



EMasjid, Islamic Chatbot Application With RAG (Retrieval Augmented Generation)

Kholid Fathoni^a, Aris Sudaryanto^{b}, Iwan Syarif^c, Zakha Maisat Eka D.^d, Andhik Ampuh Yunanto^e, Prasetyo^f, Rendra Suprobo Aji^g, Aryazaky Iman Fauzy^h, Rizky Yuniar Hakkunⁱ*

^{a,b,c,g,h,i} Departemen Teknologi Multimedia Kreatif, Politeknik Elektronika Negeri Surabaya, Surabaya, Indonesia

^{c,e,f} Departemen Teknik Informatika Komputer, Politeknik Elektronika Negeri Surabaya, Surabaya, Indonesia

email: ^a kholid@pens.ac.id, ^b sudaryanto@pens.ac.id, ^c iwansyarif@pens.ac.id, ^d zakha@pens.ac.id, ^e andhik@pens.ac.id, ^f pras@pens.ac.id, ^g rendra@pens.ac.id, ^h aryazaky@gt.student.pens.ac.id, ⁱ rizky@pens.ac.id

*Corresponding Author

Article History:

Received 1 December 2024

Reviewed 20 December 2024

Revised 18 January 2025

Accepted 2 May 2025



Lisensi: cc-by-sa

ABSTRACT

The development of a religious chatbot is crucial for fostering good understanding and mitigating potential social conflicts arising from religious misunderstandings. This study introduces the Intelligent Religious Assistant ChatBot, developed using the Retrieval-Augmented Generation (RAG) method within a Large Language Model (LLM). RAG combines information retrieval from authoritative sources integrated into a database with text generation to produce user-specific answers. The application has been fully developed and performance-tested, demonstrating its capability to provide accurate and contextually appropriate answers based on authoritative Islamic references endorsed by major Indonesian Islamic organizations, including Nahdlatul Ulama, Muhammadiyah, and Indonesian Ulema Council (MUI). The score of EMasjid quality answer that not use naive RAG is 0,355 while the score of EMasjid quality answer that use naive RAG is 0,681. Its mean that Naive RAG that used in EMasjid can increase its answer quality.

Keywords: Religious Chatbot; Religious Understanding; RAG (Retrieval-Augmented Generation); Artificial Intelligence

1.1. Introduction

Misinterpretation of religious teachings often becomes a source of societal issues, ranging from individual or group conflicts and the spread of misinformation to social fragmentation. Several studies affirm that radicalism stems from distorted or excessive religious understanding. For instance, KH. Said Aqil Siradj, former Chairman of PBNU, highlights that radicalism originates from ghuluw, an extreme reaction to religious matters, leading to unreasonable behaviors and radical actions [1]. Similarly, Nurjannah states that religious texts, inherently neutral, are often interpreted selectively, prioritizing violent over peaceful verses, fueling radical ideologies [2]. A report by Pew Research Center identifies religion-based tensions as a major conflict driver, particularly in religiously diverse regions [3]. Furthermore, the anonymity afforded by digital platforms increases the likelihood of aggressive and harmful behavior, as noted by Kuncoro et al. [4].

The solution to these challenges lies in the development of a system capable of delivering valid, reliable, and universally accessible information. Research highlights that providing valid electronic information is an effective measure to combat hoaxes and misinformation [5]. Such a system must offer accurate, credible answers and be easily accessible. Artificial intelligence-based technologies, such as

chatbots, present a viable solution due to their accessibility and the ability to integrate authoritative references.

In response, we have developed EMasjid, a Smart Assistant ChatBot application. EMasjid is designed to assist users in obtaining accurate religious information aligned with the teachings of the four major Islamic schools of thought: Hanafi, Maliki, Shafi'i, and Hanbali. The application functions as a virtual assistant, responding to user queries based on pre-stored authoritative texts. All responses are derived from a database populated with validated texts from the four schools of thought, to ensuring its accuracy and reliability.

Religious ChatBot applications based on AI (Artificial Intelligence) have been widely used and proven to facilitate public access to religious knowledge, and have been proven to be able to provide relevant answers [6]. Other studies have shown that the use of AI or other information technology media can facilitate and expand access to learning materials [7], and allow personalization of information materials according to each user [8], in addition, the use of AI has also been proven to produce interactive information materials [9]. Delivering informative material through gadgets or other information media is important, considering that exposure to gadgets (or other information media) has been proven to influence human behavior [10]. One example of delivering informative material is by modifying it into a game, for example a number recognition game with integral projection to increase user interaction [11], or a role play game, such as the Habertan waste management education game which can be used to educate the public in waste management [12]. Delivering material in an interactive way such as playing games or also known as gamification has been proven to increase user enthusiasm [13]. With interactive delivery, as well as valid references, EMasjid is expected to help the public understand Islam better and reduce the potential for misunderstandings that can ultimately lead to conflict.

To generate accurate and relevant responses, EMasjid use the Retrieval-Augmented Generation (RAG) method. RAG allows the chatbot to extract relevant information from external sources and integrate it into the text generation process. Research indicates that RAG not only delivers precise answers but also provides additional insights beneficial to users [14]. Using RAG method, EMasjid combines the capabilities of language models with valid references from religious texts. This is to ensures that responses given by EMasjid are align with users' questions while being contextually accurate and grounded in valid sources, so that enhancing its reliability and trust.

We conducted interviews with representatives of the largest Islamic community organizations in Indonesia, they are Nahdlatul Ulama and Muhammadiyah [15] and MUI (Council of Indonesian Ulema), to make sure the data that used for this application. Last, we use the RAGAS framework to test the quality of EMasjid's answers. RAGAS is a framework designed to evaluate and improve the performance of LLM and RAG systems. The RAG used in EMasjid is naive RAG.

With the advancement of artificial intelligence, language model-based and machine learning approaches have become more dominant. These approaches indicate that chatbots capable of processing complex queries require the integration of NLP techniques to provide more accurate and relevant answers. In the context of religious knowledge, chatbots can serve as effective tools to educate the public by provide credible information. AI-based chatbots have been extensively studied and developed in several sectors such as the provision of religious information. For example, one study that focused on developing a chatbot for Islamic jurisprudence (fiqh) related to buying and selling using the LangChain method [6]. This research aimed to facilitate and expedite users in obtaining answers for their religious questions, particularly regarding fiqh on buying and selling. The LangChain method was able to match between questions and answers, ensuring that the provided responses closely aligned with the users' queries.

Another study employed the RAG (Retrieval-Augmented Generation) method and LangChain to develop a web-based Hadith Q&A system, utilizing nine Hadith books as reference data. The system's advantage lies in its ability to not only answer user queries but also accompanied by its references from the nine fiqh books [16]. This research indicates that chatbots using RAG, integrated with other methods,

perform effectively in answering queries based on the sources stored in the database. In this study, we ensured that the references used are authoritative religious texts from various sects, recommended by prominent Islamic organizations in Indonesia, including Nahdlatul Ulama, Muhammadiyah, and MUI (Indonesian Council of Ulama).

One of the objectives of developing the Islamic chatbot application, EMasjid, is to educate the public and prevent religious misunderstandings by providing valid religious information derived from authoritative sources. A study analyzing the impact of AI chatbots on Islamic education in Indonesia concluded that AI chatbots offer positive outcomes, including fast information delivery, enhanced user engagement, improved understanding of Islamic concepts, and content personalization [8]. Another study demonstrated that AI in religious education can boost motivation through personalized learning and broader access to resources, though challenges like technological infrastructure and educator preparedness remain [17].

The Retrieval-Augmented Generation (RAG) method is a prominent approach designed to enhance a chatbot's ability to provide accurate and relevant responses for user's questions. RAG integrates two main techniques: retrieval of information from external data sources and the generation of text by combining the retrieved information. Research by Lewis et al. (2020) introduced RAG, demonstrating its effectiveness in tasks requiring deep and specific knowledge. RAG excels in extracting information from external databases or texts and enriching the generated content [18], significantly improving answer accuracy in fields such as healthcare or law.

In another study by Izacard and Grave, RAG was implemented in a document retrieval model to generate deeper, more accurate content-based responses. Their results showed that RAG outperforms end-to-end models relying solely on generative approaches [19]. The ability of RAG to incorporate external sources of context enhances its reliability, particularly for scenarios requiring detailed explanations from valid sources.

Further, research by Karpukhin et al on machine learning-based information retrieval (Dense Passage Retrieval, DPR) supports the idea that integrating retrieval and generation strengthens chatbot output quality. RAG allows chatbots to access both internal and external data sources, improving the relevance and richness of responses [20].

Previous research has demonstrated that the RAG method significantly contributes to enhancing chatbot performance, particularly in providing answers to user queries with reliable references. In this study, the sources referenced include authoritative texts from the four major Islamic schools of thought, which are also recommended by prominent Islamic organizations in Indonesia, such as Nahdlatul Ulama, Muhammadiyah, and the Indonesian Council of Ulama (MUI).

The ability of RAG to access and retrieve information from external sources [21] enables the chatbot to provide responses that are not only generative but also grounded in valid and contextual references. This method enhances the accuracy and relevance of the answers by ensuring they are backed by reliable, valid sources, improving both the quality and trustworthiness of the information provided.

Thus, integrating the RAG method into the development of the EMasjid Smart Assistant Chatbot holds significant potential to enhance the accuracy, relevance, and credibility of the responses provided to users [22]. Additionally, RAG enables the chatbot to deliver answers alongside their respective sources, minimizing the potential for bias or misinterpretation, and ensuring that the information provided is both trustworthy and contextually accurate. This approach contributes to the overall reliability and effectiveness of the chatbot in delivering religious knowledge.

2. Method

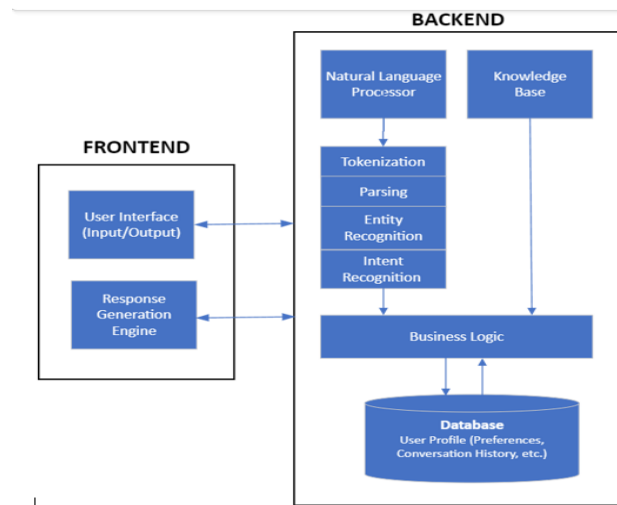


Fig 1. Architecture of Emasjid Islamic Chatbot

The Islamic chatbot application developed in this study is called EMasjid. Initially, the researcher inputs data from the authoritative texts of the four main Islamic schools of thought into the EMasjid database (knowledge base). Once the database is sufficiently populated, EMasjid can answer users' questions about Islam. The chatbot processes the input (user's questions) using Natural Language Processing (NLP) to understand the meaning and context of the queries. It then generates appropriate answers based on both the users' questions and the information stored in the database, which includes authoritative texts that also recommended by major Indonesian Islamic organizations such as Nahdlatul Ulama, Muhammadiyah, and MUI (Indonesian Council of Ulama).

This research aims to develop an intelligent assistant chatbot named EMasjid, which provides answers to religious knowledge questions based on references from the four major Islamic schools of thought (Hanafi, Maliki, Shafi'i, and Hanbali), utilizing the Retrieval-Augmented Generation (RAG) method.

This study involves the development of a software application designed to create a chatbot capable of providing accurate answers to religious questions. The chatbot is built to be a reliable tool for users seeking information related to religious knowledge, leveraging advanced methods to ensure the responses are valid and contextually appropriate.

The primary data used in EMasjid consists of texts from the authoritative religious books of the four major Islamic schools of thought, including Al-Hidayah (Hanafi), Al-Mudawwanah Al-Kubra (Maliki), Al-Umm (Shafi'i), and Al-Mughni (Hanbali). These texts were gathered from valid digital sources such as e-books and digitized manuscripts. After collection, the texts underwent preprocessing, which included tokenization, text normalization, removal of irrelevant words (stop words), and lemmatization, to optimize the texts for use in the RAG model.

The development of the Retrieval-Augmented Generation (RAG) model aims to combine the information retrieval capabilities from external sources with the generative capabilities of Large Language Models (LLM). The LLM model consists of two main components: the retriever, which is responsible for searching relevant texts from the processed books, and the generator, which produces answers based on the information retrieved. LLMs used in this research are GPT 4o Mini, SeaLLMs, Goekdeniz Qwen, Sailor, SmoLLM, IBM Granite, and Qwen2.5-1.5B. Then the RAG method used in this research is naive RAG.

The RAG model is trained to identify parts of the religious texts that align with the user's questions. This process involves selecting a retriever model capable of searching long texts and a generator model that can formulate responses based on the relevant data retrieved. The retriever ensures the retrieval of the most pertinent sections from the dataset, while the generator synthesizes a coherent and contextually accurate response using this information.

EMasjid is developed as an application utilizing a framework for both the backend and user interface. The design of EMasjid prioritizes ease of use for the general public, allowing users to input text-based questions. The system then generates answers based on references from the four major islamic schools of thought. This simple yet effective structure ensures accessibility while maintaining accuracy and validity in the provided responses.

To measure the answer quality of EMasjid, we test it using RAGAS Platform. RAGAS is a framework that designed to evaluate and improve the performance of LLM and RAG systems. We eparate the test by two categories, LLM system without RAG and LLM system with RAG. LLM system without RAG is tested by three parameters, semantic similarity, answer relevancy, and rouge score. EMasjid's LLM system with RAG is tested by six parameters, context recall, LLM context precision with reference, faithfulness, semantic similarity, answer relevancy, and rouge score.

Semantic similiarity show the similiarity between EMasjid's answer with ground truth. Value of semantic similiarity is in the range of 0 to 1. With bigger score signifies better alignment between answer and ground truth. Answer relevancy indicate how relevance the answer to its question. A lower score mean the answer are incomplete or contain redundant information and higher scores mean better relevancy. Rouge score is a set of metrics used to measures the overlap between answres and the reference.

Context recall is a metric to measures how many relevant documents or information were successfully retrieved. LLM context precision with reference is a metric to measures the relevance of answers based on its questions and external databases. The faithfulness metric measures how factually consistent a response is with the retrieved context.

3. Results and Discussion

EMasjid's system consist of main part, that is LLM (Large Language Model) and RAG (Retrieval Augmented Generation). LLM that used in EMasjid are GPT4o Mini, SeaLLMs, Qwen, Sailor, TBSmolLM, and Granite. While the RAG method that used in EMasjid is Naive RAG.

To meausre the quality of EMasjid's anwer, we test it by semantic similiarity, answer relevancy and rouge score metrics for EMasjid's system LLM without RAG. While EMasjid's system that using RAG is tested by context recall, LLM context precision with reference, faithfulness and semantic similiarity metrics.

The chart of testing result for EMasjid's LLM without RAG are shown in figure 2. The metrics used in this testing are semantic similiarity, answer relevancy and rouge score. The average score for all LLM is 0,355 while thehighest score is reached by GPT 4o Mini with 0,467.

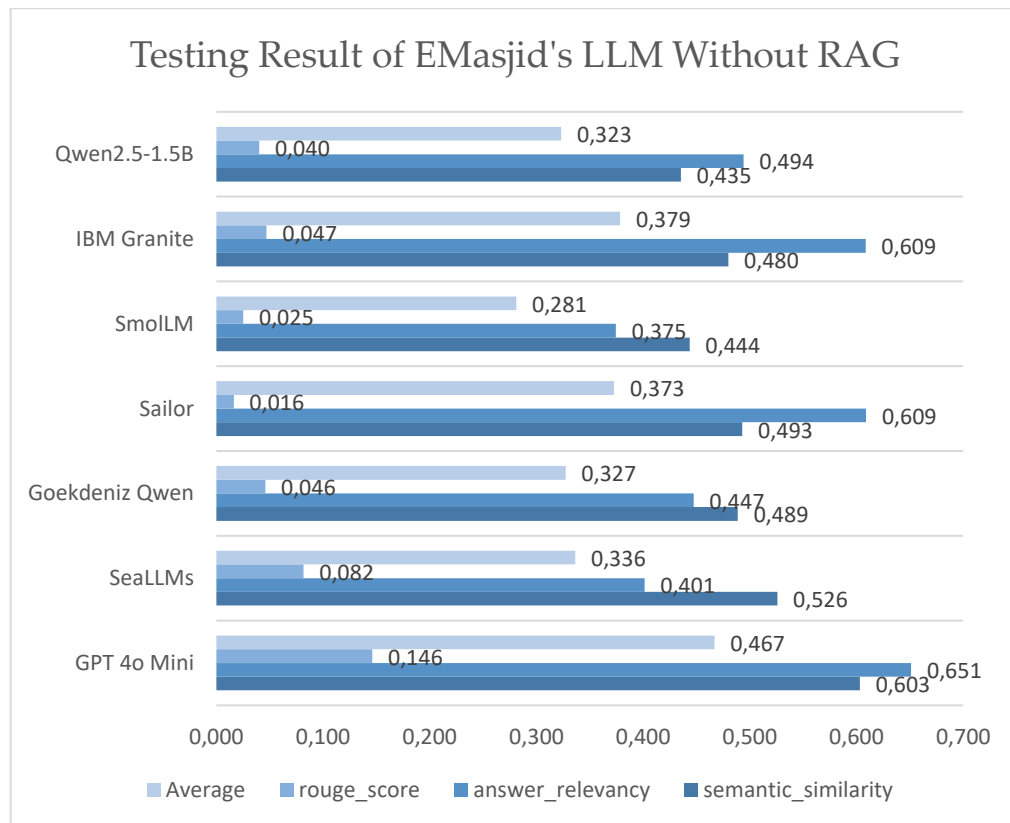


Fig 2. Testing Result of EMasjid's LLM Without RAG

The chart of testing result for EMasjid's LLM with naive RAG are shown in figure 3. The metrics used in this testing are context recall, LLM context precision with reference, faithfulness, semantic similarity, answer relevancy and rouge score. The average score for all LLM with naive RAG is 0,681 while the highest score is reached by GPT 4o Mini with 0,696.

The highest test score for EMasjid's LLM with naive RAG is reached by GPT 4o Mini with 0,696. Actually GPT 4o Mini is not lead at every metric, but the delta score for faithfulness an rouge score is so high so that can make GPT 4o Mini overall score is higher than others.

Average of all test score for EMasjid's LLM without RAG is 0,355 while average of all test score for EMasjid's LLM with naive RAG is 0,681. Its mean that naive RAG can increase the quality answer of EMasjid 91,8% than without RAG.

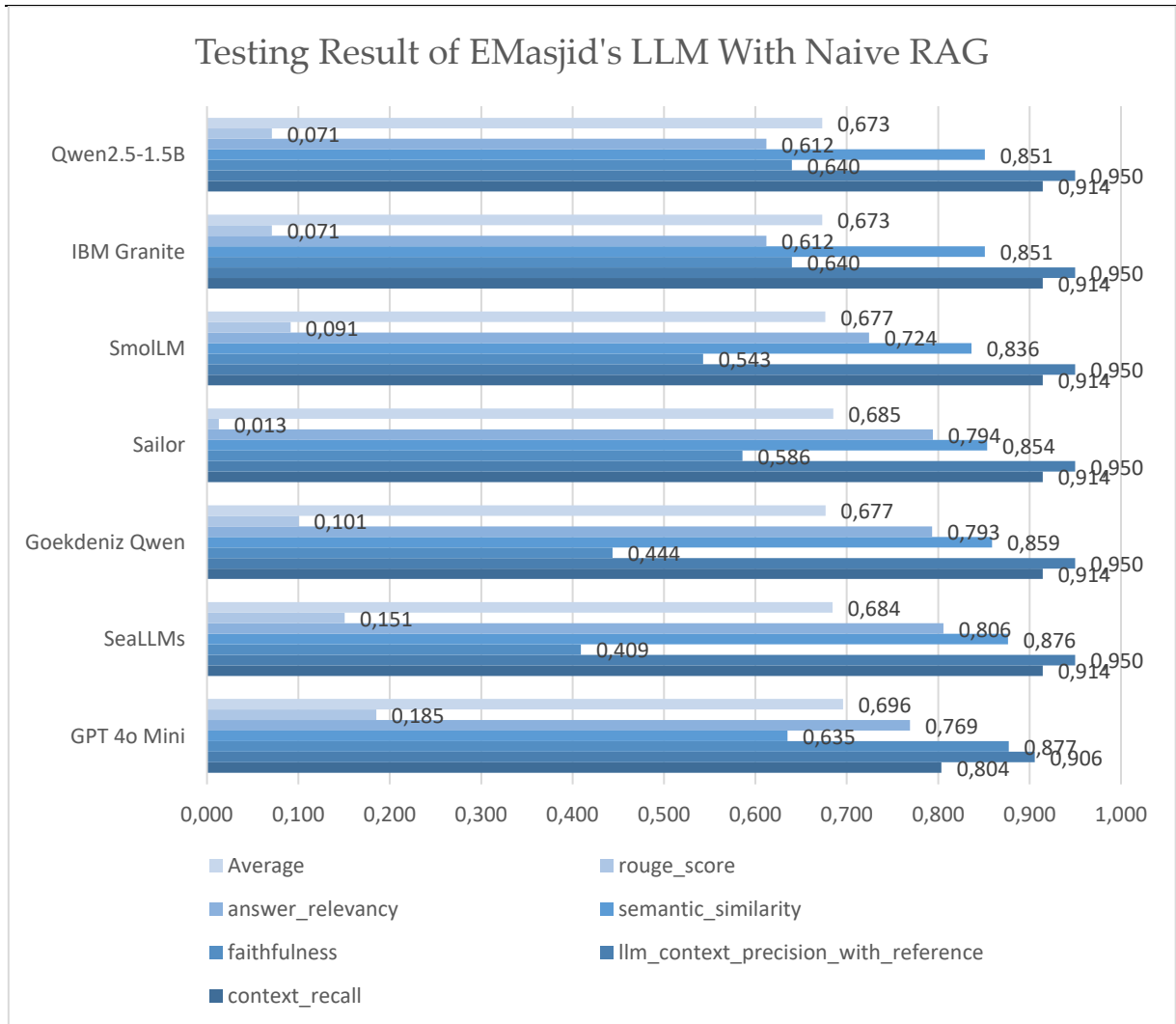


Fig 3. Testing Result of EMasjid's LLM With Naive RAG

4. Conclusions and Recommendations

Based on the results and analyses conducted, we can conclude that the EMasjid Islamic Chatbot Application has been successfully developed and operates effectively. Average of all test score for EMasjid's LLM without RAG is 0,355 while average of all test score for EMasjid's LLM with naive RAG is 0,681. Its mean that naive RAG can increase the quality answer of EMasjid 91,8% than without RAG.

For further development we suggest to enhance the dataset to expand the collection of authoritative texts and references integrated into the system. This expansion would provide greater diversity and depth of information, enabling the chatbot to offer more comprehensive and nuanced responses to user inquiries.

Improving language capabilities to enhance the chatbot's ability to comprehend the context and subtleties of user queries, particularly those posed in Arabic.

5. Acknowledgment

This research was funded through a vocational matching fund grant. For that we would like to thank to all levels of the Ministry of Education, Culture, Research and Technology (Kemendikbudristek), all level leader of the Politeknik Elektronika Negeri Surabaya (PENS), the Center

for Research and Community Service of PENS, PT Gerakan Ekonomi Indonesia Mandiri (PT Geni Mandiri), Nahdlatul Ulama, Muhammadiyah, MUI (Majelis Ulama Indonesia) and all who have assisted in this research.

7. References

- [1] S. Rohman and S. Nurhasanah, "Paham Radikalisme Berdasarkan Perspektif Agama," *Journal OF Terrorism Studies*, vol. 1, no. 1, pp. 18–32, 2019.
- [2] Nurjannah, "Faktor Pemicu Radikalisme Islam Atas Nama Dakwah," *Dakwah*, vol. XIV, no. 2, pp. 177–198, 2013, [Online]. Available: <https://ejournal.uin-suka.ac.id/dakwah/jurnaldakwah/article/view/2013.14202/260>
- [3] P. R. Center, "Globally, Social Hostilities Related to Religion Decline in 2019, While Government Restrictions Remain at Highest Levels," Pew Research Center. Accessed: Nov. 20, 2024. [Online]. Available: <https://www.pewresearch.org/religion/2021/09/30/globally-social-hostilities-related-to-religion-decline-in-2019-while-government-restrictions-remain-at-highest-levels/>
- [4] H. R. Kuncoro, D. Lupitasari, K. Hasanah, and E. Kurniawati, *Mengurai Ancaman Sekuritisasi Melalui Lensa Framing Dan Diskursus Di Media Sosial*, 1st ed., vol. 1, no. 1. LPPM UPNVY Press, 2023. [Online]. Available: http://scioteca.caf.com/bitstream/handle/123456789/1091/RED2017-Eng-8ene.pdf?sequence=12&isAllowed=y%0Ahttp://dx.doi.org/10.1016/j.regsciurbeco.2008.06.005%0Ahttps://www.researchgate.net/publication/305320484_SISTEM_PEMBETUNGAN_TERPUSAT_STRATEGI_MELESTARI
- [5] O. Lutfi and J. Jumino, "Peran Perpustakaan Nasional Republik Indonesia dalam Penyediaan Sumber Daya Informasi Elektronik sebagai Upaya Mengatasi Infodemi pada Masa Pandemi Covid-19," *ANUVA*, vol. 6, no. 2, pp. 141–162, 2022, [Online]. Available: <https://ejournal2.undip.ac.id/index.php/anuva/article/view/14390/7371>
- [6] Nurhapiza, N. S. Harahap, M. Fikry, and M. Affandes, "Penerapan Chatbot pada Aplikasi Web Tanya Jawab Tentang Fiqih," *Journal of Computer System and Informatics (JoSYC)*, vol. 5, no. 3, pp. 548–557, 2024, doi: 10.47065/josyc.v5i3.5148.
- [7] N. Nabila, E. I. Pramesti, and A. Sudaryanto, "Pengaruh Penerapan G-Suite bagi Perkembangan Pembelajaran di SMAN 1 Kalidawir Sidoarjo," *Jurnal Manajemen dan Ilmu Administrasi Publik (JMIAP)*, vol. 3, no. 4, pp. 329–332, Apr. 2021, doi: 10.24036/jmiap.v3i4.366.
- [8] R. N. T. N. S. P. N. A. A. A. M. Urva, "Dinamika Pembelajaran Pendidikan Agama Islam Berbasis Artificial Intelligence (AI)," *Sentikjar*, vol. 3, pp. 1–7, 2024, doi: 10.47435/sentikjar.v3i0.3131.

-
- [9] M. F. Zulrahman and H. Syahputra, "PEMANFAATAN ARTIFICIAL INTELLIGENCE MARKUP LANGUAGE (AIML) DAN LATENT SEMANTIC ANALYSIS (LSA) DALAM PENGEMBANGAN CHATBOT E-EDUCATION," *Journal of Information Technology and Computer Science (INTECOMS)*, vol. 6, no. 1, pp. 36–43, 2023, [Online]. Available: <https://journal.ipm2kpe.or.id/index.php/INTECOM/article/view/5459/3295>
- [10] J. F. Yusuf and R. Destiwati, "PENGARUH GADGET TERHADAP PERILAKU ANAK SELAMA MASA PANDEMI COVID-19 PADA SISWA KELAS 2 – 4 DI SEKOLAH DASAR ISLAM TERPADU ANNI'MAH," *e-Proceeding of Management*, vol. 8, no. 5, 2021, [Online]. Available: <https://annimah.sch.id>
- [11] A. Sudaryanto *et al.*, "Learning number game with interactive projection for students of Sepuluh Nopember Surabaya Kindergarten," *KACANEGARA Jurnal Pengabdian pada Masyarakat*, vol. 6, no. 4, p. 481, Nov. 2023, doi: 10.28989/kacanegara.v6i4.1681.
- [12] A. Sudaryanto, K. Fathoni, R. Y. Hakkun, and F. Revindasari, "Peningkatan Kesadaran Kebersihan elalui Game ' Habertan ' pada Siswa Siswi Pesantren Muhyiddin Gebang Putih , Sukolilo , Surabaya," *JURPIKAT*, vol. 5, no. 4, pp. 1425–1440, 2024, doi: <https://doi.org/10.37339/jurpikat.v5i4.2032>.
- [13] M. F. Hasan, F. N. Fajri, and M. Muafi, "Game Edukasi Simulasi Pengenalan Reaksi Unsur Kimia Dengan Lingkungan Berbasis Virtual Reality," *NJCA (Nusantara Journal of ...)*, vol. 9, no. 1, pp. 10–18, 2024, doi: <http://dx.doi.org/10.36564/njca.v9i1.365>.
- [14] I. Ichsanudin, R. Pratama, and B. Sisephaputra, "Pengembangan Sistem Helpdesk Menggunakan Chatbot Dengan Metode Retrieval-Augmented Generation (RAG)," *JINACS*, vol. 06, no. 03, pp. 696–710, 2024, [Online]. Available: <https://ejournal.unesa.ac.id/index.php/jinacs/article/view/64408/48747>
- [15] A. R. Sitoresmi, "Daftar Organisasi Islam di Indonesia yang Paling Banyak Pengikutnya," *Liputan6.com*. [Online]. Available: <https://www.liputan6.com/hot/read/5357846/daftar-organisasi-islam-di-indonesia-yang-paling-banyak-pengikutnya>
- [16] Muhammad Irfan Syah, Nazruddin Safaat Harahap, Novriyanto, and Suwanto Sanjaya, "Penerapan Retrieval Augmented Generation Menggunakan Langchain Dalam Pengembangan Sistem Tanya Jawab Hadis Berbasis Web," *ZONAsi: Jurnal Sistem Informasi*, vol. 6, no. 2, pp. 370–379, 2024, doi: 10.31849/zn.v6i2.19940.
- [17] A. R. Ramadhan, "Phenomena Of Chatbot Artificial Intelligence And Its Impact On Islamic Religious Education," in *4rd UIN Imam Bonjol International Conference on Islamic Education*, 2023, p. 67. [Online]. Available: <https://ibicie.uinib.ac.id/index.php/ibicie/article/view/62%0Ahttps://ibicie.uinib.ac.id/index.php/ibicie/article/download/62/59>
-

- [18] P. Lewis *et al.*, "Retrieval-augmented generation for knowledge-intensive NLP tasks," *Adv Neural Inf Process Syst*, vol. 2020-Decem, no. NeurIPS, 2020.
- [19] G. Izacard and E. Grave, "Leveraging passage retrieval with generative models for open domain question answering," in *EACL 2021 - 16th Conference of the European Chapter of the Association for Computational Linguistics, Proceedings of the Conference*, Association for Computational Linguistics, 2021, pp. 874–880. doi: 10.18653/v1/2021.eacl-main.74.
- [20] V. Karpukhin *et al.*, "Dense passage retrieval for open-domain question answering," in *Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, 2020, pp. 6769–6781. doi: 10.18653/v1/2020.emnlp-main.550.
- [21] K. Pranusha, P. Vamsi, and K. Raja, "External Information on Large Linguistic Models Utilizing Retrieval Enhanced Generation (RAG)," *International Journal of Engineering Innovations in Advanced Technology*.
- [22] D. Soong *et al.*, "Improving accuracy of GPT-3/4 results on biomedical data using a retrieval-augmented language model," *PLOS Digital Health*, vol. 3, no. 8, Aug. 2024, doi: 10.1371/journal.pdig.0000568.