

Chatbot Fiqih Berbasis Retrieval-Augmented Generation pada Kitab Safinatun Najah

^{1*}Muhammad Ichsan, ²Ardi sanjaya, ³Ahmad Bagus Setiawan

^{1,2,3} Teknik Informatika, Universitas Nusantara PGRI Kediri

E-mail: ¹muhammadichsan11tkj@gmail.com, ²dersky@gmail.com,

³ahmadbagus@unpkediri.ac.id

Penulis Korespondens : Muhammad Ichsan

Abstrak— Perkembangan teknologi kecerdasan buatan mendorong pemanfaatan *chatbot* sebagai media penyedia informasi keagamaan berbasis digital. Namun, penggunaan *Large Language Model* (LLM) masih menghadapi permasalahan berupa *hallucination*, yaitu kondisi ketika model menghasilkan informasi yang tidak sesuai dengan sumber rujukan. Penelitian ini bertujuan mengembangkan *chatbot* fiqih berbasis metode *Retrieval-Augmented Generation* (RAG) dengan Kitab Safinatun Najah sebagai sumber pengetahuan utama. Sistem dikembangkan menggunakan model ADDIE (*Analysis, Design, Development, Implementation, Evaluation*) dan diimplementasikan menggunakan Streamlit, LangChain, ChromaDB, model *embedding multilingual-e5-large*, serta Gemini sebagai *Large Language Model*. Metode RAG diterapkan dengan menggabungkan proses *retrieval* dokumen relevan dan proses generasi jawaban sehingga respons yang dihasilkan tetap mengacu pada isi kitab. Evaluasi sistem difokuskan pada kinerja proses *retrieval* menggunakan 250 data pengujian yang berasal dari 50 pertanyaan utama dengan lima variasi pertanyaan pada setiap data. Hasil pengujian menunjukkan nilai *retrieval accuracy*, *precision*, *recall*, dan *F1-score* sebesar 76,80%, serta *average similarity* sebesar 0,7644. Hasil penelitian menunjukkan bahwa metode RAG mampu memahami berbagai variasi pertanyaan pengguna dan menemukan informasi yang relevan berdasarkan isi Kitab Safinatun Najah. Dengan demikian, pendekatan RAG berpotensi meningkatkan kualitas sistem tanya jawab fiqih sekaligus mengurangi risiko *hallucination* pada model generatif.

Kata Kunci— *Chatbot, Fiqih, Kecerdasan Buatan, Retrieval-Augmented Generation, Safinatun Najah.*

Abstract— The rapid advancement of artificial intelligence technology has encouraged the adoption of chatbots as a medium for providing digital Islamic knowledge services. However, Large Language Models (LLMs) still face the challenge of hallucination, where generated responses may not be consistent with the referenced knowledge source. This study aims to develop a fiqh chatbot based on the Retrieval-Augmented Generation (RAG) approach using the Safinatun Najah book as the primary knowledge source. The system was developed using the ADDIE (*Analysis, Design, Development, Implementation, Evaluation*) model and implemented with Streamlit, LangChain, ChromaDB, the multilingual-e5-large embedding model, and Gemini as the Large Language Model. The RAG approach combines document retrieval and response generation processes to ensure that generated answers remain consistent with the source material. System evaluation focused on retrieval performance using 250 testing queries derived from 50 primary questions, each expanded into five paraphrased variations. The evaluation results achieved a retrieval accuracy, precision, recall, and F1-score of 76.80%, with an average similarity score of 0.7644. These findings indicate that the RAG approach is capable of understanding various user query formulations and retrieving relevant information from the Safinatun Najah knowledge base. Therefore, RAG has the potential to improve the quality of fiqh question-answering systems while reducing the risk of hallucination in generative language models.

Keywords— *Artificial Intelligence, Chatbot, Fiqh, Retrieval-Augmented Generation, Safinatun Najah.*

This is an open access article under the CC BY-SA License.



I. PENDAHULUAN

Perkembangan teknologi *Artificial Intelligence* telah mendorong pemanfaatan sistem chatbot pada berbagai bidang, termasuk pendidikan dan penyediaan informasi keagamaan. Sistem chatbot memungkinkan pengguna memperoleh informasi secara cepat melalui interaksi berbasis bahasa alami. Namun, model bahasa generatif masih memiliki kelemahan berupa fenomena *hallucination*, yaitu kondisi ketika sistem menghasilkan informasi yang tidak sesuai dengan sumber rujukan yang valid. Permasalahan tersebut menjadi tantangan utama dalam pengembangan sistem tanya jawab berbasis pengetahuan karena dapat menurunkan tingkat keakuratan informasi yang diberikan kepada pengguna[1],[2]. Oleh karena itu, diperlukan pendekatan yang mampu menghubungkan model bahasa dengan sumber pengetahuan yang terpercaya agar jawaban yang dihasilkan tetap relevan dan sesuai dengan referensi yang digunakan.

Salah satu pendekatan yang banyak digunakan untuk mengurangi permasalahan tersebut adalah *Retrieval-Augmented Generation* (RAG). Metode ini menggabungkan proses pencarian dokumen yang relevan dengan kemampuan generatif *Large Language Model* sehingga jawaban yang dihasilkan lebih akurat, kontekstual, dan sesuai dengan sumber pengetahuan yang tersedia [1], [3]. Pada tahap pencarian informasi, teknik *semantic embedding* digunakan untuk mengukur kemiripan makna antara pertanyaan pengguna dan dokumen yang tersimpan dalam basis pengetahuan [4].

Dalam bidang pendidikan Islam, khususnya pembelajaran fiqih, proses pencarian informasi masih banyak dilakukan secara manual melalui kitab-kitab rujukan. Kitab kuning merupakan salah satu sumber utama yang digunakan dalam proses pembelajaran di lingkungan pesantren. Salah satu kitab yang banyak digunakan adalah *Safinatun Najah*, yang memuat berbagai pembahasan dasar mengenai *thaharah*, sholat, puasa, zakat, dan hukum-hukum fiqih lainnya. Namun, proses pencarian informasi dalam kitab secara manual sering memerlukan waktu yang cukup lama, terutama bagi pengguna yang belum memahami struktur pembahasan kitab secara menyeluruh. Seiring perkembangan teknologi digital, diperlukan inovasi pembelajaran yang mampu mempermudah akses terhadap sumber-sumber keilmuan Islam tanpa mengurangi keakuratan materi yang dipelajari [5].

Kitab *Safinatun Najah* merupakan salah satu kitab fiqih dasar mazhab Syafi'i yang banyak digunakan dalam pembelajaran di pesantren. Pencarian informasi pada kitab tersebut umumnya masih dilakukan secara manual sehingga membutuhkan waktu yang relatif lama. Oleh karena itu, penelitian ini mengembangkan sistem *chatbot* fiqih berbasis metode *Retrieval-Augmented Generation* (RAG) dengan memanfaatkan Kitab *Safinatun Najah* sebagai sumber pengetahuan utama.

Penelitian ini bertujuan untuk mengimplementasikan metode *Retrieval-Augmented Generation* pada sistem *chatbot* fiqih guna membantu proses pencarian informasi secara lebih cepat dan kontekstual. Selain itu, penelitian ini juga mengevaluasi kinerja proses *retrieval* menggunakan 250 data pengujian yang berasal dari 50 pertanyaan utama dengan lima variasi pertanyaan pada setiap data. Evaluasi dilakukan menggunakan metrik *retrieval accuracy*, *precision*, *recall*, *F1-score*, dan *average similarity* untuk mengukur kemampuan sistem dalam menemukan dokumen yang relevan berdasarkan isi Kitab *Safinatun Najah*. Hasil penelitian diharapkan dapat menjadi referensi dalam pengembangan sistem informasi keagamaan berbasis kecerdasan buatan yang lebih akurat dan sesuai dengan sumber rujukan yang digunakan[3],[6].

II. METODE

Penelitian ini menggunakan metode *Research and Development* (R&D) dengan model ADDIE (*Analysis, Design, Development, Implementation, Evaluation*). Model ADDIE digunakan karena memiliki tahapan yang sistematis dalam pengembangan aplikasi berbasis teknologi informasi. Tahapan penelitian meliputi analisis kebutuhan, perancangan sistem, pengembangan aplikasi, implementasi sistem, dan evaluasi performa *chatbot*.

A. Analysis

Tahap analisis dilakukan untuk mengidentifikasi kebutuhan pengguna dan permasalahan dalam pencarian informasi fiqih yang masih dilakukan secara manual melalui Kitab *Safinatun Najah*. Pada tahap ini dilakukan pengumpulan sumber data berupa materi kitab yang akan digunakan sebagai basis pengetahuan *chatbot*. Selain itu, dilakukan analisis kebutuhan perangkat lunak dan perangkat keras yang diperlukan dalam pengembangan sistem.

B. Design

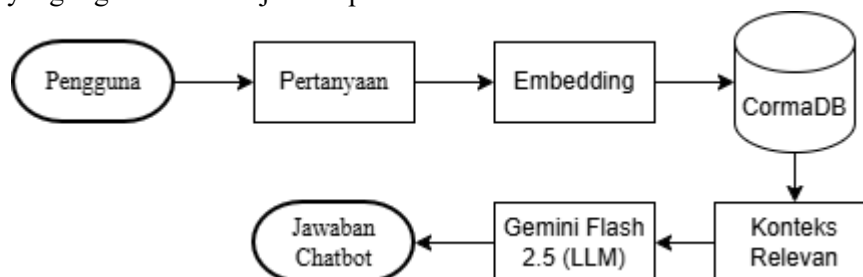
Tahap perancangan dilakukan dengan menyusun arsitektur sistem *chatbot*, alur proses *Retrieval-Augmented Generation* (RAG), struktur basis pengetahuan, serta rancangan antarmuka pengguna. Pada tahap ini dirancang pula mekanisme pencarian dokumen menggunakan teknik *semantic embedding* dan penyimpanan data pada basis data vektor.

C. Development

Tahap pengembangan dilakukan menggunakan bahasa pemrograman *Python*. Isi Kitab *Safinatun Najah* diproses menjadi data digital, kemudian dilakukan proses *chunking*, pembentukan *embedding* menggunakan model *multilingual-e5-large*, serta penyimpanan data ke dalam basis data vektor ChromaDB.

Penelitian ini menerapkan metode *Retrieval-Augmented Generation* (RAG) dalam pengembangan *chatbot* fiqih. Metode RAG menggabungkan proses pencarian informasi (*retrieval*) dengan kemampuan generatif *Large Language Model* (LLM) sehingga jawaban yang dihasilkan dapat mengacu pada sumber pengetahuan yang tersedia [1],[7].

Pada saat pengguna mengajukan pertanyaan, sistem akan mengubah pertanyaan menjadi representasi vektor, kemudian melakukan pencarian dokumen yang memiliki tingkat kemiripan tertinggi pada basis data vektor. Dokumen yang relevan digunakan sebagai konteks bagi model Gemini untuk menghasilkan jawaban sesuai isi Kitab *Safinatun Najah*. Alur sistem yang digunakan ditunjukkan pada Gambar 1 Alur sisitem.



Gambar 1. Alur sisitem

Tahap pencarian dokumen menggunakan metode *Cosine Similarity* untuk menghitung tingkat kemiripan antara vektor pertanyaan dan vektor dokumen. Persamaan *Cosine Similarity* ditunjukkan pada Persamaan (1).

$$\text{Cosine Similarity}(A, B) = \frac{A \cdot B}{|A||B|} \quad (1)$$

D. Implementation

Tahap implementasi dilakukan dengan menerapkan sistem *chatbot* yang telah dikembangkan ke lingkungan pengujian. Pengguna dapat mengajukan pertanyaan menggunakan bahasa alami melalui antarmuka sistem. Selanjutnya sistem melakukan proses pencarian dokumen yang relevan dan menghasilkan jawaban berdasarkan konteks yang ditemukan pada basis pengetahuan.

Model Gemini digunakan sebagai *Large Language Model* dengan parameter *temperature* sebesar 0 untuk menghasilkan jawaban yang lebih konsisten dan deterministik[8]. Penggunaan konteks hasil *retrieval* bertujuan untuk mengurangi kemungkinan terjadinya *hallucination* dan meningkatkan kesesuaian jawaban terhadap sumber rujukan [1], [9].

E. Evaluation

Tahap *evaluation* dilakukan untuk mengukur kinerja sistem chatbot fiqih berbasis *Retrieval-Augmented Generation* (RAG) dalam menemukan informasi yang relevan dari basis pengetahuan Kitab Safinatun Najah. Evaluasi difokuskan pada proses *retrieval*, yaitu kemampuan sistem dalam mengambil dokumen yang sesuai dengan pertanyaan pengguna.

Data pengujian terdiri atas 50 pertanyaan utama yang disusun berdasarkan materi fiqih dalam Kitab Safinatun Najah dan divalidasi oleh ahli pada bidang fiqih. Pertanyaan-pertanyaan tersebut mencakup beberapa topik utama, seperti bersuci, wudhu, tayamum, salat, puasa, zakat, dan haji. Untuk menguji kemampuan sistem dalam memahami variasi bahasa alami, setiap pertanyaan utama dikembangkan menjadi lima variasi pertanyaan dengan makna yang sama tetapi menggunakan susunan kalimat yang berbeda. Dengan demikian, total data pengujian yang digunakan dalam penelitian ini berjumlah 250 pertanyaan.

Variasi pertanyaan digunakan untuk mensimulasikan berbagai bentuk pertanyaan yang mungkin diajukan oleh pengguna pada saat berinteraksi dengan *chatbot*. Distribusi data pengujian ditujukan pada Tabel 1. Distribusi Data Pengujian.

Tabel 1. Distribusi Data Pengujian

No	Keterangan	Jumlah
1	Pertanyaan Pertanyaan utama	50
2	Variasi pertanyaan per data	5
3	Total data pengujian	250

Pertanyaan pada proses pengujian, setiap pertanyaan diubah menjadi representasi vektor menggunakan model *multilingual-e5-large*. Selanjutnya sistem melakukan *semantic search* pada basis data vektor untuk memperoleh dokumen dengan tingkat kemiripan tertinggi. Hasil *retrieval* kemudian dibandingkan dengan dokumen acuan (*ground truth*) berdasarkan identitas dokumen (*id*).

Kinerja sistem diukur menggunakan metrik *retrieval accuracy*, yaitu perbandingan antara jumlah *retrieval* yang benar dengan total data pengujian. Persamaan *retrieval accuracy* ditunjukkan pada Persamaan (2).

$$Accuracy = \frac{Jumlah\ Retrieval\ Benar}{Total\ Data\ Pengujian} \quad (2)$$

Selain itu, penelitian ini menghitung nilai *average similarity* untuk mengetahui tingkat kedekatan semantik antara pertanyaan dan dokumen yang ditemukan sistem. Nilai yang semakin mendekati 1 menunjukkan tingkat kesesuaian semantik yang semakin tinggi.

Melalui evaluasi tersebut dapat diketahui tingkat keberhasilan metode *Retrieval-Augmented Generation* dalam memahami variasi pertanyaan pengguna dan menemukan informasi fiqih yang relevan berdasarkan isi Kitab *Safinatun Najah*.

III. HASIL DAN PEMBAHASAN

A. Implementasi Sistem

Implementasi sistem *chatbot* fiqih berbasis *Retrieval-Augmented Generation* (RAG) dilakukan dengan mengintegrasikan model *embedding multilingual-e5-large*, basis data vektor ChromaDB, serta model generatif Gemini. Sistem dikembangkan menggunakan bahasa pemrograman Python dan antarmuka berbasis Streamlit untuk memudahkan interaksi pengguna.

Pada saat pengguna mengajukan pertanyaan, sistem terlebih dahulu mengubah teks pertanyaan menjadi representasi vektor menggunakan model *embedding*. Vektor pertanyaan kemudian dibandingkan dengan seluruh vektor dokumen yang tersimpan pada ChromaDB melalui proses *semantic search*. Dokumen yang memiliki tingkat kemiripan tertinggi dipilih sebagai hasil *retrieval* dan digunakan sebagai konteks bagi model Gemini untuk menghasilkan jawaban.

Penerapan metode *Retrieval-Augmented Generation* memungkinkan sistem menghasilkan jawaban yang lebih relevan karena model bahasa memperoleh konteks langsung dari dokumen yang tersimpan pada basis pengetahuan [1], [10]. Pendekatan ini juga dapat mengurangi risiko *hallucination* yang sering muncul pada model bahasa generatif tanpa dukungan sumber informasi yang jelas [10], [9].

Sumber pengetahuan yang digunakan dalam penelitian ini berasal dari Kitab *Safinatun Najah* yang telah melalui proses *chunking*, pembentukan *embedding*, dan penyimpanan ke dalam basis data vektor. Dengan mekanisme tersebut, sistem tidak hanya melakukan pencarian berbasis kata kunci, tetapi juga mempertimbangkan kesamaan makna antar kalimat melalui representasi vektor semantik [4].

B. Hasil Pengujian Retrieval

Pengujian dilakukan untuk mengevaluasi kemampuan sistem dalam menemukan dokumen yang sesuai dengan pertanyaan pengguna pada proses *retrieval*. Data pengujian yang digunakan berjumlah 250 pertanyaan yang berasal dari 50 pertanyaan utama dengan lima variasi pertanyaan pada setiap data. Setiap pertanyaan diuji terhadap basis data vektor dan hasil *retrieval* dibandingkan dengan dokumen acuan (*ground truth*) berdasarkan identitas dokumen (*id*).

Hasil pengujian menunjukkan bahwa dari 250 data yang diuji, sistem berhasil menemukan dokumen yang sesuai pada 192 pertanyaan, sedangkan 58 pertanyaan lainnya menghasilkan dokumen yang tidak sesuai dengan dokumen acuan.

Hasil pengujian sistem ditunjukkan melalui Tabel 3. Hasil Pengujian Retrieval berikut:

Tabel 3. Hasil Pengujian *Retrieval*

Parameter	Nilai
Total Data Pengujian	250
Retrieval Benar	192
Retrieval Salah	58
Retrieval Accuracy	0,7680
Average Similarity	0,7644

Berdasarkan hasil Tabel 3. Hasil Pengujian *Retrieval*, diperoleh nilai *retrieval accuracy* sebesar 0,7680 atau 76,80%. Nilai tersebut menunjukkan bahwa sebagian besar variasi pertanyaan yang diajukan pengguna dapat dipetakan ke dokumen yang sesuai pada basis pengetahuan. Selain itu, nilai *average similarity* sebesar 0,7644 menunjukkan bahwa secara umum pertanyaan pengguna memiliki tingkat kedekatan semantik yang cukup tinggi dengan dokumen yang ditemukan sistem.

Meskipun demikian, masih ditemukan beberapa kasus *retrieval* yang tidak sesuai. Kesalahan terjadi ketika dokumen yang ditemukan sistem berbeda dengan dokumen acuan (*ground truth*), meskipun memiliki kemiripan makna yang tinggi.

Sebagai contoh, pada pertanyaan "Apa saja yang bisa membuat wudhu kita batal?", sistem mengembalikan dokumen dengan *retrieved_id* = 12 sehingga hasil pengujian dinyatakan salah. Kondisi ini menunjukkan bahwa terdapat kemiripan semantik antar topik fiqh sehingga sistem memilih dokumen yang bukan merupakan dokumen acuan.

Secara keseluruhan, hasil pengujian menunjukkan bahwa metode *Retrieval-Augmented Generation* (RAG) mampu melakukan pencarian informasi fiqh dengan tingkat keberhasilan yang baik. Namun, optimalisasi proses *embedding*, *chunking*, dan *retrieval* masih diperlukan untuk meningkatkan akurasi pencarian dokumen [1], [4].

C. Analisis Hasil Retrieval

Berdasarkan hasil pengujian, sistem memperoleh *retrieval accuracy* sebesar 76,80% dengan 192 data berhasil ditemukan sesuai *ground truth* dan 58 data tidak sesuai. Hasil ini menunjukkan bahwa metode *Retrieval-Augmented Generation* (RAG) mampu menemukan sebagian besar informasi yang relevan dari basis pengetahuan Kitab Safinatun Najah.

1. Analisis Keberhasilan Sistem

Sebanyak 192 data berhasil ditemukan dengan benar. Keberhasilan ini menunjukkan bahwa model *multilingual-e5-large* mampu memahami berbagai variasi pertanyaan pengguna dan memetakan pertanyaan tersebut ke dokumen yang sesuai melalui proses *semantic search* [4].

2. Analisis Kegagalan Sistem

Meskipun memperoleh hasil yang cukup baik, masih terdapat 58 data yang mengalami kesalahan *retrieval*. Kesalahan terjadi ketika dokumen yang memiliki tingkat kemiripan tertinggi bukan merupakan dokumen acuan yang telah ditetapkan pada data pengujian. Kondisi ini umumnya disebabkan oleh kemiripan makna antar topik fiqh yang tersimpan dalam basis data vektor.

3. Analisis Kasus Kesalahan Retrieval

Salah satu contoh kesalahan retrieval terjadi pada pertanyaan "Apa saja yang bisa membuat wudhu kita batal?". Pada data pengujian, pertanyaan tersebut seharusnya mengarah pada dokumen yang membahas pembatal wudhu. Namun sistem mengembalikan dokumen dengan *retrieved_id* = 12 dan nilai *score* sebesar 0,2872 sehingga status pengujian dinyatakan salah.

Proses kesalahan tersebut terjadi ketika pertanyaan pengguna terlebih dahulu diubah menjadi vektor oleh model *multilingual-e5-large*. Selanjutnya sistem mencari dokumen yang memiliki jarak vektor paling dekat pada basis data ChromaDB. Pada kasus ini, dokumen dengan *retrieved_id* = 12 memiliki kemiripan semantik yang dianggap lebih tinggi oleh model dibandingkan dokumen acuan yang sebenarnya. Akibatnya sistem memilih dokumen yang salah meskipun topik yang dibahas masih memiliki keterkaitan dengan materi fiqih yang sama.

Kondisi tersebut menunjukkan bahwa kemiripan makna antar materi fiqih dapat menyebabkan terjadinya *semantic overlap* sehingga dokumen yang dipilih sistem tidak selalu sesuai dengan dokumen acuan.

4. Analisis Metrik Evaluasi

Nilai *accuracy*, *precision*, *recall*, dan *F1-score* pada penelitian ini sama-sama sebesar 76,80%. Hal tersebut terjadi karena evaluasi dilakukan menggunakan skema *top-1 retrieval*, yaitu setiap pertanyaan hanya menghasilkan satu dokumen hasil pencarian. Pada kondisi tersebut, jumlah dokumen yang berhasil ditemukan dengan benar menjadi dasar utama perhitungan seluruh metrik sehingga nilai yang dihasilkan identik.

Secara keseluruhan, hasil penelitian menunjukkan bahwa metode RAG cukup efektif untuk pencarian informasi fiqih. Namun, optimasi pada proses *embedding*, *chunking*, dan *retrieval* masih diperlukan untuk meningkatkan akurasi pencarian serta mengurangi jumlah kesalahan *retrieval* [1], [10].

D. Keterbatasan Penelitian

Penelitian ini memiliki beberapa keterbatasan yang dapat mempengaruhi kinerja sistem. Pertama, basis pengetahuan yang digunakan masih terbatas pada satu kitab, yaitu Kitab Safinatun Najah, sehingga cakupan informasi fiqih yang dapat dijawab oleh sistem masih terbatas pada materi yang terdapat dalam kitab tersebut. Kedua, proses *retrieval* masih menggunakan pendekatan *top-1 retrieval* sehingga sistem hanya mengambil satu dokumen dengan tingkat kemiripan tertinggi. Kondisi ini berpotensi menyebabkan kesalahan *retrieval* ketika terdapat beberapa dokumen yang memiliki kemiripan semantik yang berdekatan. Ketiga, proses *chunking* dan parameter pencarian pada basis data vektor belum dioptimalkan secara khusus untuk domain fiqih. Selain itu, variasi pertanyaan yang digunakan dalam pengujian masih disusun berdasarkan 50 pertanyaan utama sehingga belum sepenuhnya merepresentasikan seluruh kemungkinan variasi bahasa yang digunakan oleh pengguna.

E. Penelitian Selanjutnya

Pengembangan penelitian selanjutnya dapat dilakukan dengan memperluas basis pengetahuan menggunakan lebih dari satu kitab fiqih sehingga informasi yang tersedia menjadi lebih lengkap. Selain itu, optimasi model *embedding* yang lebih spesifik terhadap domain keislaman dapat dilakukan untuk meningkatkan kemampuan sistem dalam memahami istilah-istilah fiqih. Pengembangan metode *hybrid retrieval* yang

mengombinasikan *semantic search* dan pencarian berbasis kata kunci juga dapat diterapkan untuk meningkatkan akurasi pencarian dokumen. Penelitian selanjutnya juga dapat menambahkan evaluasi berbasis pengguna (*human evaluation*) untuk mengukur kualitas jawaban yang dihasilkan sistem dari sisi relevansi, kejelasan, dan kemudahan dipahami.

IV. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, sistem *chatbot* fiqih berbasis *Retrieval-Augmented Generation* (RAG) berhasil diimplementasikan menggunakan model *multilingual-e5-large*, basis data vektor ChromaDB, dan model Gemini sebagai *Large Language Model*. Sistem mampu melakukan pencarian informasi berdasarkan isi Kitab Safinatun Najah melalui pendekatan *semantic search* sehingga jawaban yang dihasilkan lebih relevan terhadap konteks pertanyaan pengguna.

Hasil pengujian menggunakan 250 variasi pertanyaan yang berasal dari 50 pertanyaan utama menunjukkan bahwa sistem berhasil melakukan *retrieval* yang sesuai sebanyak 192 data dan mengalami kesalahan *retrieval* sebanyak 58 data. Berdasarkan hasil tersebut diperoleh nilai *retrieval accuracy*, *precision*, *recall*, dan *F1-score* sebesar 76,80%, serta nilai *average similarity* sebesar 0,7644. Hasil ini menunjukkan bahwa sistem mampu memahami sebagian besar variasi pertanyaan pengguna dan menemukan informasi yang relevan pada basis pengetahuan yang digunakan.

Meskipun demikian, masih ditemukan beberapa kasus kesalahan *retrieval* yang disebabkan oleh kemiripan semantik antar topik fiqih sehingga dokumen yang dipilih sistem tidak selalu sesuai dengan dokumen acuan. Oleh karena itu, optimasi proses *embedding*, *chunking*, dan strategi *retrieval* masih diperlukan untuk meningkatkan akurasi pencarian pada penelitian selanjutnya. Secara keseluruhan, penelitian ini menunjukkan bahwa pendekatan RAG efektif diterapkan pada *chatbot* fiqih sebagai media pencarian informasi keagamaan berbasis Kitab Safinatun Najah.

DAFTAR PUSTAKA

- [1] W. Fan *et al.*, "A Survey on RAG Meeting LLMs: Towards Retrieval-Augmented Large Language Models," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 6491–6501, 2024, doi: 10.1145/3637528.3671470.
- [2] M. Alshammary, M. N. Uddin, and L. Khan, "RFPG: Question-Answering from Low-Resource Language (Arabic) Texts using Factually Aware RAG," *Proc. - 2024 IEEE 10th Int. Conf. Collab. Internet Comput. CIC 2024*, no. Cic, pp. 107–116, 2024, doi: 10.1109/CIC62241.2024.00023.
- [3] G. Lang and T. Gulpinar, "AI-Powered Learning Support: A Study of Retrieval-Augmented Generation (RAG) Chatbot Effectiveness in an Online Course," *Inf. Syst. Educ. J.*, vol. 23, no. 2, pp. 4–13, 2025, doi: 10.62273/zklk5988.
- [4] O. Nacar and A. Koubaa, "Enhancing Semantic Similarity Understanding in Arabic NLP with Nested Embedding Learning," *Stud. Comput. Intell.*, vol. 1214, pp. 179–216, 2025, doi: 10.1007/978-3-031-90573-5_6.
- [5] S. Syukran, "Strategi Pembelajaran Kitab Kuning dalam Pendidikan Agama Islam di Era Digital: Studi Kasus di Pondok Pesantren Darul Hijrah Al-Aziziyah," *Edu Soc. J. Pendidikan, Ilmu Sos. Dan Pengabd. Kpd. Masy.*, vol. 5, no. 2, pp. 2095–2108, 2025, doi: 10.56832/edu.v5i2.1688.

- [6] T. Q. Ramadhani, N. Q. Nada, and N. D. S, “Penerapan Metode Retrieval-Augmented Generation (RAG) Pada Chatbot E-Commerce Berbasis Gemini Ai,” *J. Ilm. Ilk. - Ilmu Komput. Inform.*, vol. 8, no. 2, pp. 301–313, 2025, doi: 10.47324/ilkoinfo.v8i2.384.
- [7] A. Mishra and N. Brahmanapally, “A Comparative Performance Analysis of Locally Deployed Large Language Models Through a Retrieval-Augmented Generation Educational Assistant Application for Textual Data Extraction,” *AI*, vol. 6, no. 6, pp. 0–26, 2025, doi: 10.3390/ai6060119.
- [8] M. Renze and E. Guven, “The Effect of Sampling Temperature on Problem Solving in Large Language Models,” *EMNLP 2024 - 2024 Conf. Empir. Methods Nat. Lang. Process. Find. EMNLP 2024*, pp. 7346–7356, 2024, doi: 10.18653/v1/2024.findings-emnlp.432.
- [9] L. Huang *et al.*, “A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions,” *ACM Trans. Inf. Syst.*, vol. 43, no. 2, pp. 1–58, 2025, doi: 10.1145/3703155.
- [10] W. Zhang and J. Zhang, “Hallucination Mitigation for Retrieval-Augmented Large Language Models: A Review,” *Mathematics*, vol. 13, no. 5, 2025, doi: 10.3390/math13050856.