Indonesian Journal of Electrical Engineering and Computer Science
Vol. 39, No. 2, August 2025, pp. 987~995
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v39.i2.pp987-995

The development of contextual chat interactions with retrieval-augmented generation system for facilitating learning hadith

Rio Nurtantyana[1,2,3], Yudi Priyadi[1,3], Eko Darwiyanto[3,4]
[1]Center of Excellence of Artificial Intelligence for Learning and Optimization, Telkom University, Bandung, Indonesia
[2]Research Center for Data and Information Sciences, National Research and Innovation Agency (BRIN), Bandung, Indonesia
[3]School of Computing, Telkom University, Bandung, Indonesia
[4]CoE HUMIC, School of Computing, Telkom University, Bandung, Indonesia

Article Info

Article history:

Received Jan 5, 2025
Revised Mar 18, 2025
Accepted Jul 1, 2025

Keywords:

Contextual answer
Interactive learning
Large language model
Learning of hadith
RAG system

ABSTRACT

This study explores the development and implementation of a retrieval-augmented generation (RAG) system using the large language model (LLM) to enhance the learning of hadith through a chat interface for high school students. This study addresses challenges in optimizing RAG configurations and problems associated with traditional educational methods that lack interactivity. In addition, the RAG system was designed to replace real teacher interactions, offering a chat feature that provides contextual answers to real-life scenarios related to Hadith. Various configurations were tested, with a focus on the Matn component, achieving a high accuracy score with a mean of .754 and demonstrating efficiency in context relevance with a mean of .797. Results indicated significant accessibility using our RAG system for learning hadith via WhatsApp's chat interface. Hence, this study highlights the potential of RAG systems in transforming educational environments and offers insights into the development of technology for interactive Hadith learning solutions.

Corresponding Author:

Rio Nurtantyana
Research Center for Data and Information Sciences, National Research and Innovation Agency (BRIN)
Bandung, Indonesia
Email: akunerio@gmail.com

1. INTRODUCTION

In the digital age, traditional methods of searching and retrieving information are increasingly being supplemented or replaced by more sophisticated technologies [1]. Key among these advancements is the use of large language models (LLMs) to powering AI systems like retrieval-augmented generation (RAG), which promise to revolutionize the way users search for and interact with information [1]-[3]. While traditional search systems often return general results, the RAG systems are designed to deliver precise and contextually relevant responses [1], [4]. However, platforms like ChatGPT can also sometimes provide responses that are not sufficiently grounded in related references [5], [6], which is particularly problematic when dealing with sensitive and precise texts, such as Hadith data.

Hadith literature in Islam is a fundamental component of religious tradition, comprising the sayings, actions, and approvals of the Prophet Muhammad [7]. These texts serve as primary sources for Islamic law and moral guidance, second only to the Quran, and their interpretation requires meticulous accuracy and context [8], [9]. One of the most renowned collections of hadith in Islamic literature is Hadith Bukhari [10]. However, the growing complexity of hadith literature with its substantial volume and the need for precise interpretation can present a significant challenge, especially when the hadith have been translated into different languages. Based on the global Muslim population, Indonesia is the country with the largest Muslim

population in the world [11], [12]. In detail, 87.2% of the population in Indonesia, or 207 million Indonesians is identified as Muslim in 2023 [11]. It means that Islamic literature like Hadith Bukhari should be translated into the Indonesian language to reach Muslim people in Indonesia.

Currently, traditional digital search methods and even AI models like ChatGPT can fail to retrieve accurate or context-relevant Hadith content due to hallucinations, which leads to misunderstandings or misinterpretations [13]. To recognize these challenges, recent technological advancements, such as RAG systems could offer promising solutions [14]. By using the RAG system, it could enable the creation of intelligent searching or chatbots that can perform rapid and precise searches within large datasets, effectively overcoming the limitations of generic AI systems by anchoring their responses in specific terminology [1].

Hence, this study investigates the application of the RAG system for the precise and efficient retrieval of Hadith Bukhari, which is translated into the Indonesian language [15]. Specifically, it addresses two main challenges such as determining whether there are significant differences in performance when employing various RAG configurations and evaluating the benefits of using RAG for developing a chat interaction dedicated to learning hadith. Through this study, we aim to enhance the precision and relevance of AI-driven educational tools using RAG in delivering translated Hadith Bukhari content.

2. LITERATURE REVIEW

2.1. The potential of LLMs in learning hadith

Hadith literature as a foundational component of Islam, requires meticulous authenticity and precision in its retrieval and interpretation [8], [10]. Traditional search systems, which often rely solely on keyword matching are inadequate for handling the complexity and nuance inherent in hadith content [9]. However, AI like ChatGPT can result in inaccuracies due to hallucinations and lack of the contextual understanding necessary for meaningful engagement compared to the original contexts [3], [16]. The vastness of the hadith corpus and its critical role in Islamic jurisprudence necessitate more sophisticated retrieval mechanisms that ensure accuracy and contextual integrity [1], [10], [13]. Recent advancements in LLMs present new opportunities for improving how hadith can be accessed and understood. AI models like OpenAI's GPT series have shown the potential to generate human-like responses by understanding the context at a deeper level [17]. However, when dealing with religious content like in hadith literature, it is crucial that these AI models are not only accurate but also closely aligned with original sources to reduce the hallucinations [4].

2.2. RAG systems as solutions for learning hadith

RAG systems empowered with LLMs offer promising enhancements. These technologies leverage the semantic capabilities of AI by embedding texts in a manner that facilitates effective context-based searches and retrievals [3], [18]. Furthermore, RAG systems integrate structured knowledge databases with a generation process with LLM, thus ensuring that the information retrieved is not only contextually relevant but also precise and factually accurate [19]. Hence, such systems can effectively point out responses to validated hadith references [1]. Implementing RAG systems specifically tailored for hadith retrieval can significantly improve educational experiences, particularly for learners in digital environments. By enabling precise and rapid searches within authorized hadith datasets, these systems can provide users with contextually rich and factually accurate answers to their queries [1], [5]. This method enhances the engagement and learning process, mimicking the benefits of direct interaction with a knowledgeable teacher [20].

In terms of learning Hadith, the RAG system needs to integrate with interactive interactions for example using the chatting interface [2], [21]. One of the common chatting interface systems is Whatsapp (WA) that mostly used for daily online chatting and the previous study also mentioned that the WA is the effective communication to support learning [22], [23]. In addition, the WA platform offers an application programming interface (API) that allows developers to integrate with other systems [24]. A previous study also mentioned that using the interface for chatting could reduce their cognitive load [25] and they could focus on content only in their learning [26]. Hence, the RAG system as a backend could integrate with WA as the interface to provide related contextual hadith using a chatting mechanism. Therefore, this mechanism not only supports a dynamic learning environment but also makes studying hadith more accessible.

2.3. The RAG evaluation

The evaluation of RAG systems is crucial, particularly when applied to sensitive domains like hadith literature [8]. There were several evaluation metrics for the RAG system, for example TruLens [27]. The TruLens framework emphasizes three key metrics such as context relevance, groundedness, and answer relevance, as shown in Figure 1 [27]. Context relevance assesses how well the generated responses align with the situational context of the user query. Groundedness measures the extent to which the responses are

anchored in recognized and authoritative hadith sources, preventing misinformation. Lastly, answer relevance evaluates the precise correlation between the query and the information provided, ensuring that responses are not only relevant but also directly concise to the queries posed [4].

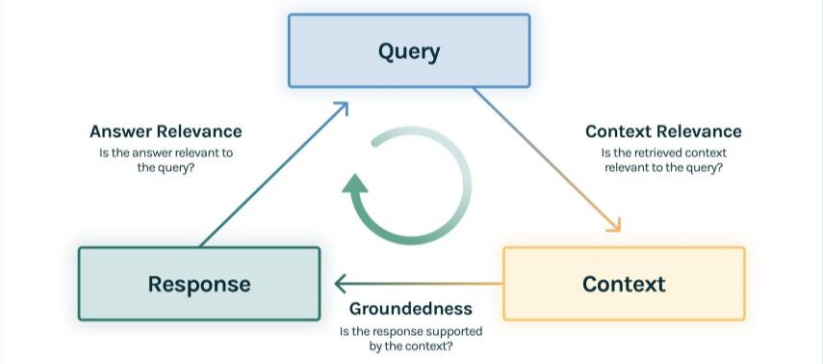


Figure 1. The RAG triad evaluation [27]

2.4. GAP analysis

Since the Hadith Bukhari data must be authentic, necessitating a search mechanism rather than generation by LLMs [3], [4], [9], [16]. Previous studies indicate that Islamic search systems often rely on traditional query methods like input one or two keywords, which limited in capturing semantic meanings [9], [28]. While information retrieval algorithms like latent semantics, cosine similarity [29], and TF-IDF [30] have been applied, precision remains low at 36.25% for searching the hadith data [29]. Conversely, embedding search mechanisms with LLMs show promise in semantic retrieval from the Quranic texts [13], [31]. Hence, this study contributes by employing AI with a RAG system and embedding search mechanisms [32] for authentic hadith Indonesian translation retrieval [1]. Unlike prior studies, we propose RAG system with chat-based interaction using LLMs for a user-friendly search experience via a WA interface for daily use and learning hadith as well [22]-[25]. Furthermore, we also evaluated semantic data in the RAG system performance across different hadith data simulations in the RAG system.

3. METHOD

The methodology for this study employed a prototyping approach within the software development life cycle (SDLC) model to iteratively develop and refine an RAG system for learning hadith [33], as shown in Figure 2. In the first, the communication with a survey method was conducted among twenty-one high school students from an Islamic private school in Bandung, Indonesia. These students were randomly selected to provide input on the prototype requirements through a structured survey form.

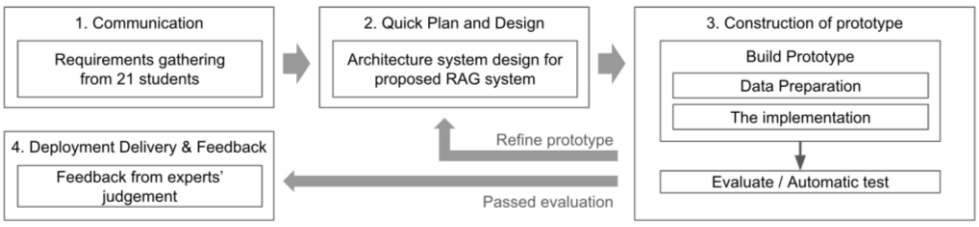


Figure 2. The implementation of prototyping methodology in this study

The second and the third steps were iteration steps to refine a prototype. We built the architecture system that included the RAG and then it used three different configurations involving Matn and Sanad, Matn only, and Matn and Chapter of the Hadith dataset for the construction of the prototypes. The hadith dataset of Hadith Bukhari was obtained from open-source data in Github [15]. For evaluation, two testing methods were employed such as automatic and expert evaluation. The automatic evaluation used the RAG Triad framework based on the TruLens technical report, which assessed system-generated answers for

precision and context relevance from 10 questions. The TruLens framework emphasizes three key metrics such as context relevance, groundedness, and answer relevance, as shown in Figure 1 [27].

After we iterated three times and passed the automatic evaluation, the system deployed and got the feedback from teachers as experts evaluation. Three experts asked 20 logical questions, like daily life-related questions in the system to generate answers including relevant hadith references. Three raters then assessed the generated responses using a scoring rubric adapted from the Langchain technical report for evaluating LLMs [34]. Scores ranged from one to ten, with one indicating inaccurate and irrelevant answers and ten signifying precise answers accurately related to the hadith references. This dual evaluation approach ensured both the technical performance and practical applicability of the RAG system were thoroughly assessed, leading to a prototype that met students' educational needs.

4. RESULTS AND DISCUSSION

In this section, the results and discussion are organized through the prototyping methodologies in four steps, such as communication, quick plan & design, prototype construction, and deployment and feedback.

3.1. Communication

The results of a survey which was obtained from twenty-one high school students as the users showed that most of the female students (*n*=13) participated in this study compared to the male students (n=8). Based on their background, the 13 students learn hadith from the hard-copy book, a student learns hadith from an Android application, and 7 students didn't learn hadith before both from the book and the mobile application. Interestingly, most of them usually used smartphones for daily life tasks and learning (*n*=20). Moreover, all of them want to access and learn hadith from their smartphone since it is easier compared to the laptop computer and they use it daily (*n*=20). Furthermore, they didn't want to use a mobile application with a search feature or search hadith in a search engine since the experience is different compared to learning with a teacher directly. However, they need to have a human teacher with a chat feature that is able to answer the question of real-life scenarios, which are related to the hadith (*n*=15).

Hence, it needs to develop a mobile application tailored for high school students to learn hadith, leveraging their preference for smartphones due to convenience and daily usage. This application should offer an interactive learning experience that mimics the engagement found in traditional teacher-student interactions, incorporating a chat feature for live assistance from knowledgeable teachers to address real-life scenarios related to hadith. The design must prioritize a user-friendly and intuitive interface, facilitating easy access, and it can accommodate diverse learning preferences.

3.2. Quick plan and design

Based on the communication phase, there was one main requirement in the mobile application such as the chat feature for live assistance from knowledge teachers that could answer questions related to hadith. Hence, we used the RAG mechanisms that could replace the mechanism of live assistance from knowledge based on hadith data from the teacher chatting due to limited time. Furthermore, the WA could be used as a user interface for live chatting due to its robust features and ease of use for high school students [35].

Therefore, we designed the architecture system to provide the answer based on the students' inquiry regarding the hadith as shown in Figure 3. The architecture system, managed under the LlamaIndex framework begins with data preparation and ingestion, where 7008 items of Hadith Bukhari in Bahasa Indonesian language are processed.

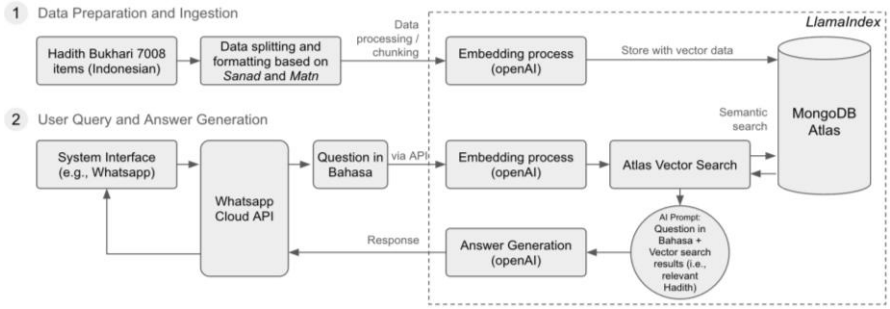


Figure 3. The RAG system for contextual chat interactions based on hadith data

In the first phase, it involves splitting and formatting the data based on the Sanad and Matn, which refers to the transmission chain and the text of the Hadith, respectively. After formatting, the data undergoes an embedding process using OpenAI to convert it into vector data. This vector data is then stored in MongoDB Atlas for semantic searching. In the second phase, user queries and answer generation are facilitated through a system interface, such as WhatsApp (WA). When a user submits a question in Indonesian language, the WA Cloud API forwards it for processing. The query undergoes an embedding process to transform it into a format compatible with the stored vector data. An atlas vector search is conducted to find relevant Hadiths by performing a semantic search through MongoDB Atlas. The combined data, consisting of the user's question and the search results, is then used in an answer generation process, leveraging OpenAI models to create a comprehensive response. The final answer is sent back to the user through the interface, enabling interactive and informative engagement.

3.3. Prototype construction

3.3.1. Data preparation

In the prototype construction, the data preparation phase is crucial for structuring and organizing the Hadith content for semantic analysis and retrieval. The translated Hadith Bukhari dataset, in this case, is prepared using a manual process that involves splitting each entry into several components that allow for both individual and combined analysis as shown in Table 1, such as a) the Matn (text of the Hadith), b) Sanad (chain of narrators), c) original Arabic text, d) ID number, and e) chapter title. Each Hadith item is also categorized by its ID number and chapter title, providing a comprehensive framework that facilitates the indexing and retrieval tasks during semantic searches. By maintaining the integrity and context of each component, this preparation phase ensures that the prototype system can effectively deliver accurate and contextually relevant responses to user queries. This structured approach is foundational for conducting experiments and refining the model's ability to interpret and engage with the Hadith data in this study.

Table 1. The example of Hadith Bukhari number 1 in translated Indonesian language

No	Component	Example (Indonesian language)
1	Hadith Translation - Matn	*Semua perbuatan tergantung niatnya, dan (balasan) bagi tiap-tiap orang (tergantung) apa yang diniatkan; Barangsiapa niat hijrahnya karena dunia yang ingin dicapainya atau karena seorang perempuan yang ingin dinikahinya, maka hijrahnya adalah kepada apa dia diniatkan.*
2	Hadith Translation - Sanad	*Telah menceritakan kepada kami [Al Humaidi Abdullah bin Az Zubair] dia berkata, Telah menceritakan kepada kami [Sufyan] yang berkata, bahwa Telah menceritakan kepada kami [Yahya bin Sa'id Al Anshari] berkata, telah mengabarkan kepada kami [Muhammad bin Ibrahim At Taimi], bahwa dia pernah mendengar [Alqamah bin Waqash Al Laitsi] berkata; saya pernah mendengar [Umar bin Al Khaththab] diatas mimbar berkata; saya mendengar Rasulullah shallallahu 'alaihi wasallam bersabda.*
3	Hadith - Arabic	حَدَّثَنَا الْحُمَيْدِيُّ عَبْدُ اللَّهِ بْنُ الزُّبَيْرِ ، قَالَ: حَدَّثَنَا سُفْيَانُ ، قَالَ: حَدَّثَنَا يَحْيَى بْنُ سَعِيدٍ الْأَنْصَارِيُّ ، قَالَ: أَخْبَرَنِي مُحَمَّدُ بْنُ إِبْرَاهِيمَ التَّيْمِيُّ ، أَنَّهُ سَمِعَ عَلْقَمَةَ بْنَ وَقَّاصٍ اللَّيْثِيَّ ، يَقُولُ: سَمِعْتُ عُمَرَ بْنَ الْخَطَّابِ رَضِيَ اللَّهُ عَنْهُ عَلَى الْمِنْبَرِ، قَالَ: سَمِعْتُ رَسُولَ اللَّهِ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ، يَقُولُ: " إِنَّمَا الْأَعْمَالُ بِالنِّيَّاتِ، وَإِنَّمَا لِكُلِّ امْرِئٍ مَا نَوَى، فَمَنْ كَانَتْ هِجْرَتُهُ إِلَى دُنْيَا يُصِيبُهَا أَوْ إِلَى امْرَأَةٍ يَنْكِحُهَا، فَهِجْرَتُهُ إِلَى مَا هَاجَرَ إِلَيْهِ
4	ID	1
5	Book	1
6	Chapter	*Permulaan wahyu*

3.3.2. The reconstruction of the prototype

Regarding the reconstruction of the prototype, the use of LLMs through the RAG system provides significant advantages in handling Hadith data, as shown in Table 2. By combining Python for RAG with NodeJS for the WhatsApp cloud API, the system leverages advanced technologies like OpenAI's GPT-4 for answer generation and the text-embedding-3-large model for embedding processes. This architecture, supported by the LlamaIndex framework and MongoDB Atlas as the vector database, ensures efficient data retrieval and contextual response generation. The RAG system excels over traditional methods by not only performing semantic searches to retrieve relevant Hadith but also generating human-like answers that align with the student's queries. This capability is essential for providing fact-based, contextually enriched answers that resonate well with users, offering a more interactive and engaging experience that mimics learning from a human expert [2].

Table 2. The use of LLMs for the RAG system in this study

Component	Stack
Language	Python (RAG) and NodeJS (Whatsapp cloud API)
Model	gpt-4 model for answer generation and text-embedding-3-large for embedding process from openAI
Architecture	LlmaIndex
Vector database	MongoDB Atlas

3.3.3. The experiment with different configurations

Furthermore, a series of experiments were conducted to evaluate different mechanisms for embedding Hadith data into a vector database as shown in Table 3. These experiments were designed to assess how splitting the Hadith data affects query performance and retrieval accuracy. The experiments involved three configurations such as Matn and Sanad (E1), Matn only (E2), and Matn and Chapter (E3). Each setup was tested using two types of questions-user-generated (zero-shot) answers not available in the vector database and predefined questions with answers available in the vector database. TruLens was used for a comprehensive evaluation, measuring factors such as groundedness, answer relevance, and context relevance. The results indicated that E3 had the highest answer relevance score with a mean of 0.960, while E2 showed better context relevance with a mean of 0.797. Additionally, metrics such as total tokens, cost, and latency were also monitored, revealing that while E2 minimized latency and cost, E3 optimized relevance. This detailed comparison highlights the trade-offs and efficiencies associated with different data preparation strategies, guiding the optimal design for scalable, responsive, and accurate Hadith retrieval systems. In the TruthLens evaluation, the RAG performs better when searching the preset question (M = .829; M = .824, M = .821) compared to the users' questions (M = .658; M = .683; M = .644). It is because the preset question was constructed based on the hadith data, meanwhile, the user's questions were constructed based on the users' perspectives or cases. This result was in line with the previous research that searching with similarity algorithms will result in quicker and higher precision [36]. Furthermore, we averaged the RAG results of both preset and user questions on each experiment (E1-E3) for further analysis. In the overall TruLens evaluation with three dimensions, the E2 was better than E1 and E3 (M = .754). It is because we save the Manad only with an embedding vector, which is saved in the vector store. The similarity searching will perform better and concisely related to the question. In detail, the groundedness and the context relevance of E2 (M = .529; M = .797) perform better than the E1 and E3.

Table 3. The result of the comparison of the experiment with different configurations.

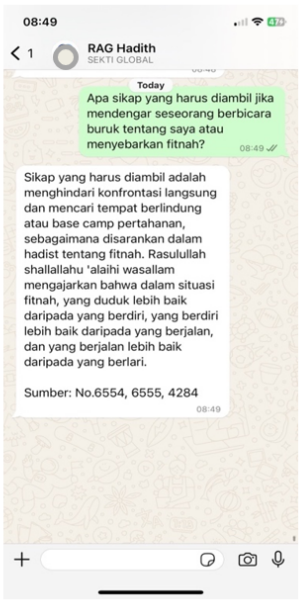
Experiment		Matn and Sanad (E1)			Matn only (E2)			Matn and Chapter (E3)		
Question		Preset	User	Mean	Preset	User	Mean	Preset	User	Mean
Tru Lens	Groundedness	.685	.350	.518	.665	.393	**.529**	.705	.350	.528
	Answer relevance	.970	.930	.950	.930	.940	.935	.970	.950	**.960**
	Context relevance	.833	.693	.763	.877	.717	**.797**	.789	.633	.711
	Mean	.829	.658	.744	.824	.683	**.754**	.821	.644	.733
Total token		1004.7	890.2	947.45	581.9	639.9	610.9	1044.8	873.7	959.25
Total cCost (USD)		.006	.005	.006	.004	.004	.004	.006	.005	.006
Latency (s)		7.9	5.1	6.5	4.5	3.9	4.2	5.7	4.3	5

Note about the question section:
Preset indicates that the questions asked by users, which answers available in the vector database, and the user indicates that the question asked by users, which answers not available in the vector database

Furthermore, E2, which focused on using only the Matn component of the Hadith, demonstrated impressive efficiency in terms of lower token usage, reduced latency, and cost effectiveness. This approach minimizes the data processed during query handling, resulting in faster response times and lower computational costs. By concentrating solely on the Matn, the system can streamline searches and retrievals without sacrificing the core content needed to generate relevant answers. Metrics such as context relevance and groundedness are crucial in the RAG system because they ensure that the indexed data aligns closely with the source references, providing accurate and contextually appropriate responses [27]. Given these advantages, E2 emerges as the best configuration for deployment, offering a balance of speed, cost efficiency, and robust context alignment, making it ideal for deployment in this study [28].

3.4. Deployment and feedback

In the deployment phase of this study, we implemented a dual approach to evaluation by using both quantitative and qualitative feedback mechanisms. Alongside the quantitative evaluation metrics such as TrueLens evaluation, we incorporated human evaluation to gather nuanced insights into our system's effectiveness and user satisfaction. This involved three independent raters who assessed the system via the WA application. The raters achieved a moderate agreement score, as determined by the Inter-Class Correlation (ICC), reflecting a reliable level of consistency in their evaluations. The final results averaged 8.36 from 10 scores, indicating a generally positive reception. To gain more insight into the deployment process, see Figure 4, which illustrates the deployment of the RAG system using the WA application interface that integrates data from our E2.



Data sources:
The data source from the top 3 search results on the vector database of E2 and then RAG helped to generate the answer.

1. ID: 6554; similarity score=0.789
2. ID: 6555; similarity score=0.778
3. ID: 4284; similarity score=0.763

Translated Q&A:
Question:
What attitude should I take if I hear someone talking badly about me or spreading slander?
Answer:
The attitude that must be taken is to avoid direct confrontation and look for a place of shelter or a defense base camp, as suggested in the hadith about fitnah. Prophet Muhammad taught that in a situation of fitnah, the one who sits is better than the one who stands, the one who stands is better than the one who walks, and the one who walks is better than the one who runs.
Source No. 6554, 6555, 4284

Figure 4. The deployment of the RAG system using the WA application interface

5. CONCLUSION

This study demonstrated that using an RAG system could perform semantic searching and develop a live chat for learning translated Hadith Bukhari that offers significant benefits compared to the traditional searching mechanism, particularly in enhancing user engagement and accessibility for educational tools in the future. By focusing on the Matn component only of the translated Hadith Bukhari data, the RAG system provided accurate and efficient responses, highlighting differences in RAG configurations' performance. Hence, the response was relevant to the authentic Hadith data and it will be useful for education tools like learning hadith in school or home. However, the study were limitations, such as a small sample size and potential subjective biases in human evaluation. To improve and validate the findings, future study should involve larger and more diverse groups. In addition, the experiment in future should explore long-term impacts on knowledge retention with quasi-experimental design for learning hadith in the school or home. Regarding the RAG system, it might be possible to use different frameworks for search engines via LLM. Additionally, integrating adaptive learning techniques and collaborating with educators could enhance the system's educational effectiveness.

FUNDING INFORMATION

This research was supported by Directorate of Research & Community Service, Telkom University, Indonesia under Grant KWR4.018/LIT06/PPM-LIT/2024.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Rio Nurtantyana	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			
Yudi Priyadi		✓		✓	✓				✓	✓	✓	✓	✓	✓
Eko Darwiyanto	✓		✓	✓	✓	✓				✓	✓	✓	✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

Derived data supporting the findings of this study are available from the corresponding author Nurtantyana on request.

REFERENCES

[1] X. Shi, J. Liu, Y. Liu, Q. Cheng, and W. Lu, "Know where to go: make LLM a relevant, responsible, and trustworthy searchers," *Decision Support Systems*, vol. 188, 2025, doi: 10.1016/j.dss.2024.114354.

[2] B. Paneru, B. Thapa, and B. Paneru, "Leveraging AI in ayurvedic agriculture: a RAG chatbot for comprehensive medicinal plant insights using hybrid deep learning approaches," *Telematics and Informatics Reports*, vol. 16, 2024, doi: 10.1016/j.teler.2024.100181.

[3] S. Vidivelli, M. Ramachandran, and A. Dharunbalaji, "Efficiency-driven custom chatbot development: unleashing LangChain, RAG, and performance-optimized LLM fusion," *Computers, Materials and Continua*, vol. 80, no. 2, pp. 2423–2442, 2024, doi: 10.32604/cmc.2024.054360.

[4] M. Arslan, H. Ghanem, S. Munawar, and C. Cruz, "A survey on RAG with LLMs," *Procedia Computer Science*, vol. 246, no. C, pp. 3781–3790, 2024, doi: 10.1016/j.procs.2024.09.178.

[5] A. Casheekar, A. Lahiri, K. Rath, K. S. Prabhakar, and K. Srinivasan, "A contemporary review on chatbots, AI-powered virtual conversational agents, ChatGPT: Applications, open challenges and future research directions," *Computer Science Review*, vol. 52, 2024, doi: 10.1016/j.cosrev.2024.100632.

[6] E. Lozić and B. Štular, "Fluent but not factual: a comparative analysis of ChatGPT and other AI Chatbots' proficiency and originality in scientific writing for humanities," *Future Internet*, vol. 15, no. 10, 2023, doi: 10.3390/fi15100336.

[7] A. B. Kamran, B. Abro, and A. Basharat, "SemanticHadith: An ontology-driven knowledge graph for the hadith corpus," *Journal of Web Semantics*, vol. 78, pp. 1–13, 2023, doi: 10.1016/j.websem.2023.100797.

[8] B. Sulistio, A. Ramadhan, E. Abdurachman, M. Zarlis, and A. Trisetyarso, "The utilization of machine learning on studying Hadith in Islam: A systematic literature review," *Education and Information Technologies*, vol. 29, no. 5, pp. 5381–5419, 2024, doi: 10.1007/s10639-023-12008-9.

[9] S. Hakak *et al.*, "Digital Hadith authentication: recent advances, open challenges, and future directions," *Transactions on Emerging Telecommunications Technologies*, vol. 33, no. 6, 2022, doi: 10.1002/ett.3977.

[10] L. El-Ali, "Hadith Corpus," in *No Truth Without Beauty: God, the Qur'an, and Women's Rights*, 2022, pp. 11–17, doi: 10.1007/978-3-030-83582-8_2.

[11] A. Rofii and N. Hosen, "Freedom of religion and religious minorities in indonesia: the local beliefs case," *Freedom of Religion and Religious Diversity*, pp. 271–298.

[12] Prayer Times, "Global Muslim Population," *Times Prayer*, 2024. https://timesprayer.com/en/muslim-population/ (accessed Dec. 30, 2024).

[13] T. M. T. Sembok and S. Wani, "Is ChatGPT not appropriate for religious use?," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 14322 LNCS, pp. 595–605, 2024, doi: 10.1007/978-981-99-7339-2_48.

[14] T. Feridooni *et al.*, "Development of a vascular surgery-specific artificial intelligence chat interface using retrieval-augmented generation: VASC.AI, a specialized vascular surgery chatbot," *JVS-Vascular Insights*, vol. 2, p. 100137, 2024, doi: 10.1016/j.jvsvi.2024.100137.

[15] Github, "Hadith - API," 2020. https://github.com/gadingnst/hadith-api (accessed Oct. 10, 2024).

[16] M. Chelli *et al.*, "Hallucination rates and reference accuracy of ChatGPT and bard for systematic reviews: comparative analysis," *Journal of Medical Internet Research*, vol. 26, 2024, doi: 10.2196/53164.

[17] T. B. Brown *et al.*, "Language models are few-shot learners," *Advances in neural information processing systems*, vol. 33, pp. 1877–1901, 2020.

[18] J. Johnson and M. Douze, "Billion-scale similarity search with GPUs," *ArXiV*, 2017.

[19] P. Lewis *et al.*, "Retrieval-augmented generation for knowledge-intensive nlp tasks," *Advances in neural information processing systems*, vol. 33, pp. 17105–17116, 2020.

[20] R. Swarup, "Understanding the Hadith: the sacred traditions of Islam," *Choice Reviews Online*, vol. 40, no. 09, pp. 40-5184-40–5184, 2003, doi: 10.5860/choice.40-5184.

[21] A. Bora and H. Cuayáhuitl, "Systematic analysis of retrieval-augmented generation-based LLMs for medical chatbot applications," *Machine Learning and Knowledge Extraction*, vol. 6, no. 4, pp. 2355–2374, 2024, doi: 10.3390/make6040116.

[22] H. Mulyono, G. Suryoputro, and S. R. Jamil, "The application of WhatsApp to support online learning during the COVID-19 pandemic in Indonesia," *Heliyon*, vol. 7, no. 8, 2021, doi: 10.1016/j.heliyon.2021.e07853.

[23] M. Conde, F. J. Rodríguez-Sedano, F. J. Rodríguez Lera, A. Gutiérrez-Fernández, and Á. M. Guerrero-Higueras, "WhatsApp or Telegram. Which is the best instant messaging tool for the interaction in teamwork?," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12784 LNCS, pp. 239–249, 2021, doi: 10.1007/978-3-030-77889-7_16.

[24] G. R. Gunnam, D. Inupakutika, R. Mundlamuri, S. Kaghyan, and D. Akopian, "Assessing performance of cloud-based heterogeneous chatbot systems and a case study," *IEEE Access*, vol. 12, pp. 81631–81645, 2024, doi: 10.1109/ACCESS.2024.3397053.

[25] B. Urien, A. Erro-Garcés, and A. Osca, "WhatsApp usefulness as a communication tool in an educational context," *Education and Information Technologies*, vol. 24, no. 4, pp. 2585–2602, 2019, doi: 10.1007/s10639-019-09876-5.

[26] N. Al-Qaysi, N. Mohamad-Nordin, and M. Al-Emran, "Developing an educational framework for using WhatsApp based on social constructivism theory," *Studies in Systems, Decision and Control*, vol. 295, pp. 243–252, 2021, doi: 10.1007/978-3-030-47411-9_14.

[27] TruLens, "The RAG triad." https://www.trulens.org/getting_started/core_concepts/rag_triad/ (accessed Oct. 10, 2024).

[28] S. Alowaidi, E. Atwel, and M. A. Alsalka, "Survey of semantic web," *International Journal on Islamic Applications in Computer Science And Technology*, vol. 12, no. 4, 2024.
[29] W. Darmalaksana, C. Slamet, W. B. Zulfikar, I. F. Fadillah, D. S. adillah Maylawati, and H. Ali, "Latent semantic analysis and cosine similarity for hadith search engine," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 18, no. 1, pp. 217–227, 2020, doi: 10.12928/TELKOMNIKA.V18I1.14874.
[30] K. A. F. A. Samah, N. F. S. Norhisam, S. F. A. Fesol, and R. Aminuddin, "TF-IDF and data visualization for syafie madhhab hadith scriptures authenticity," *ISCAIE 2021 - IEEE 11th Symposium on Computer Applications and Industrial Electronics*, pp. 65–70, 2021, doi: 10.1109/ISCAIE51753.2021.9431831.
[31] M. Alqarni, "Embedding search for Quranic texts based on large language models," *The International Arab Journal of Information Technology*, vol. 21, no. 2, 2024, doi: 10.34028/iajit/21/2/7.
[32] M. A. Abdulrahman, "The future of Hadith studies in the digital age: opportunities and challenges," *Journal of Ecohumanism*, vol. 3, no. 8, pp. 2792–2800, 2024, doi: 10.62754/joe.v3i8.4927.
[33] R. S. Pressman, "Software Engineering : a practitioner's approach." New York: McGraw Hill, 2010.
[34] Langchain, "Introduction of the documentation," 2024. https://python.langchain.com/docs/introduction/ (accessed Oct. 10, 2024).
[35] N. Yuviler-Gavish, R. Halutz, and L. Neta, "How whatsappization of the chatbot affects perceived ease of use, perceived usefulness, and attitude toward using in a drive-sharing task," *Computers in Human Behavior Reports*, vol. 16, 2024, doi: 10.1016/j.chbr.2024.100546.
[36] S. K. Ali, Z. M. Aydam, and B. M. Rashed, "Similarity metrics for classification: a review," *IOP Conference Series: Materials Science and Engineering*, vol. 928, no. 3, 2020, doi: 10.1088/1757-899X/928/3/032052.

BIOGRAPHIES OF AUTHORS

Rio Nurtantyana received the Ph.D. degree. He is currently working as a researcher with the research center for data and information sciences, National Research and Innovation Agency (BRIN), Bandung, Indonesia. He is also a lecturer in software engineering courses with the School of Computing, Telkom University, Bandung. His research interests include the IoT, AI, and mobile learning with AR/VR to transform the education in digital era. He was selected as an Honorary Member of the Phi Tau Phi Scholastic Honor Society of the Republic of China, in 2023. He can be contacted at email: akunerio@gmail.com.

Yudi Priyadi is currently active as a researcher and lecturer at the department of software engineering, Telkom University. He has the competence of teaching and practitioners in requirement engineering, text preprocessing, data management, and software development. He can be contacted at email: whyphi@telkomuniversity.ac.id.

Eko Darwiyanto is currently working as a lecturer in the Department of Software Engineering at Telkom University, Bandung, Indonesia. His expertise in software development methodologies and software engineering. He can be contacted at email: ekodarwiyanto@telkomuniversity.ac.id.