, multi-phase systematic search and selection methodology, adapted from established best-practice guidelines for systematic literature reviews [12], [13]. The overall process was meticulously designed to identify all relevant, peer-reviewed scientific publications applying computational and AI-driven methods to Qur’anic studies published between January 1, 2016, and October 24, 2025.

Our search strategy integrated three complementary and interlocking phases to maximize both breadth and depth of coverage. In the first phase, we conducted a systematic, database-driven exploration across seven high-coverage academic databases, selected to balance broad interdisciplinary reach (Google Scholar, Scopus, and Web of Science) with domain-specific precision in computer science and computational linguistics (IEEE Xplore, ACM Digital Library, ACL Anthology, and arXiv.org). A standardized search

string—iteratively refined through pilot testing and combining Qur’anic-specific terms with computational and AI-related keywords using Boolean operators—was applied uniformly across all databases, yielding an initial pool of 982 potentially relevant records.

In the second phase, to identify studies that might not have been retrieved through the primary database search, we employed a bidirectional snowballing strategy [14]. Backward snowballing involved examining reference lists of a seed set of 45 core publications, while forward snowballing traced subsequent citations to these works. These procedures collectively contributed an additional 85 candidate studies. In parallel, a targeted manual search was conducted across selected conference proceedings and workshops—namely OSACT, IslamicEval, ArabicNLP, and COLING-Rel—to capture recent contributions published in specialized or emerging venues.

In the third phase, study selection and screening were conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. As summarized in Figure 1, the combined search phases resulted in a total of 1,067 identified records. After the removal of duplicate entries, 870 unique records remained and were independently screened by two reviewers using predefined inclusion and exclusion criteria. Eligible studies were required to focus on computational or AI-based approaches applied to the Qur’anic text, to have been published between 2016 and 2025, to present original empirical or methodological contributions, and to be written in English.

Following title and abstract screening, 340 studies advanced to the next stage, from which 155 articles were selected for full-text assessment. During this evaluation, studies were excluded if they lacked methodological transparency, relied exclusively on pre-existing tools without introducing novel contributions, or failed to meet established quality standards. Any disagreements between reviewers were resolved through structured discussion and consensus. This multi-stage filtering process ultimately resulted in a final corpus of 112 primary studies, which form the analytical foundation of this survey.

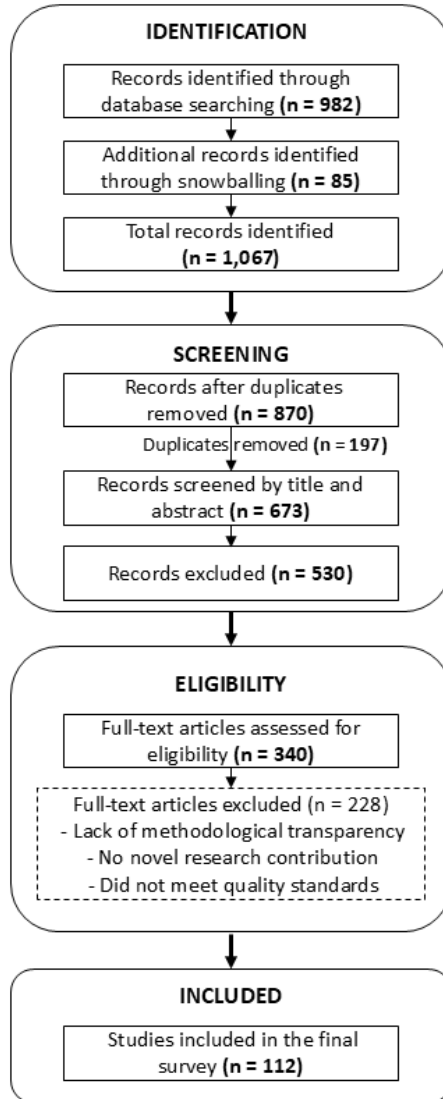


Figure 1: PRISMA flowchart illustrating the study selection procedure

Subsequently, each of the included studies underwent a qualitative quality assessment to gauge methodological rigor. This assessment was not used as an exclusion criterion but rather as a contextual weighting factor to support the interpretation of findings. For each study, detailed metadata were systematically extracted using a predefined data extraction schema, encompassing research tasks, methodologies, datasets, evaluation metrics, and key findings. These data were then synthesized, cross-compared, and

thematically coded, forming the empirical basis for the longitudinal and thematic analyses presented in the subsequent sections.

3. The Evolutionary Trajectory of Research: A Chronological Analysis

The decade between 2016 and 2025 did not witness a linear or incremental accumulation of knowledge in computational Qur'anic studies; rather, it underwent a dynamic, multi-phased, and transformative evolution. Our systematic analysis of 112 primary studies reveals three distinct yet interlinked eras, each defined by its dominant research paradigms, methodological priorities, and the gradual consolidation of its technical infrastructure. This chronological framework provides a narrative and analytical lens through which to understand the field's journey—from foundational, proof-of-concept explorations to the methodologically mature integration of state-of-the-art AI.

3.1. Phase I (2016–2019): The Foundational Era – The Preoccupation with Semantic Structure

The initial years of our survey period were characterized by a singular and foundational preoccupation: imposing a computationally tractable structure upon a text of exceptional linguistic and semantic complexity. This phase represents both a continuation and consolidation of the dominant trends from the preceding decade. Our baseline analysis of the period leading up to 2016 revealed a research landscape heavily dominated by knowledge engineering, with ontological methods constituting nearly 60% of all efforts.

Research between 2016 and 2019—although modest in volume—continued this orientation. Pioneering studies sought to formally model the conceptual architecture of the Qur'an [6], [15], representing a shift toward semantic retrieval. Parallel efforts during this phase, as cataloged in the Appendix, focused on constructing foundational linguistic resources and domain-specific ontologies for concepts such as nature and finance [7], [16]–[25]. Concurrently, early surveys and general text mining studies began to map the broader landscape of Arabic and Islamic NLP, establishing initial classifications for the field [26]–[31].

This ontological groundwork provided the scaffolding and conceptual infrastructure for the first generation of application-oriented systems, predominantly rule-based or hybrid early neural network-driven Question-Answering (QA) models [32]. Yet, this formative era was defined more by

its constraints than its breakthroughs. It was marked by: (1) an overreliance on small-scale, often privately curated datasets that limited reproducibility and generalization; (2) the absence of standardized evaluation protocols or shared benchmarks, which hindered cross-comparative progress; and (3) a heavy dependence on English translations, a pragmatic shortcut that circumvented the formidable challenges of Classical Arabic morphology while simultaneously diluting critical linguistic subtleties.

The principal contribution of this phase, therefore, was not the production of high-performance systems, but rather the epistemic demonstration that the Qur'anic corpus was indeed amenable to structured, machine-interpretable representation. In doing so, it laid the essential groundwork for the methodological and conceptual leap that would define the next era.

3.2. Phase II (2020–2023): The Transformer Turn – From Individual Efforts to a Benchmark-Driven Community

This period marks a decisive inflection point in the field's evolution, wherein research began to coalesce from a dispersed collection of isolated efforts into a coherent, benchmark-oriented scientific community. This transformation was catalyzed by two mutually reinforcing developments: the technological disruption brought by transformer architectures and the organizational momentum generated by community-led initiatives.

The advent of large, pre-trained Arabic transformer models such as AraBERT [33] furnished researchers with robust, contextually sensitive tools for semantic representation, thereby catalyzing a definitive shift away from the limitations of static embeddings and conventional machine learning pipelines.

The single most pivotal development of this era was the Qur'an QA 2022 Shared Task [8]. The release of the Qur'anic Reading Comprehension Dataset (QRCD) galvanized the community, attracting numerous participating teams. These contributions—detailed in the Appendix—collectively advanced the state of the art in Machine Reading Comprehension (MRC) through diverse approaches, ranging from ensemble BERT-based models to transfer learning strategies and data augmentation techniques [34]–[46]. Beyond accelerating progress in QA, this shared task redefined the methodological baseline for the entire field.

Beyond accelerating progress in QA, this shared task redefined the methodological baseline for the entire field, anchoring it in the fine-tuning of Arabic-specific transformer models and introducing reproducibility as a shared community value. Concurrently, independent research matured in both depth and analytical sophistication, with notable works applying deep learning to verse authentication [47] and doctoral theses exploring deep semantic modeling and interpretation [48].

This era thus signifies the field's critical transition—from exploratory, fragmented experimentation to systematic, benchmark-driven, and community-coordinated scientific inquiry. It represents not only a technological maturation, but also a sociological consolidation of the research community around shared datasets, methodologies, and evaluative norms.

3.3. Phase III (2024–2025): The Acceleration Era – The Synthesis of Technology and Theological Prudence

The most recent period is characterized by a remarkable synthesis between rapid technological acceleration and an emergent sense of theological and ethical self-awareness. This era witnessed an exponential surge in research output and a second major paradigm shift, defined by the integration of Large Language Models (LLMs) into Qur'anic computational studies. Crucially, this was not a mere technological enthusiasm or uncritical adoption of a global trend. Rather, it reflected a deliberate, domain-conscious adaptation of cutting-edge AI frameworks to a sacred linguistic and interpretive tradition.

The widespread implementation of Retrieval-Augmented Generation (RAG) architectures for Qur'anic Question Answering (QA) and semantic retrieval [9], [49] represents a methodologically mature and theologically cautious innovation. RAG emerged as a direct, contextually appropriate response to the persistent problem of model hallucination—a technological flaw that carries profound theological and ethical implications when applied to sacred text. The field, at this stage, demonstrates a heightened reflexivity, consciously engineering for accuracy, transparency, and fidelity rather than mere performance gains.

This period also witnessed a significant thematic diversification. Research in Qur'anic recitation analysis progressed with end-to-end architectures [50], bridging the gap between linguistic form and phonetic embodiment. The scope of analysis expanded dramatically in 2024–2025 (see Appendix for full listing) to encompass multilingual translation quality, sentiment analysis, and

semantic clarity in Urdu and English interpretations [51]–[57]. Furthermore, broader applications emerging in this phase included specialized studies on Hadith knowledge graphs, hallucination detection, and Explainable AI (XAI) for theological validation [11], [58]–[64].

Finally, this era is marked by a transformative leap in research infrastructure, signifying the field’s transition from methodological maturity to institutional consolidation. The development of the QuranMorph corpus, featuring detailed morphological annotations linked to over 110 classical lexicons [10], and the compilation of a massive, curated digital corpus encompassing 80 canonical tafsīr (exegetical) works [65], collectively constitute a paradigm-defining achievement. The research focus has thus evolved from analyzing the Qur’anic text in isolation to computationally modeling its vast ecosystem of interpretive scholarship.

This shift signals the emergence of a self-reflective discipline—one mature enough to bridge the centuries-old divide between algorithmic analysis and classical hermeneutics, and to position the Qur’an not merely as a dataset but as a dynamic site of dialogue between theology, technology, and epistemology.

4. Cross-Cutting Thematic Analysis: Mapping the Topologies of a Maturing Field

Beyond a simple chronological progression, the evolution of computational Qur’anic studies is best understood through a cross-cutting thematic analysis of its intellectual trajectories and methodological convergences. This section dissects the field’s development by mapping the symbiotic interplay between its research objectives, methodological paradigms, and foundational infrastructures. Our analysis of the 112 primary studies reveals that the field’s trajectory has not been random or opportunistic but strategically shaped by a reciprocal dynamic between pragmatic needs, technological affordances, and the gradual construction of a coherent research ecosystem.

4.1. A Functional Evolution of Research Tasks: From Data Structuring to Problem Solving

The intellectual focus of the field has undergone a marked functional evolution, transitioning from an early preoccupation with static knowledge representation to a more dynamic, problem-solving orientation. The distribution of primary research tasks by publication volume over the

decade is shown in Figure 2; this trajectory reveals a distinct epistemic shift. The data illustrates a clear progression from the initial dominance of “Ontology & Knowledge Representation” toward the eventual centrality of “QA & Information Retrieval,” culminating in a recent surge of pragmatically motivated tasks such as “Verse Authentication and Classification,” This pattern underscores the move from mere data organization toward enabling meaningful, context-aware human engagement with the Qur’anic text.

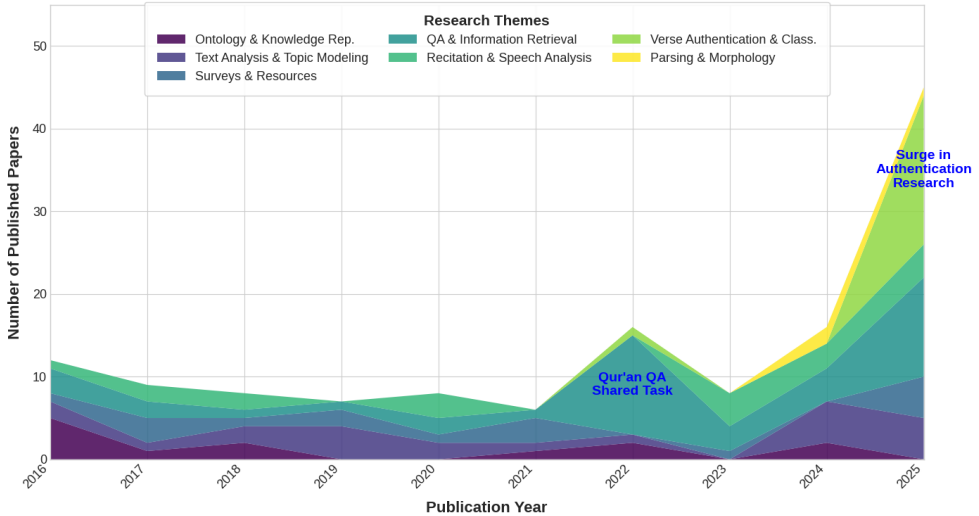


Figure 2: Thematic evolution of research tasks by publication volume (2016–2025)

The chart immediately reveals that the Foundational Era (2016–2019) was overwhelmingly characterized by efforts in Knowledge Representation, where ontology development dominated the research landscape. Numerous studies sought to formally encode Qur’anic concepts to move beyond keyword-based retrieval and toward semantic, relation-oriented modeling [6], [7], [15]–[21].

A pivotal functional shift emerged around 2022 as the research agenda diversified in response to pressing, real-world challenges. Figure 2 visualizes the rapid emergence and expansion of the Verse Authentication and Classification task—driven by the urgent societal need to counter the proliferation of digital misinformation and unauthorized textual alterations. Nearly absent before 2022, this line of research witnessed an explosive growth by 2025, with numerous studies leveraging transformer-based architectures for high-accuracy verification and textual integrity checking [47], [66].

In parallel, Recitation and Tajwīd Analysis evolved into a substantial and pedagogically grounded research stream, reflecting the field's responsiveness to educational applications and oral tradition. This is evident in the steady publication output focusing on acoustic modeling, phonetic segmentation, and error detection throughout the decade. Numerous studies, whose methodological details are provided in the Appendix, have employed techniques ranging from HMMs and SVMs to advanced Deep Neural Networks (DNNs) to automate reciter identification and Tajwīd rule verification [67]–[80]. Simultaneously, computational efforts extended to varied text analysis tasks, including morphology, clustering, and evaluating translation coherence [81]–[110].

Collectively, this functional reorientation demonstrates the field's transition from a theoretical and exploratory exercise to a mature, application-driven science, characterized by task diversity, user-centric design, and tangible societal relevance.

4.2. The Methodological Ascent: A Story of Adaptation and Ethical Engineering

The functional diversification of research tasks was inseparably linked to—and often propelled by—a rapid methodological ascent up the ladder of technological sophistication. As visualized in the stacked bar chart in Figure 3, this trajectory reveals a clear succession of dominant computational paradigms. The data captures the gradual decline of early statistical methods and static embeddings in favor of two decisive pivots: the overwhelming dominance of “Transformer-based” models post-2022 and the recent emergence of “LLMs & RAG,” underscoring the field's transition toward greater technological sophistication.

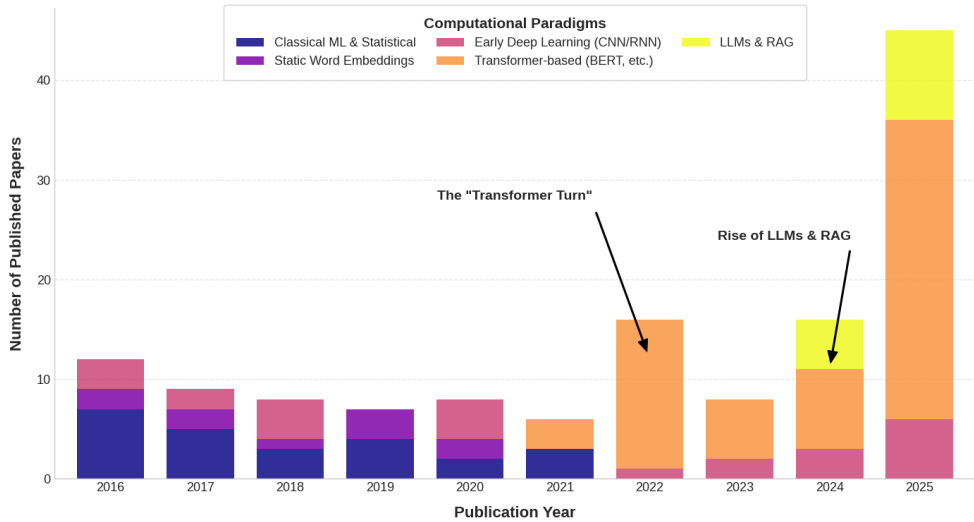


Figure 3: A decade of methodological shifts in computational Qur'anic studies

The early years (2016–2019) were dominated by Classical Machine Learning and Statistical Models, as evidenced by foundational works employing Neural Networks, SVMs, and HMMs [32], [68], [83]. The subsequent introduction of Static Word Embeddings represented a transitional phase toward more nuanced semantic modeling and contextual vectorization [24], [111].

The most profound paradigm shift—clearly observable in Figure 3's post-2020 transformation—was the Transformer Turn. Pre-trained transformer models rapidly became the dominant methodological backbone, demonstrating their superior capacity for contextual, syntactic, and discourse-level understanding [34], [35], [44], [48].

The most consequential innovation, however, has been the integration of Large Language Models (LLMs) through the architectural lens of Retrieval-Augmented Generation (RAG). The field's swift yet deliberate embrace of RAG (2024–2025) can be described as an act of ethical engineering—a purposeful design strategy to mitigate the risk of model hallucination, a failure mode that carries profound theological and epistemological implications in Qur'anic research [9], [49], [62], [108], [112].

This trajectory thus reflects more than a technical modernization; it embodies a principled evolution—one where technological advancement and theological accountability are not opposing forces but mutually reinforcing imperatives.

4.3. The Infrastructure–Innovation Nexus: A Virtuous Cycle of Community and Code

The field’s rapid and sustained maturation was fundamentally enabled by a virtuous cycle linking research innovation with the progressive development of shared community infrastructure. This transformation—from a “resource desert” to a thriving, data- rich ecosystem—is visualized in Figure 4, which plots the trajectory of key milestone events over the decade. By distinguishing between the publication of foundational papers, the release of critical datasets, and the launch of community-wide shared tasks, the timeline illustrates how collaborative, open-science initiatives have acted as the primary accelerators of methodological and conceptual advancement.

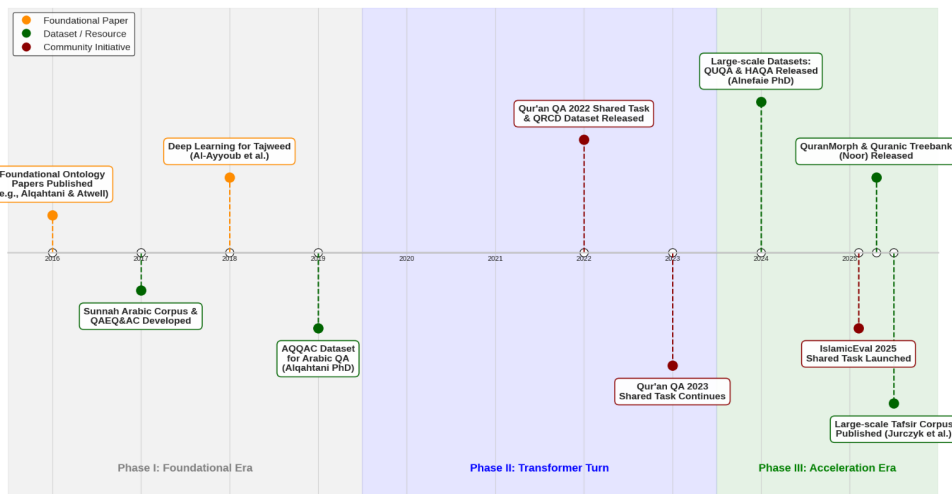


Figure 4: Timeline of key infrastructural milestones and community initiatives (2016–2025)

While early foundational resources such as the Quranic Arabic Corpus provided an indispensable starting point, the subsequent establishment of community-driven shared tasks—most notably the Qur’an QA (2022, 2023) and IslamicEval (2025) workshops—proved to be transformative catalysts. These initiatives, prominently marked on the timeline, not only standardized methodological baselines but also created a central research nexus by providing curated, openly accessible datasets such as QAEQ&AC [113], AQQAC [114], QRCD [8], and QUQA [9].

The direct impact of these infrastructural investments is empirically captured in Figure 5. This heatmap provides a quantitative snapshot of the research landscape (2022–2025) by cross-tabulating publication frequency at the intersection of primary research tasks and dominant technologies. The visualization reveals a distinct epicenter of activity: the intense concentration of studies combining Question Answering (QA) with Transformer-based and RAG architectures highlights the measurable legacy of community-driven benchmarks. Moreover, the subsequent development of highly specialized, large-scale resources in 2025, such as the QuranMorph corpus [10], the Extended Quranic Treebank [115], and the Comprehensive Tafsīr Corpus [65], demonstrates that the field has now entered a phase of infrastructural self-sufficiency and epistemic maturity.

This symbiotic relationship between code and community, resource and research, has become the engine of cumulative innovation—driving not only faster progress but also higher standards of reproducibility, interoperability, and collective scholarly accountability.

4.4. The Global and Collaborative Fabric of a Modern Scholarly Field

Finally, it is essential to recognize that computational Qur’anic studies has evolved into an inherently global, interdisciplinary, and collaborative research community. This geographical dispersion is vividly illustrated in the choropleth map in Figure 6, which tracks scholarly activity based on the first author’s affiliation. The visualization highlights a shift away from centralization, revealing vibrant research hubs indicated by color intensity. Primary concentrations are evident in Saudi Arabia and Malaysia, followed by significant contributions from Jordan, the United Kingdom, Pakistan, and Indonesia. This distribution confirms the field’s transnational and multicultural character, with each region contributing a consistent stream of publications [6], [48], [51], [116].

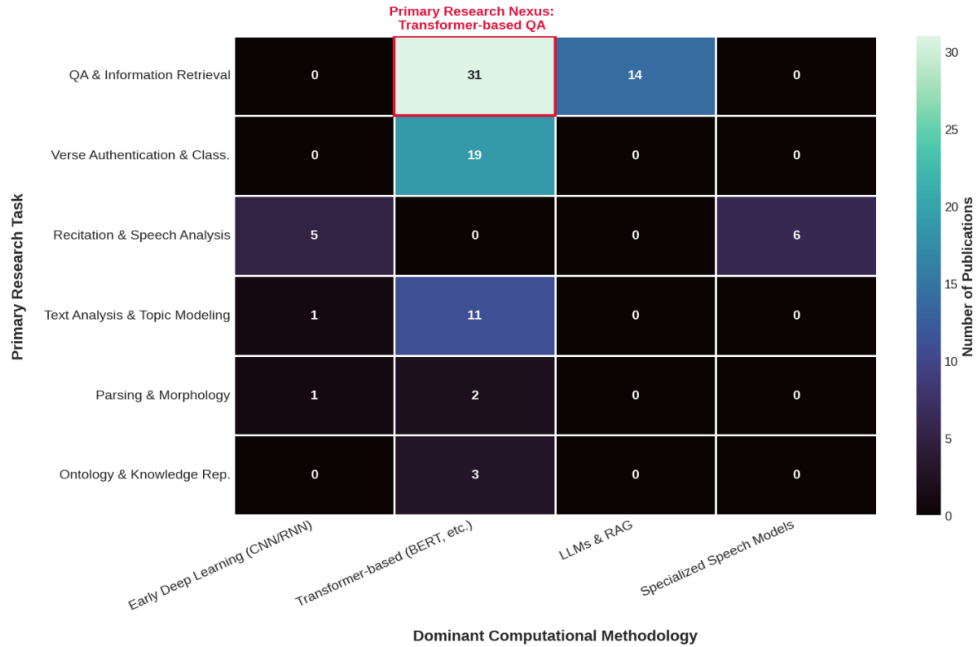


Figure 5: Task–technology concentration matrix for Qur’anic NLP research (2022–2025)

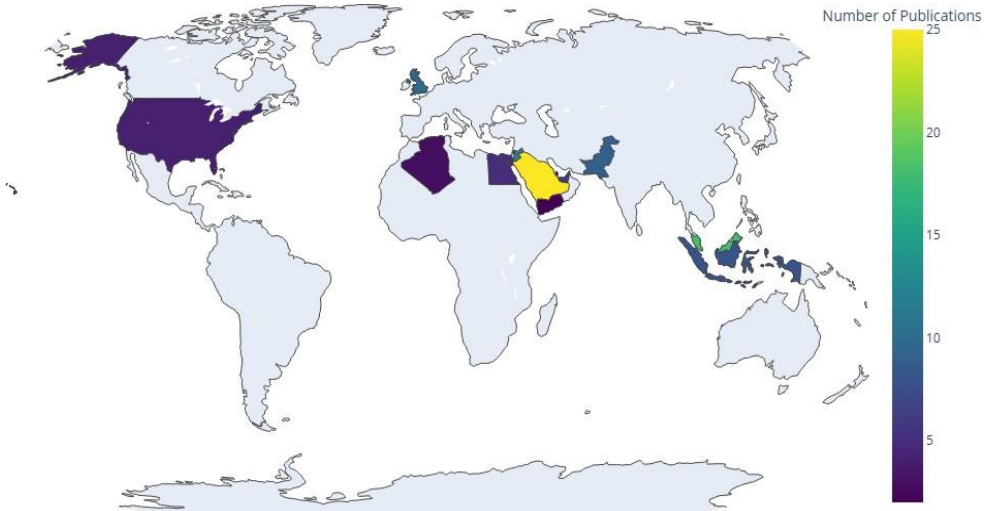


Figure 6: Geographic distribution of primary research publications by first author affiliation (2016–2025)

This geographic diversity is mirrored in the field's collaborative fabric, as visualized in the network graph in Figure 7. In this model, node size corresponds to publication volume, while edge thickness indicates the strength of inter-institutional collaboration. The topology reveals a healthy and increasingly interconnected ecosystem, identifying key bridging institutions that connect otherwise distinct regional clusters. Within this framework, expertise and methodologies circulate across borders, with strong clusters emerging around multi-institutional teams participating in shared tasks. Large-scale resource projects, such as that of Nashir et al. [117], further exemplify the depth and durability of these global partnerships.

This distributed and cooperative structure is arguably the field's greatest strength. It ensures that the evolution of computational Qur'anic studies is enriched by a plurality of intellectual traditions, grounded in both technological innovation and cross-cultural dialogue. As such, the field is not merely expanding in scale—it is maturing into a globally networked discipline, capable of advancing responsible, inclusive, and culturally informed AI research.

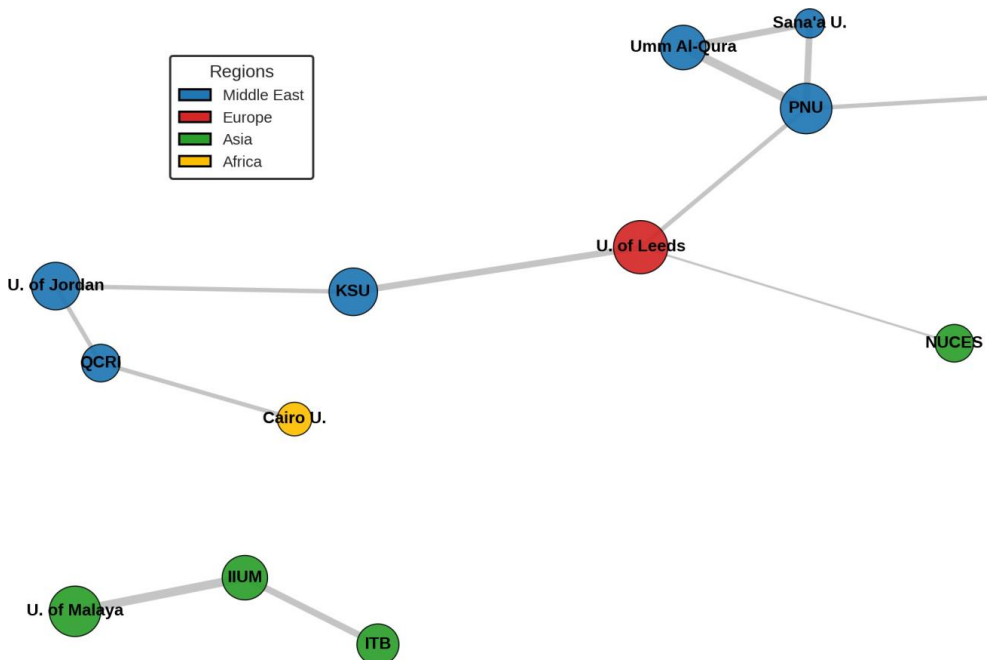


Figure 7: Inter-institutional collaboration network and co-authorship linkages

5. Discussion and Synthesis: From Nascent Field to Emerging Discipline

Our decade-long survey reveals a field undergoing a profound epistemic transformation, evolving from a collection of fragmented, exploratory efforts into the incipient stages of a cohesive and theoretically self-aware scientific discipline. The data presented in the preceding sections does not merely document an increase in publication volume; it narrates a story of conceptual maturation, infrastructural consolidation, and growing intellectual sophistication in addressing the unique methodological and theological challenges posed by a sacred text. This section synthesizes the overarching findings, critically examines the symbiotic interplay between technological innovation and theological prudence, and delineates the structural gaps and emergent frontiers that will shape the next phase of research in computational Qur’anic studies.

5.1. The Trajectory of a Maturing Field: Three Eras, Two Pivots

The evolutionary arc of computational Qur’anic studies, as charted in Section 2, does not follow a path of linear accumulation but one of punctuated equilibrium—a series of qualitative leaps defined by two critical pivots.

The initial Foundational Era (2016–2019) was characterized by a “structuralist” orientation, where the primary challenge lay in imposing machine-readable order upon the Qur’anic text through ontological modeling (Figure 2). The field was resource-poor, methodologically cautious, and its contributions were largely proof-of-concept prototypes exploring feasibility rather than performance.

The first pivot, known as the “Transformer Turn” (c. 2020–2023), represented more than a technological upgrade; it was a conceptual and infrastructural realignment, catalyzed by the establishment of community-driven shared tasks such as Qur’an QA (Figure 4). For the first time, the field possessed a unified research agenda, shared datasets, and standardized benchmarks. As transformer-based architectures rapidly became dominant (Figure 3), the research question evolved from “Can we process this text?” to “How effectively can we perform a specific, measurable task?”—signaling a new era of methodological rigor, comparability, and reproducibility.

The second pivot, defining the Acceleration Era (2024–2025), was the integration of Large Language Models (LLMs). This period signifies the point at which the community transcended the role of technological consumer to

become an active innovator, adapting and re-engineering LLMs to meet the ethical, linguistic, and theological specificities of Qur'anic scholarship. The convergence of research around specific task–technology pairings, as visualized in Figure 5, demonstrates the emergence of mature sub-domains with shared best practices—particularly the nexus of transformer-based architectures for Question Answering and Verse Authentication. Furthermore, the global and collaborative structure of this work, evidenced in the geographic and institutional networks (Figures 6 & 7), affirms the field's transition into a modern, international research community defined by openness, diversity, and interdisciplinary exchange.

5.2. Theological Prudence as a Catalyst for Innovation: The Symbiosis of Ethics and Engineering

Perhaps the most profound insight from our survey is that the unique theological constraints of the Qur'anic text have served not as limitations, but as drivers of innovation. In a domain where interpretive error carries not just analytical but moral and epistemic weight, the pursuit of fidelity, traceability, and ethical soundness has compelled researchers to move beyond generic AI pipelines and engage in domain-aware, ethically grounded system design.

This is most evident in the community's measured response to LLMs. While other fields grappled with the risks of model hallucination, Qur'anic AI research rapidly coalesced around Retrieval-Augmented Generation (RAG) as a non-negotiable architectural paradigm [9]. RAG here represents not just a technical enhancement, but an act of ethical engineering—a deliberate safeguard ensuring that AI-generated content remains anchored in verifiable, authoritative sources. This alignment between computational design and theological epistemology exemplifies the field's emerging maturity.

Similarly, the growing interest in Explainable AI (XAI) is not a peripheral academic curiosity [11], but a structural necessity in a scholarly tradition where the chain of reasoning (*isna'd*) and methodological transparency are paramount. The demand for systems that can “show their interpretive work” is fostering innovations in model interpretability, reasoning visualization, and justification tracing. In this sense, the ethical and theological imperatives of the domain are actively propelling the field toward more responsible, transparent, and trustworthy AI systems—demonstrating that in Qur'anic studies, ethics is not an afterthought but an engine of progress.

5.3. Persistent Gaps and the Frontiers of Future Research

Despite remarkable progress, the field remains defined by several systematic and structural gaps. Our analysis identifies four critical frontiers demanding urgent attention in the coming decade:

- **The Linguistic Monolith: Neglect of Qira`a`at and Multilingualism.** The vast majority of research treats the Qur`an as linguistically homogeneous, focusing almost exclusively on the H. afs. (an (A` s.im recitation and on Arabic or English texts. The rich diversity of canonical recitation traditions (Qira`a`at)—which encode subtle but meaningful linguistic, phonological, and exegetical variations—remains virtually untouched computationally, representing a major scholarly lacuna. Furthermore, with over 80% of Muslims being non-native Arabic speakers, the paucity of multilingual and cross-lingual studies (beyond pioneering work in Urdu) underscores a missed opportunity to serve the global Muslim community. The development of multi-dialectal, multilingual parallel corpora thus stands as an urgent research priority.
- **The Hermeneutic Divide: Lack of Deep Integration with Tafs`ir.** While the compilation of the Tafs`ir Corpus [65] is a monumental infrastructural milestone, most computational systems still operate in interpretive isolation, detached from the Qur`an's fourteen-century hermeneutic continuum. Current QA systems can retrieve verses but cannot contextualize or reason about exegetical diversity. The next frontier lies in developing models that can represent, compare, and synthesize arguments across classical and modern tafs`ir—a pursuit that might rightly be termed computational hermeneutics, and which represents the grand intellectual challenge of the field.
- **The "Black Box" Problem: Deficit in Evaluation and Explainability.** The field currently lacks a comprehensive benchmark suite analogous to GLUE or SuperGLUE in general NLP. Evaluation remains task-specific, rarely addressing theological coherence, interpretive fidelity, or potential for semantic distortion. This is compounded by the scarcity of robust XAI frameworks. Without standardized evaluation metrics and transparent interpretability mechanisms, system outputs will struggle to gain institutional trust and scholarly legitimacy.
- **The Untouched Depths: Absence of Rhetorical and Discourse Analysis.** While significant progress has been achieved at the lexical and verse levels, the higher-order literary and rhetorical architecture (bala`ghah) of the Qur`an remains largely uncharted. Computational exploration of discourse coherence, argumentation structure, intertextuality (tana`s. s.), and rhetorical devices represents a vast and cognitively rich frontier.

Addressing these challenges will require discourse-aware architectures, multi-layer annotation frameworks, and deep collaboration with classical Arabic rhetoricians and literary scholars.

Collectively, these challenges underscore that computational Qur'anic studies is entering a phase where technical sophistication must be matched by interpretive depth. The coming decade will likely define whether the field consolidates as a distinct interdisciplinary discipline—one capable of integrating AI innovation with theological integrity and of setting global standards for the responsible datafication of sacred knowledge.

6. Future Research Directions: A Strategic Roadmap for the Next Decade

Our systematic survey of the past decade has led to three definitive conclusions: first, the field has successfully migrated from static ontology-based methods to dynamic transformer-based architectures; second, progress has been uneven, heavily favoring Question Answering tasks while neglecting the linguistic complexity of canonical recitations (Qira'at); and third, the emergence of LLMs offers unprecedented opportunities that are currently hindered by a lack of domain-specific evaluation benchmarks. Based on these core findings, and recognizing that the field now stands at the precipice of its most significant challenges, we propose a strategic roadmap. This agenda is organized into three interconnected horizons designed to bridge the identified gaps: solidifying the foundational infrastructure, linking AI with classical scholarship, and developing transformative societal applications.

6.1. Solidifying the Foundations: Infrastructure and Domain-Specific Models

The immediate future of the field hinges on addressing its most critical bottleneck: the scarcity of large-scale, domain-specific resources. While recent progress has been remarkable, it has been built on a surprisingly narrow infrastructural base.

The overwhelming reliance on models pre-trained on Modern Standard Arabic (MSA), for instance, remains a significant limitation, as evidenced by the variable performance of models like AraBERT in tasks requiring deep stylistic understanding [66]. Therefore, a primary short-term priority must be the development of foundation models pre-trained specifically on a massive corpus of Classical Arabic (CA). Recent breakthroughs in constructing large-

scale annotated literary corpora, such as the 14-million-verse “Diwan” corpus [118], demonstrate the feasibility of aggregating massive CA datasets essential for training these next-generation domain-specific models.

Concurrently, the community must build on the proven success of shared tasks [8] by expanding and standardizing its evaluation benchmarks. The current focus on Question-Answering, while fruitful, leaves other critical areas under-evaluated. Collaborative efforts are urgently needed to launch new shared tasks and public benchmarks for Quranic Semantic Textual Similarity (Q-STs), multilingual QA, and, crucially, automatic Tajweed rule detection, an area that has seen renewed interest but lacks a standard evaluation corpus [80]. Furthermore, research must move beyond its current monolithic treatment of the Quranic text. The near-total neglect of the canonical recitation traditions (Qiraʿat) represents a major scholarly lacuna. The creation of a multi-Qiraʿat parallel corpus is a challenging but essential next step to enable the first large-scale computational studies of these variations.

6.2. Building Bridges to Classical Scholarship: Towards Computational Hermeneutics

With a more robust foundation in place, the field’s medium-term ambition should be to build deeper, more meaningful bridges to the classical Islamic scholarly tradition. This requires moving from surface-level text processing to the computational modeling of hermeneutics. A key objective in this horizon is the construction of a large-scale, integrated Islamic Knowledge Graph. Pioneering efforts like SemanticTafsir have demonstrated the feasibility of modeling a single exegetical work [119], but the next step is to create a unified graph that connects the Qur’an to the Hadith, and critically, to the arguments and interpretations of multiple, diverse schools of tafsīr. Such a resource would transform the field, enabling researchers to computationally trace the evolution of theological concepts and legal rulings across centuries of scholarship.

This infrastructural development would, in turn, enable the pursuit of the field’s grand challenge: computational hermeneutics. This does not imply creating an “Al Mufassir” to replace human scholars, but rather developing sophisticated Tafsīr Assistance Tools. These systems would support, rather than supplant, traditional scholarship by automatically synthesizing exegetical opinions on a given verse, identifying relevant intertextual connections across the vast Islamic corpus, and modeling the complex rhetorical structures

(balaḡhah) that are central to Quranic meaning. In this context, recent frameworks developed for the multi-label thematic classification of Classical Arabic poetry, such as the Maqasid architecture [120], offer a transferable methodological template for capturing the layered semantic and rhetorical dimensions of the Classical Arabic text. As these systems grow in complexity, the need for transparency becomes paramount.

6.3. A Long-Term Vision: Responsible AI for Global Communities

The long-term vision for computational Quranic studies should focus on translating these advanced capabilities into robust, accessible, and ethically-grounded applications. The ultimate goal is the creation of advanced, multimodal conversational AI systems for personalized Quranic education. These systems would move far beyond simple QA, acting as interactive tutors that can answer complex questions by citing and comparing multiple scholarly sources, provide real-time feedback on a student's recitation [50], and adapt learning pathways to individual needs.

Achieving this vision responsibly requires the establishment of robust ethical and governance frameworks, a research area in its own right. As AI's role in mediating religious experience grows, the community must proactively collaborate with Islamic legal scholars (fuqaha') and ethicists to address critical questions of interpretive authority, algorithmic bias, and data sovereignty. Finally, as the field matures, it will be uniquely positioned to contribute to broader cross-cultural and interfaith understanding. With powerful multilingual models and integrated knowledge graphs, long-term projects could undertake computational comparative analysis of themes, narratives, and legal concepts across the Qur'an and other sacred texts, creating data-driven tools to support and enrich interfaith dialogue in an increasingly interconnected world.

7. Conclusion

This systematic, decade-long survey has charted the remarkable maturation of computational Quranic studies from a collection of fragmented, proof-of-concept experiments into a cohesive, benchmark-driven scientific discipline. Our analysis of 112 primary studies reveals a distinct three-act narrative: a Foundational Era (2016–2019) preoccupied with knowledge structuring, a pivotal "Transformer Turn" (2020–2023) that consolidated the field around shared benchmarks, and a recent Acceleration Era (2024–2025) defined by the integration of Large Language Models. Crucially, this evolution was not

merely a passive adoption of mainstream AI trends. We find that the unique epistemological and ethical demands of the Quranic text have acted as a powerful catalyst for innovation, compelling the field to pioneer solutions in areas such as Retrieval-Augmented Generation (RAG) to ensure theological fidelity and to explore Explainable AI (XAI) to build scholarly trust. The field's trajectory thus offers a compelling case study in how domain-specific constraints can drive, rather than hinder, technological advancement.

Despite this progress, the field stands at a new frontier, facing challenges that are no longer purely technical but deeply hermeneutic and cross-cultural. The significant lacunae in multilingual applications, the computational analysis of canonical recitation traditions (Qira'`a`-t), and the deep integration of classical exegetical scholarship (tafs`ir) represent the major scholarly challenges for the next decade. Ultimately, the journey of computational Quranic studies over the past decade is a story of a community building not just tools, but also the intellectual and ethical frameworks to wield them responsibly. The challenge moving forward is to continue this trajectory—to build systems that do not merely process text, but that augment human understanding, respect centuries of scholarly tradition, and serve a global community's engagement with its foundational scripture in a manner that is both intellectually rigorous and theologically sound.

References

- [1] S. E. Jones, *The Emergence of the Digital Humanities*. Routledge, 2021.
- [2] Pew Research Center, "The global religious landscape: A report on the size and distribution of the world's major religious groups as of 2010," Pew Research Center, 2012.
- [3] A. Sinno, *Muslims in Western Politics*. Indiana University Press, 2022.
- [4] M. H. Bashir et al., "Arabic natural language processing for Qur'anic research: A systematic review," *Artificial Intelligence Review*, vol. 56, no. 7, pp. 6801–6854, 2023.
- [5] K. Shahzad and A. Ashraf, "Artificial intelligence (AI) and natural language processing (NLP) for religious texts—A systematic mapping study," *Journal of Integrated Sciences*, vol. 5, no. 4, 2025.
- [6] M. Alqahtani and E. Atwell, "Arabic Quranic search tool based on ontology," in *Proc. 21st Int. Conf. on Applications of Natural Language to Information Systems (NLDB 2016)*, 2016, pp. 478–485.

- [7] A. S. M. Rusli, F. Ridzuan, Z. M. Zaki, M. Sayuti, and M. S. Abdullah, "A systematic review on semantic-based ontology for Quranic knowledge," *International Journal of Engineering & Technology*, vol. 7, no. 2.29, pp. 151–157, 2018.
- [8] R. Malhas, W. Mansour, and T. Elsayed, "Qur'an QA 2022: Overview of the first shared task on question answering over the Holy Qur'an," in *Proc. 5th Workshop on Open-Source Arabic Corpora and Processing Tools, 2022*, pp. 79–87.
- [9] S. S. M. Alnefaie, "Artificial intelligence for answering questions from the Holy Quran and Hadith," Ph.D. dissertation, Univ. of Leeds, 2024.
- [10] D. Akra, T. Hammouda, and M. Jarrar, "QuranMorph: Morphologically annotated Quranic corpus," *arXiv preprint arXiv:2506.18148*, 2025.
- [11] A. M. Mustafa, S. Nakhleh, and R. Irsheidat, "Interpreting Arabic transformer models: A study on XAI interpretability for Qur'anic semantic-search models," *Jordanian Journal of Computers and Information Technology*, vol. 10, no. 3, pp. 312–327, 2024.
- [12] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," *Keele Univ. and Durham Univ. Joint Report*, 2007.
- [13] K. Petersen, S. Vakkalanka, and L. Kuzniarz, "Guidelines for conducting systematic mapping studies in software engineering: An update," *Information and Software Technology*, vol. 64, pp. 1–18, 2015.
- [14] C. Wohlin, "Guidelines for snowballing in systematic literature studies and a replication in software engineering," in *Proc. 18th Int. Conf. on Evaluation and Assessment in Software Engineering, 2014*, Art. 38, pp. 1–10.
- [15] A. Ta'a, Q. A. Abed, B. M. Ali, and M. Ahmad, "Ontology-based approach for knowledge retrieval in Al-Quran holy book," *International Journal of Computer Engineering Research*, vol. 6, no. 3, pp. 8–15, 2016.
- [16] R. Ahmed and E. Atwell, "Developing an ontology of concepts in the Qur'an," *International Journal on Islamic Applications in Computer Science and Technology*, vol. 4, no. 4, pp. 1–8, 2016.
- [17] M. F. Noordin, T. M. T. Sembok, R. Othman, and R. H. Gusmita, "Constructing an ontology-based and graph-based knowledge representation of English Quran," *Jurnal Teknologi (Sciences & Engineering)*, vol. 78, no. 8-2, pp. 11–18, 2016.

- [18] A. B. M. S. Sadi, T. Anam, M. Abdirazak, and F. Anan, "Applying ontological modeling on Quranic "Nature" domain," in 2016 7th Int. Conf. on Information and Communication Systems (ICICS), 2016, pp. 151–155.
- [19] A. Ta'a, Q. A. Abed, and M. Ahmad, "Al-Quran ontology based on knowledge themes," *Journal of Fundamental and Applied Sciences*, vol. 9, no. 5S, pp. 800–817, 2017.
- [20] M. M. Alqahtani and E. Atwell, "Developing bilingual Arabic-English ontologies of Al-Quran," in 2018 IEEE 2nd Int. Workshop on Arabic and Derived Script Analysis and Recognition (ASAR), 2018, pp. 64–69.
- [21] A. Daud, M. H. Ullah, A. R. Banjar, and A. A. Alshdadi, "Ontological modeling and semantic search in Quran," *International Journal of Computer Science and Network Security*, vol. 22, no. 5, pp. 771–779, 2022.
- [22] A. Ta'a et al., "Semantic approach for Al-Quran knowledge representation in Islamic knowledge management framework," *Journal of Quranic Studies Research*, vol. 6, no. 2, pp. 66–76, 2021.
- [23] N. Basir et al., "Retrieval performance for USIM's Quranic search engine," *Journal of Engineering and Applied Sciences*, vol. 13, Special Issue 11, pp. 8894–8899, 2018.
- [24] H. Z. Rad, S. Tiun, and S. Saad, "Lexical scoring system of lexical chain for Quranic document retrieval," *GEMA Online Journal of Language Studies*, vol. 18, no. 2, pp. 59–78, 2018.
- [25] F. S. Utomo, Y. Purwati, M. S. Azmi, L. Shafira, and A. Mudzakir, "Word embedding and imbalanced learning impact on Indonesian Quran ontology population," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 21, no. 2, pp. 156–165, 2025.
- [26] S. A. Pitchay and F. Ridzuan, "A systematic review analysis for Quran verses retrieval," *Journal of Engineering and Applied Sciences*, vol. 11, Special Issue, pp. 629–634, 2016.
- [27] S. A. Salloum, A. Q. AlHamad, M. Al-Emran, and K. Shaalan, "A survey of Arabic text mining," in *Intelligent natural language processing: Trends and applications*, Springer, Cham, 2017, pp. 417–431.
- [28] D. Suleiman, A. Awajan, and W. Al Etaiwi, "The use of Hidden Markov Model in natural Arabic language processing: A survey," *Procedia Computer Science*, vol. 113, pp. 240–247, 2017.
- [29] A. M. Azmi, A. O. Al-Qabbany, and A. Hussain, "Computational and natural language processing based studies of hadith literature: A survey," *Artificial Intelligence Review*, vol. 52, no. 2, pp. 1369–1405, 2019.

- [30] K. Darwish et al., "A panoramic survey of natural language processing in the Arab world," *Communications of the ACM*, vol. 64, no. 4, pp. 72–81, 2021.
- [31] Q. Bsoul, R. Abdul Salam, J. Atwan, and M. Jawarneh, "Arabic text clustering methods and suggested solutions for theme-based Quran clustering: Analysis of literature," *Journal of Information Science Theory and Practice*, vol. 9, no. 4, pp. 15–38, 2021.
- [32] S. K. Hamed and M. J. Ab Aziz, "A question answering system on Holy Quran translation based on question expansion technique and neural network classification," *Journal of Computer Science*, vol. 12, no. 4, pp. 169–177, 2016.
- [33] W. Antoun, F. Baly, and H. Hajj, "AraBERT: Transformer-based model for Arabic language understanding," in *Proc. 4th Workshop on Open-Source Arabic Corpora and Processing Tools*, 2020, pp. 9–15.
- [34] M. ElKomy and A. M. Sarhan, "TCE at Qur'an QA 2022: Arabic language question answering over Holy Qur'an using a post-processed ensemble of BERT-based models," in *Proc. 5th Workshop on Open-Source Arabic Corpora and Processing Tools*, 2022, pp. 154–161.
- [35] D. Premasiri, T. Ranasinghe, W. Zaghouni, and R. Mitkov, "DTW at Qur'an QA 2022: Utilising transfer learning with transformers for question answering in a low-resource domain," in *Proc. 5th Workshop on Open-Source Arabic Corpora and Processing Tools*, 2022, pp. 88–95.
- [36] E. Aftab and M. K. Malik, "eRock at Qur'an QA 2022: Contemporary deep neural networks for Qur'an based reading comprehension question answers," in *Proc. 5th Workshop on Open-Source Arabic Corpora and Processing Tools*, 2022, pp. 96–103.
- [37] B. Ahmed, M. Saad, and E. A. Refaee, "QQATeam at Qur'an QA 2022: Fine-tuning Arabic QA models for Qur'an QA task," in *Proc. 5th Workshop on Open-Source Arabic Corpora and Processing Tools*, 2022, pp. 130–135.
- [38] A. Alsaleh et al., "LK2022 at Qur'an QA 2022: Simple transformers model for finding answers to questions from Qur'an," in *Proc. 5th Workshop on Open-Source Arabic Corpora and Processing Tools*, 2022, pp. 120–125.
- [39] A. Keleg and W. Magdy, "SMASH at Qur'an QA 2022: Creating better faithful data splits for low-resourced question answering scenarios," in *Proc. 5th Workshop on Open-Source Arabic Corpora and Processing Tools*, 2022, pp. 136–145.

- [40] Y. Mellah et al., "LARSA22 at Qur'an QA 2022: Text-to-text transformer for finding answers to questions from Qur'an," in Proc. 5th Workshop on Open-Source Arabic Corpora and Processing Tools, 2022, pp. 112–119.
- [41] A. Mostafa and O. Mohamed, "GOF at Qur'an QA 2022: Towards an efficient question answering for the Holy Qu'ran in the Arabic language using deep learning-based approach," in Proc. 5th Workshop on Open- Source Arabic Corpora and Processing Tools, 2022, pp. 104–111.
- [42] N. Singh, "niksss at Qur'an QA 2022: A heavily optimized BERT based model for answering questions from the Holy Qu'ran," in Proc. 5th Workshop on Open-Source Arabic Corpora and Processing Tools, 2022, pp. 126–129.
- [43] A. Sleem et al., "Stars at Qur'an QA 2022: Building automatic extractive question answering systems for the Holy Qur'an with transformer models and releasing a new dataset," in Proc. 5th Workshop on Open-Source Arabic Corpora and Processing Tools, 2022, pp. 146–153.
- [44] S. Alnefaie, A. Alsaleh, E. Atwell, M. Alsalka, and A. Altahhan, "LKAU23 at Qur'an QA 2023: Using transformer models for retrieving passages and finding answers to questions from the Qur'an," in Proc. ArabicNLP 2023, 2023, pp. 720–727.
- [45] M. A. Ateeq, S. Tiun, H. Abdelhaq, and N. Rahhal, "Arabic narrative question answering (QA) using transformer models," IEEE Access, vol. 12, pp. 2760–2777, 2023.
- [46] M. R. Rizqullah, A. Purwarianti, and A. F. Aji, "QASiNa: Religious domain question answering using Sirah Nabawiyah," in 2023 10th Int. Conf. on Advanced Informatics: Concept, Theory and Application (ICAICTA), 2023, pp. 1–6.
- [47] Z. Touati-Hamad, M. R. Laouar, I. Bendib, and S. Hakak, "Arabic Quran verses authentication using deep learning and word embeddings," The International Arab Journal of Information Technology, vol. 19, no. 4, pp. 681–688, 2022.
- [48] M. H. K. Alshammeri, "Deep learning and distributional semantics for the Qur'an," Ph.D. dissertation, Univ. of Leeds, 2022.
- [49] Z. Saadaoui, G. Tlig, and F. Jarray, "LLMs based approach for Quranic question answering," in Proc. 20th Int. Conf. on Web Information Systems and Technologies (WEBIST 2024), 2024, pp. 397–404.
- [50] S. Alfadhli, H. Alharbi, and A. Cherif, "qArl: A hybrid CTC/attention-based model for Quran recitation recognition using bidirectional LSTM in an end-to-end architecture," IEEE Access, vol. 12, pp. 90274–90289, 2024.

- [51] A. Zafar, M. Wasim, S. Zulfiqar, T. Waheed, and U. Raza, "Transformer-based topic modeling for Urdu translations of the Holy Quran," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 23, no. 9, Art. 196, 2024.
- [52] A. Abbasi, M. J. Awan, and M. Ismail, "A deep learning approach to semantic clarity in Urdu translations of the Holy Quran," *Linguistica Antverpiensia*, 2025, Art. no. 389676054.
- [53] M. A. A. Danish Khaleeq, M. Tariq, and J. Iqbal, "Machine translation of Quranic verses: A transformer-based approach to Urdu rendering," *International Journal of Islamic Studies and Technology*, 2025.
- [54] K. Gaanoun and M. Alsuhaibani, "Sentiment preservation in Quran translation with artificial intelligence approach: Study in reputable English translation of the Quran," *Humanities and Social Sciences Communications*, vol. 12, no. 1, pp. 1–15, 2025.
- [55] M. Tariq, M. A. Awan, and D. Khaleeq, "Developing a Quranic QA system: Bridging linguistic gaps in Urdu translation using NLP and transformer model," *Linguistica Antverpiensia*, 2025.
- [56] Y. Aftab, M. A. Awan, D. Khaleeq, and T. Ismail, "Enhancing Quranic ethics and morality: An NLP-based semantic search model for Urdu translation," *International Journal of Islamic Studies and Technology*, 2025.
- [57] U. Shahid, M. Z. Hussain, and W. Sayers, "Computational analysis of Quran text using machine learning and large language models," in *2025 8th Int. Conf. on Data Science and Machine Learning Applications (CDMA)*, 2025, pp. 18–24.
- [58] A. B. Kamran, N. A. Butt, and A. Basharat, "Semantic enrichment of Hadith corpus: Knowledge graph generation from Islamic text," *Semantic Web Journal*, vol. 15, no. 4, pp. 1247–1276, 2024.
- [59] S. Aftar, A. El Ganadi, L. Gagliardelli, F. Ruozi, and S. Bergamaschi, "RoBERT2VecTM: A novel approach for topic extraction in Islamic studies," in *Findings of the Association for Computational Linguistics: EMNLP 2024*, 2024, pp. 9112–9123.
- [60] M. Asgari-Bidhendi, M. A. Ghaseminia, and A. Rezaei, "Rezwan: Leveraging large language models for comprehensive hadith text processing: A 1.2 M corpus development," *arXiv preprint arXiv:2510.03781*, 2025.
- [61] A. Omayrah, S. Alkhereyf, A. Abdelali, and H. Hassan, "HUMAIN at IslamicEval 2025 shared task 1: A three-stage LLM-based pipeline for detecting and correcting hallucinations in Quran and Hadith," in *Proc. IslamicEval @ ArabicNLP 2025*, 2025.

- [62] A. Al Adel, A. B. Soliman, M. S. Sawan, R. Al-Najjar, and A. Ashraf, "BurhanAl at IslamicEval 2025 shared task: Combating hallucinations in LLMs for Islamic content; Evaluation, correction, and retrieval-based solution," in Proc. IslamicEval @ ArabicNLP 2025, 2025.
- [63] A. Salah, M. Mahdi, and A. A. Badawy, "Exploring thematic structures in Surah Al-Baqarah using TF-IDF, dimensionality reduction, and K- means clustering," *Journal of Computing and Communication Systems*, vol. 1, no. 1, pp. 15–28, 2025.
- [64] A. N. An, F. Q. A. Tamami, Z. Daud, and N. M. Salleh, "Understanding the integration of deep learning and artificial intelligence in Quranic education and research through bibliometric analysis," *Educational Process: International Journal*, vol. 14, no. 1, pp. 45–62, 2025.
- [65] T. Jurczyk, R. Seidel, A. Bernhard, T. Scheffler, and A. El-Sayed, "Text mining Tafsir: Compilation and preliminary explorations of a curated corpus of 80 Qur'anic commentaries," *Journal of Digital Islamic Research*, vol. 3, no. 1, pp. 97–125, 2025.
- [66] K. A. Alam et al., "Automated authentication of Quranic verses using BERT (Bidirectional Encoder Representations from Transformers) based language models," in Proc. New Horizons in Computational Linguistics for Religious Texts (Coling-Rel), 2025, pp. 59–66.
- [67] T. M. H. Asda, T. S. Gunawan, M. Kartiwi, and H. Mansor, "Development of Quran reciter identification system using MFCC and neural network," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 1, no. 1, pp. 168–175, 2016.
- [68] A. A. Q. Al-Bakeri and A. A. Basuhail, "ASR for Tajweed rules: Integrated with self-learning environments," *International Journal of Information Engineering and Electronic Business*, vol. 9, no. 6, pp. 1–9, 2017.
- [69] F. Ahmad, S. Z. Yahya, and Z. Saad, "Tajweed classification using artificial neural network," in 2018 Int. Conf. on Computational Approach in Smart Systems Design and Applications (ICASSDA), 2018, pp. 1–6.
- [70] M. Al-Ayyoub, N. A. Damer, and I. Hmeidi, "Using deep learning for automatically determining correct application of basic Quranic recitation rules," *The International Arab Journal of Information Technology*, vol. 15, no. 3A, pp. 620–625, 2018.
- [71] A. M. Alagrami and M. M. Eljazzar, "Smartajweed: Automatic recognition of Arabic Quranic recitation rules," arXiv preprint arXiv:2101.04200, 2020.

- [72] K. M. Nahar, R. M. Al-Khatib, M. A. Al-Shannaq, and M. M. Barhoush, "An efficient Holy Quran recitation recognizer based on SVM learning model," *Jordanian Journal of Computers and Information Technology*, vol. 6, no. 4, pp. 392–406, 2020.
- [73] J. H. Alkhateeb, "A machine learning approach for recognizing the Holy Quran reciter," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 7, pp. 268–271, 2020.
- [74] A. A. Harere and K. A. Jallad, "Mispronunciation detection of basic Quranic recitation rules using deep learning," *arXiv preprint arXiv:2305.06429*, 2023.
- [75] A. A. Harere and K. A. Jallad, "Quran recitation recognition using end- to-end deep learning," *arXiv preprint arXiv:2305.07034*, 2023.
- [76] N. Nigar, A. Wajid, S. A. Ajagbe, and A. Al-omari, "An intelligent framework based on deep learning for online Quran learning during pandemic," *Applied Computational Intelligence and Soft Computing*, vol. 2023, Art. no. 5541699, 2023.
- [77] S. Al-Fadhli, H. Al-Harbi, and A. Cherif, "Speech recognition models for Holy Quran recitation based on modern approaches and Tajweed rules: A comprehensive overview," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 12, pp. 793–807, 2023.
- [78] M. A. Shakeel, H. A. Khattak, and M. Al-Hawawreh, "Deep acoustic modelling for Quranic recitation: Current solutions and future directions," *IPSI BgD Transactions on Internet Research*, vol. 20, no. 1, pp. 45–58, 2024.
- [79] H. M. Osman, B. S. Mustafa, and B. M. Mahmood, "A deep learning approach for recognizing the noon rule for reciting Holy Quran," *Protok: Jurnal Ilmiah Teknik Elektro*, vol. 11, no. 3, pp. 198–205, 2024.
- [80] D. Shaiakhmetov, G. Gimaletdinova, K. Momunov, and A. Rakhimzhanova, "Evaluation of the pronunciation of Tajweed rules based on DNN as a step towards interactive recitation learning," *arXiv preprint arXiv:2503.23470*, 2025.
- [81] E. M. Alsukhni, M. N. Al-Kabi, and I. M. Alsmadi, "An automatic evaluation for online machine translation: Holy Quran case study," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 6, pp. 118–125, 2016.
- [82] M. A. Hamdelsayed and E. Atwell, "Islamic applications of automatic question-answering," *SUST Journal of Engineering and Computer Science (JECS)*, vol. 17, no. 2, pp. 51–57, 2016.

- [83] M. A. Saloot et al., "Hadith data mining and classification: A comparative analysis," *Artificial Intelligence Review*, vol. 46, no. 1, pp. 113–132, 2016.
- [84] S. J. Putra et al., "A semantic-based question answering system for Indonesian translation of Quran," in *Proc. 18th Int. Conf. on Asian Language Processing (IALP)*, 2016, pp. 126–131.
- [85] A. Alosaimy and E. Atwell, "Sunnah Arabic Corpus: Design and methodology," in *Proc. 5th Int. Conf. on Islamic Applications in Computer Science and Technologies (IMAN 2017)*, 2017.
- [86] S. Bougrine, H. Cherroun, and A. Abdelali, "Altruistic crowdsourcing for Arabic speech corpus annotation," *Procedia Computer Science*, vol. 113, pp. 276–283, 2017.
- [87] A. Mahmood, H. U. Khan, and W. Khan, "Query based information retrieval and knowledge extraction using Hadith datasets," in *2017 13th Int. Conf. on Emerging Technologies (ICET)*, 2017, pp. 1–6.
- [88] M. Alhawarat and M. Hegazi, "Revisiting k-means and topic modeling, a comparison study to cluster Arabic documents," *IEEE Access*, vol. 6, pp. 42740–42751, 2018.
- [89] A. Adeleke, N. Samsudin, A. Mustapha, and S. Ahmad Khalid, "Automating Quranic verses labeling using machine learning approach," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 16, no. 2, pp. 1089–1097, 2019.
- [90] E. Atwell, "Using the web to model modern and Qur'anic Arabic," in *Arabic corpus linguistics*, T. McEnery, A. Hardie, and N. Younis, Eds. Edinburgh: Edinburgh University Press, 2019, pp. 100–119.
- [91] S. N. Olimat, "Euphemism in the Qur'an: A corpus-based linguistic approach," *International Journal of Computational Linguistics (IJCL)*, vol. 10, no. 2, pp. 16–32, 2019.
- [92] J. Olsson, "Computing Qur'ans: A suggestion for a digital humanities approach to the question of interrelations between English Qur'an translations," *Journal of Qur'anic Studies*, vol. 21, no. 2, pp. 83–115, 2019.
- [93] N. Varghese and M. Punithavalli, "Lexical and semantic analysis of sacred texts using machine learning and natural language processing," *International Journal of Scientific & Technology Research*, vol. 8, no. 12, pp. 2820–2827, 2019.
- [94] H. J. A. Alshahrani, "The semantic prosody of natural phenomena in the Qur'an: A corpus-based study," Ph.D. dissertation, Univ. of Leeds, 2020.

- [95] I. Tlemsani, F. Marir, and M. Majdalawieh, "Screening of Murabaha business process through Quran and hadith: A text mining analysis," *Journal of Islamic Accounting and Business Research*, vol. 11, no. 9, pp. 1909–1933, 2020.
- [96] Z. Touati-Hamad, M. R. Laouar, and I. Bendib, "Quran content representation in NLP," in *Proc. 10th Int. Conf. on Information Communication and Management*, 2020, pp. 42–46.
- [97] F. S. Utomo, N. Suryana, and M. S. Azmi, "Question answering systems on holy Quran: A review of existing frameworks, approaches, algorithms and research issues," in *Journal of Physics: Conf. Series*, vol. 1501, no. 1, 2020, Art. no. 012022.
- [98] M. L. Arifianto, "Utilizing the Quranic Arabic Corpus as a supplementary teaching and learning material for Arabic syntax: An overview of a web-based Arabic linguistics corpus," in *Int. Seminar on Language, Education, and Culture (ISOLEC)*, 2021, pp. 403–412.
- [99] H. Maraoui, K. Haddar, and L. Romary, "Arabic factoid Question-Answering system for Islamic sciences using normalized corpora," *Procedia Computer Science*, vol. 189, pp. 164–172, 2021.
- [100] D. I. A. Putra and M. Yusuf, "Proposing machine learning of Tafsir al-Quran: In search of objectivity with semantic analysis and Natural Language Processing," in *IOP Conf. Series: Materials Science and Engineering*, vol. 1098, no. 2, 2021, Art. no. 022101.
- [101] O. Ahmed, "Semantic search engine for the holy Quran," Master's thesis, *Metropolia Univ. of Applied Sciences*, 2024.
- [102] M. A. Alqarni, "Embedding search for Quranic texts based on large language models," *The International Arab Journal of Information Technology*, vol. 21, no. 2, pp. 243–257, 2024.
- [103] A. S. M. Khidhir, "An AI model for parsing the text of Holy Quran sentences," *Mesopotamian Journal of Quran Studies*, pp. 20–25, 2024.
- [104] F. Qamar, S. Latif, and R. Latif, "A benchmark dataset with larger context for non-factoid question answering over Islamic text," *arXiv preprint arXiv:2409.09844*, 2024.
- [105] R. Salah, M. Mukred, and L. Q. Zakaria, "A machine learning approach for named entity recognition in Classical Arabic natural language processing," *KSII Transactions on Internet and Information Systems*, vol. 18, no. 11, pp. 3156–3178, 2024.
- [106] A. Badawy, A. Salah, and M. Mahdy, "Digital Quran computing algorithms and applications," *International Journal of Computers and Informatics*, vol. 7, pp. 79–90, 2025.

- [107] A. Bengueddach, "Multi-dimensional NLP analysis of the Quran: Structure, semantics, and emotion across Qira^ˆa^ˆt," *QURANICA: Inter- national Journal of Quranic Research*, vol. 17, no. 1, pp. 32–52, 2025.
- [108] E. Elrefai, T. Khaled, and A. Soliman, "ThinkDrill at IslamicEval 2025 shared task: LLM hybrid approach for Qur'an and Hadith question answering," in *Proc. IslamicEval @ ArabicNLP 2025*, 2025.
- [109] S. Lagrini, A. Debbah, and M. Redjimi, "A new semantic search approach for the Holy Quran based on discourse analysis and advanced word representation models," *International Journal of Computing and Digital Systems*, 2025.
- [110] I. Oshallah, M. Basem, A. Hamdi, and M. Khalil, "Cross-language approach for Quranic QA," in *Int. Congress on Information and Communication Technology*, 2025, pp. 423–431.
- [111] S. Saeed, S. Haider, and Q. Rajput, "On finding similar verses from the Holy Quran using word embeddings," in *2020 Int. Conf. on Information Networking (ICOIN)*, 2020, pp. 524–528.
- [112] S. Amin, R. Aly, Y. Allam, Y. Eid, and E. H. Mohamed, "NUR at IslamicEval 2025 shared task: Retrieval-augmented LLMs for Qur'an and Hadith QA," in *Proc. IslamicEval @ ArabicNLP 2025*, 2025.
- [113] B. Hamoud and E. Atwell, "Evaluation corpus for restricted-domain question-answering systems for the Holy Quran," *International Journal of Science and Research*, vol. 6, no. 8, pp. 1924–1930, 2017.
- [114] M. M. A. Alqahtani, "Quranic Arabic semantic search model based on ontology of concepts," Ph.D. dissertation, Univ. of Leeds, 2019.
- [115] W. A. Nashir et al., "A complete, multi-layered quranic treebank dataset with hybrid syntactic annotations for classical arabic processing," *Data in Brief*, vol. 62, Art. no. 111940, 2025.
- [116] A. O. Adeleke, N. A. Samsudin, A. Mustapha, and M. T. Mustapha, "Comparative analysis of text classification algorithms for automated labelling of Quranic verses," *International Journal of Advanced Science, Engineering and Information Technology*, vol. 7, no. 4, pp. 1419–1427, 2017.
- [117] W. A. Nashir et al., "Noor: A pipeline framework for Classical Arabic parsing," *IEEE Access*, vol. 13, pp. 167538–167557, 2025.
- [118] B. B. Al-Onazi, W. A. Nashir, and A. A. Al-Shargabi, ""Diwan": Constructing the Largest Annotated Corpus for Arabic Poetry," *IEEE Access*, vol. 13, pp. 58927–58941, 2025.

- [119] A. Basharat, A. B. Kamran, and M. Rehman, "SemanticTafsir: Building a cultural heritage ontology and knowledge graph from the Quranic exegesis of al-Tabari," *Semantic Web Journal*, vol. 15, no. 6, pp. 1821–1854, 2024.
- [120] R. G. Al-Anazi, W. A. Nashir, B. B. Al-Onazi, A. A. Alhamad, and A. A. Al-Shargabi, "Maqasid: A Hybrid CNN-BiLSTM Framework for Nuanced Thematic Classification of Arabic Poetry," *IEEE Access*, vol. 13, pp. 189421–189443, 2025.

Appendix

This appendix provides a comprehensive, chronological catalogue of the 112 primary studies systematically reviewed in this survey. The table below serves as a detailed reference, summarizing each publication from 2016 to 2025 across five key dimensions: publication year, principal authors, primary research task, core computational technology employed, and the primary dataset utilized.

Year	Principal Author(s)	Primary Research Task	Core Technology Used	Primary Dataset(s) Used
2016	Ahmed & Atwell [16]	Ontology Development	Ontology Merging, Semantic Rules	Quranic Concepts
	Alqahtani & Atwell [6]	Ontology Development	Ontology/ Semantic Rules	Arabic Quran Text
	Alsukhni et al. [81]	Machine Translation Evaluation	ATEC Metric, Statistical Analysis	Commercial MT Outputs
	Asda et al. [67]	Recitation Analysis (Reciter ID)	Classical ML (ANN), MFCC	Custom Audio (5 reciters)
	Hamdelsayed & Atwell [82]	Question Answering (QA)	Rule-based System	Classical Arabic Quran Text
	Hamed & Ab Aziz [32]	Question Answering (QA)	Classical ML (NN), WordNet	English Translation (Yusuf Ali)
	Noordin et al. [17]	Ontology Development	Ontology (OWL), Graph Database (Neo4j)	English Translation
	Pitchay & Ridzuan [26]	Survey	Systematic Review	N/A (Literature)
	Putra et al. [84]	Question Answering (QA)	Ontology, Semantic Rules	Indonesian Translation
	Sadi et al. [18]	Ontology Development	Ontology (OWL), Domain Modeling	Quranic "Nature" Concepts
	Saloot et al. [83]	Text Analysis (Hadith)	Classical ML (NN, VSM)	Hadith Corpus (Bukhari)
Ta'a et al. [15]	Ontology Development	Ontology-based Retrieval	Arabic Quran Text	

Year	Principal Author(s)	Primary Research Task	Core Technology Used	Primary Dataset(s) Used
2017	Adeleke et al. [116]	Verse Classification	Classical ML (SVM), N-grams	Selected Quran Chapters (English)
	Al-Bakeri & Basuhail [68]	Recitation Analysis (Tajweed)	Classical ML (HMM), CMU Sphinx	Custom Audio (10 reciters)
	Alosaimy & Atwell [85]	Resource (Hadith)	Corpus Linguistics, Annotation	Sunnah Arabic Corpus
	Bougrine et al. [86]	Resource (Speech)	Crowdsourcing, Annotation	Arabic Speech Corpus (Kalam'DZ)
	Hamoud & Atwell [113]	Resource (QA)	Corpus Linguistics	QAEQ&AC Dataset
	Mahmood et al. [87]	Information Retrieval (Hadith)	Data Mining, Keyword Search	Hadith Datasets
	Salloum et al. [27]	Survey	Systematic Review	N/A (Literature)
	Suleiman et al. [28]	Survey	Systematic Review (HMMs)	N/A (Literature)
Ta'a et al. [19]	Ontology Development	Ontology (Prote'ge'-OWL), Semantic Search	Thematic Quranic Verses	
2018	Ahmad et al. [69]	Recitation Analysis (Tajweed)	Classical ML (ANN), MFCC	Custom Audio Recordings
	Al-Ayyoub et al. [70]	Recitation Analysis (Tajweed)	Deep Learning (CDBN), SVM	Custom Audio (3,071 recordings)
	Alhawarat & Hegazi [88]	Text Analysis (Clustering)	Classical ML (K-means), LDA	Arabic Document Collections
	Alqahtani & Atwell [20]	Ontology Development	Ontology (OWL), Data Integration	Bilingual Quranic Datasets
	Basir et al. [23]	Information Retrieval	Ontology-based Search	USIM Search Engine Data
	Rad et al. [24]	Information Retrieval	Lexical Chains, WordNet	English Translation (Yusuf Ali)
	Rusli et al. [7]	Survey	Systematic Review (Ontologies)	N/A (Literature)

Year	Principal Author(s)	Primary Research Task	Core Technology Used	Primary Dataset(s) Used
2019	Adeleke et al. [89]	Verse Classification	Classical ML (SVM, NB)	Full Quran with Manual Labels
	Alqahtani [114]	Question Answering (QA)	Ontology, Transformers (FastText)	AQQAC Dataset
	Atwell [90]	Survey/Resource	Corpus Linguistics Overview	Quranic Arabic Corpus (QAC) & others
	Azmi et al. [29]	Survey (Hadith)	Systematic Review	N/A (Literature)
	Olimat [91]	Text Analysis (Linguistics)	Corpus Linguistics	Quranic Arabic Corpus (QAC)
	Olsson [92]	Text Analysis (Translations)	Digital Humanities, Network Analysis	51 English Translations
	Varghese & Punithavalli [93]	Text Analysis (Comparative)	Classical ML, Topic Modeling	English Translations (Quran, Bible, Gita)

Year	Principal Author(s)	Primary Research Task	Core Technology Used	Primary Dataset(s) Used
2020	Alagrami & Eljazzar [71]	Recitation Analysis (Tajweed)	Classical ML (SVM)	Custom Audio
	Alkhateeb [73]	Recitation Analysis (Reciter ID)	Classical ML (ANN, KNN)	Custom Audio (11,000 samples)
	Alshahrani [94]	Text Analysis (Semantics)	Corpus Linguistics, NLP (Python)	Full Quran + 5 English Translations
	Antoun et al. [33]	NLP Foundations (Language Model)	Transformers (BERT)	AraBERT Corpus
	Nahar et al. [72]	Recitation Analysis (Qira'at)	Classical ML (SVM)	Custom Audio
	Saeed et al. [111]	Information Retrieval	Static Word Embeddings (Word2Vec)	Full Quran Text
	Tlemsani et al. [95]	Text Analysis (Islamic Finance)	Text Mining, Co-occurrence Analysis	Quran and Hadith Texts
	Touati-Hamad et al. [96]	NLP Foundations (Representation)	Static Word Embeddings, TF-IDF	Full Quran Text
Utomo et al. [97]	Survey (QA Systems)	Systematic Review	N/A (Literature)	
2021	Arifianto [98]	Pedagogical Application	Corpus Linguistics	Quranic Arabic Corpus (QAC)
	Bsoul et al. [31]	Survey (Clustering)	Systematic Review	N/A (Literature)
	Darwish et al. [30]	Survey (Arabic NLP)	Systematic Review	N/A (Literature, mentions QAC)
	Maraoui et al. [99]	Question Answering (QA)	Rule-based, TEI Standard	Normalized Islamic Corpora (Tafsir, etc.)
	Putra & Yusuf [100]	Conceptual Framework (Tafsir)	Machine Learning Concepts	N/A (Conceptual)
	Ta'a et al. [22]	Ontology Development	Semantic Modeling	Thematic Quranic Verses

Year	Principal Author(s)	Primary Research Task	Core Technology Used	Primary Dataset(s) Used
2022	Aftab & Malik [36]	Question Answering (QA)	Deep Learning	QRCD Dataset
	Ahmed et al. [37]	Question Answering (QA)	Transformers (BERT)	QRCD Dataset
	Alsaleh et al. [38]	Question Answering (QA)	Transformers (BERT)	QRCD Dataset
	Alshammeri [48]	Semantic Analysis	Deep Learning (Doc2Vec, Transform- ers)	Full Quran Text
	Daud et al. [21]	Ontology Development	Ontology, Semantic Search	Quranic Concepts
	ElKomy & Sarhan [34]	Question Answering (QA)	Transformers (BERT-variants Ensemble)	QRCD Dataset
	Keleg & Magdy [39]	QA (Data Methodology)	Data Splitting Analysis	QRCD Dataset
	Malhas et al. [8]	Community Initiative	Shared Task Organization	QRCD Dataset
	Mellah et al. [40]	Question Answering (QA)	Transformers (Text-to-Text)	QRCD Dataset
	Mostafa & Mohamed [41]	Question Answering (QA)	Deep Learning	QRCD Dataset
	Premasiri et al. [35]	Question Answering (QA)	Transformers (BERT), Transfer Learning	QRCD, SOQAL Datasets
	Singh [42]	Question Answering (QA)	Transformers (BERT)	QRCD Dataset
	Sleem et al. [43]	Question Answering (QA)	Transformers (BERT)	QRCD + Custom Dataset
	Touati-Hamad et al. [47]	Verse Authentication	Deep Learning (CNN/LSTM)	Quran Text + Arabic Learner Corpus

Year	Principal Author(s)	Primary Research Task	Core Technology Used	Primary Dataset(s) Used
2023	Al-Fadhli et al. [77]	Survey (Recitation)	Systematic Review	N/A (Literature)
	Alnefaie et al. [44]	Question Answering (QA)	Transformers (BERT-variants)	Qur'an QA 2023 Dataset
	Ateeq et al. [45]	Question Answering (QA)	Transformers (BERT-variants)	Arabic Narrative QA Datasets
	Bashir et al. [4]	Survey (Quranic NLP)	Systematic Review	N/A (Literature)
	Harere & Jallad [74]	Recitation Analysis (ASR)	Deep Learning (CNN/GRU)	Ar-DAD Dataset
	Harere & Jallad [75]	Recitation Analysis (Tajweed)	Deep Learning (LSTM)	QDAT Dataset
	Nigar et al. [76]	Recitation Analysis (Tajweed)	Deep Learning (DNNs)	Custom & Public Audio
	Rizqullah et al. [46]	Question Answering (QA)	Transformers (BERT-variants)	QASiNa Dataset (Indonesian Sirah)

Year	Principal Author(s)	Primary Research Task	Core Technology Used	Primary Dataset(s) Used
2024	Aftar et al. [59]	Topic Modeling (Hadith)	Transformers (RoBERTa), Doc2Vec	Hadith Corpus
	Ahmed [101]	Information Retrieval	Transformers (BERT), Topic Modeling (LDA)	Full Quran Text (Arabic)
	Alfadhli et al. [50]	Recitation Analysis (ASR)	Deep Learning (BiLSTM), CTC/Attention	Custom Audio
	Alnefaie [9]	Question Answering (QA)	Transformers (BERT), LLMs & RAG	QUQA, HAQA Datasets
	Alqarni [102]	Information Retrieval	LLM Embeddings (OpenAI)	Full Quran Text
	Basharat et al. [119]	Knowledge Representation	Knowledge Graph, Ontology (OWL)	Tafsir al-Tabari Corpus
	Kamran et al. [58]	Knowledge Representation (Hadith)	Knowledge Graph, Transformers (BERT-NER)	Hadith Corpus
	Khidhir [103]	Parsing & Morphology	Classical ML (NN)	Custom Annotated Quranic Data
	Mustafa et al. [11]	Explainable AI (XAI)	Transformers (BERT), Attention Viz.	Tanzil Quran Dataset
	Osman et al. [79]	Recitation Analysis (Tajweed)	Deep Learning (CNN)	Custom Audio
	Qamar et al. [104]	Resource (QA)	Corpus Linguistics	Islamic Text QA Benchmark
	Saadaoui et al. [49]	Question Answering (QA)	LLMs, Prompt Engineering	Quranic Text
	Salah et al. [105]	Text Analysis (NER)	Classical ML	Classical Arabic Texts
	Shakeel et al. [78]	Survey (Recitation)	Systematic Review	N/A (Literature)
	Zafar et al. [51]	Topic Modeling (Multilingual)	Transformers (BERTopic)	Quran-UTM (Urdu Translations)

Year	Principal Author(s)	Primary Research Task	Core Technology Used	Primary Dataset(s) Used
2025	Abbasi et al. [52]	Machine Translation (Urdu)	Deep Learning, WSD	Urdu Translations
	Akra et al. [10]	Resource	Corpus Linguistics, Annotation	QuranMorph Corpus
	Al Adel et al. [62]	QA (Hallucination Detection)	LLMs & RAG	IslamicEval 2025 Dataset
	Alam et al. [66]	Verse Authentication	Transformers (BERT-variants)	Custom Authentic/Non-authentic Data
	Amin et al. [112]	Question Answering (QA)	LLMs & RAG	IslamicEval 2025 Dataset
	An et al. [64]	Survey	Bibliometric Analysis	N/A (Literature)
	Asgari-Bidhendi et al. [60]	Resource (Hadith)	LLMs, Corpus Linguistics	Rezwan Hadith Corpus
	Badawy et al. [106]	Topic Modeling	Classical ML (LDA, K-means), UMAP	Selected Surahs (Arabic)
	Bengueddach [107]	Text Analysis (Multi-dimensional)	Transformers (AraBERT), Sentiment Analysis	Full Quran Text
	Danish Khaleeq et al. [53]	Machine Translation (Urdu)	Transformers (NMT), Multilingual BERT	Arabic-Urdu Parallel Corpus
	Elrefai et al. [108]	Question Answering (QA)	LLMs & RAG	IslamicEval 2025 Dataset
	Gaanoun & Alsuhaibani [54]	Text Analysis (Sentiment)	Sentiment Analysis	English Translations
	Jurczyk et al. [65]	Resource	Corpus Linguistics, Text Mining	Tafsir Corpus (80 commentaries)
	Lagrini et al. [109]	Information Retrieval	Transformers (BERT), Discourse Analysis	Full Quran Text

Year	Principal Author(s)	Primary Research Task	Core Technology Used	Primary Dataset(s) Used
2025	Nashir et al. [117]	Parsing & Morphology	Deep Learning (BiLSTM)	Quranic Treebank (Reconstructed)
	Nashir et al. [115]	Resource	Data Curation, DL Parsing	Extended Quranic Treebank (EQTB)
	Omayrah et al. [61]	QA (Hallucination Detection)	LLMs & RAG	IslamicEval 2025 Dataset
	Oshallah et al. [110]	Question Answering (QA)	Transformers, Cross-lingual	Quranic Text
	Salah et al. [63]	Topic Modeling	Clustering, TF-IDF	Surah Al-Baqarah
	Shaiakhmetov et al. [80]	Recitation Analysis (Tajweed)	Deep Learning (DNNs)	Custom Audio
	Shahid et al. [57]	Topic Modeling	LLM Embeddings (OpenAI), Clustering	English Translation
	Shahzad & Ashraf [5]	Survey	Systematic Mapping Study	N/A (Literature)
	Tariq et al. [55]	Question Answering (QA) (Urdu)	Transformers (BERT, RoBERTa)	Urdu Translations
	Utomo et al. [25]	Ontology Development	Static Word Embeddings, Imbalanced Learning	Indonesian Quran Ontology