

Perancangan Chatbot Fiqih Berbasis *Retrieval-Augmented Generation* Kitab Safinatun Najah

Muhammad Ichsan¹, Alvino Graha Nusantara²

^{1,2}Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Nusantara PGRI Kediri

E-mail: *¹muhammadichsan11tkj@gmail.com, ²vinoalvino40@gmail.com

Abstrak – Pembelajaran fiqih di pesantren menghadapi tantangan, seperti kesulitan santri memahami teks Arab klasik dan ketergantungan pada ustadz akibat keterbatasan waktu. Penelitian ini mengembangkan *chatbot* fiqih berbasis *Retrieval-Augmented Generation* (RAG) dengan *korpus* tertutup kitab *Safinatun Najah* untuk santri kelas 1 *Asasiyah* di Pondok Pesantren Al-Ma'ruf Kediri. Pemilihan topik ini penting untuk meningkatkan kemandirian santri dalam mencari dalil fiqih secara akurat melalui teknologi.

Metode penelitian menggunakan *Research and Development* (R&D) dengan model ADDIE, meliputi: analisis kebutuhan melalui observasi dan wawancara, perancangan arsitektur RAG dengan *embedding multilingual-e5-large*, implementasi menggunakan *Python 3.11* dan *LangChain* serta antarmuka *web* berbasis *Streamlit*, serta evaluasi melalui *black-box testing* dan *User Acceptance Test* (UAT). Dataset disusun dari kitab *Safinatun Najah* menjadi sekitar 900 *chunk* teks.

Pengujian awal menunjukkan sistem mampu memberikan jawaban relevan dengan respons cepat dan minim *hallucination*. Penelitian ini penting karena menyediakan media pembelajaran fiqih yang terkontrol, sehingga mendukung integrasi teknologi dalam pendidikan agama Islam di pesantren.

Kata Kunci — *chatbot fiqih, pesantren, retrieval-augmented generation, safinatun najah.*

1. PENDAHULUAN

Perkembangan teknologi kecerdasan buatan telah membuka peluang baru dalam berbagai bidang, termasuk pendidikan, khususnya pendidikan agama Islam. Salah satu pendekatan terkini dalam pengembangan *Large Language Model* (LLM) adalah *Retrieval-Augmented Generation* (RAG), yang mengintegrasikan proses pencarian dokumen relevan sebelum tahap generasi jawaban untuk meningkatkan kinerja pada tugas-tugas berbasis pengetahuan (*knowledge-intensive tasks*) Islam [1], [2]. Pendekatan RAG terbukti efektif dalam mengurangi fenomena *hallucination* pada LLM dengan memastikan jawaban yang dihasilkan berbasis sumber pengetahuan yang tervalidasi [1],[2]. Selain itu, perkembangan *multilingual embedding* seperti *Multilingual E5* memungkinkan pemrosesan teks lintas bahasa secara lebih akurat, termasuk bahasa Arab yang tergolong sebagai bahasa dengan sumber daya terbatas (*low-resource language*) [3]. Penelitian sebelumnya juga menunjukkan bahwa integrasi RAG dengan pendekatan yang mampu meningkatkan performa sistem tanya jawab berbasis teks Arab [4], [5].

Penerapan RAG dalam pengembangan *chatbot* Islami telah menunjukkan hasil yang positif, khususnya dalam konteks konsultasi syariah dan pembelajaran berbasis dialog [6]. Studi lain juga menegaskan bahwa *chatbot* berbasis RAG dapat meningkatkan efektivitas pembelajaran dengan menyediakan jawaban yang kontekstual dan relevan terhadap kebutuhan pengguna [7], [8]. Namun demikian, pemanfaatan teknologi AI dalam pendidikan agama Islam memerlukan perhatian serius terhadap aspek etika dan kesesuaian dengan prinsip-prinsip hukum Islam, mengingat potensi pengaruhnya terhadap pemahaman keagamaan peserta didik [9].

Retrieval-Augmented Generation (RAG) menggabungkan proses *retrieval* dari sumber pengetahuan eksternal dengan kemampuan *generation* LLM untuk meningkatkan akurasi jawaban, serta menjadi pendekatan yang semakin umum dalam berbagai domain aplikasi termasuk layanan kesehatan dan pendidikan [10].

Di lingkungan pesantren, pembelajaran fiqih masih mengandalkan metode tradisional dengan fokus pada kitab kuning seperti *Safinatun Najah*. Santri kelas 1 *Asasiyah* sering mengalami kesulitan memahami teks Arab klasik dan mencari dalil secara mandiri. Penelitian ini mengembangkan *chatbot* fiqih berbasis RAG dengan korpus tertutup kitab *Safinatun Najah* untuk mendukung kemandirian belajar santri di Pondok Pesantren Al-Ma'ruf Kediri.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan *Research and Development* (R&D) dengan model ADDIE (*Analysis, Design, Development, Implementation, Evaluation*).

1. Analysis Analisis kebutuhan dilakukan melalui observasi 2 sesi kelas dan wawancara 20 santri, mengidentifikasi kesulitan memahami teks Arab.
2. Design Perancangan arsitektur RAG dengan dataset CSV kitab *Safinatun Najah*, *embedding multilingual-e5-large*, dan antarmuka *Streamlit*.
3. Development Implementasi dilakukan menggunakan *Python 3.11*, *LangChain*, *ChromaDB*, dan *API Gemini 2.5 Flash*.
4. Implementation *Deployment* lokal dan melalui *Streamlit Community Cloud*.
5. Evaluation *Black-box testing* pada pertanyaan representatif dan *User Acceptance Test (UAT)* dengan kuesioner Likert pada santri.

2.1 Arsitektur Sistem

Arsitektur sistem *chatbot* fiqih dirancang berbasis *Retrieval-Augmented Generation (RAG)* dengan alur sebagai berikut:

1. Pengguna mengajukan pertanyaan dalam bahasa Indonesia.
2. Pertanyaan di-*embedding* menjadi *vektor* menggunakan model *multilingual-e5-large*.
3. *Vektor* pertanyaan dibandingkan dengan *vektor* dokumen di *ChromaDB* menggunakan *cosine similarity*:

$$\cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} \quad (1)$$

Dimana *A* adalah *vektor* *embedding* pertanyaan dan *B* adalah *vektor* dokumen. Dokumen relevan (*top-k* = 3) digabungkan sebagai *context* dan dikirim ke *LLM Google Gemini 2.5 Flash* dengan *temperature* = 0 untuk jawaban deterministik. *LLM* menghasilkan jawaban terstruktur disertai teks Arab dan referensi bab.

2.2 Proses Retrieval dan Generation

Proses *retrieval* dan *generation* merupakan inti dari algoritma *Retrieval-Augmented Generation (RAG)* pada sistem ini. *Retrieval* dilakukan untuk mencari dokumen relevan dari korpus kitab *Safinatun Najah*, sementara *generation* menghasilkan jawaban terstruktur berdasarkan dokumen tersebut.

a. Retrieval Dokumen

Vektor embedding pertanyaan dibandingkan dengan *vektor embedding* dokumen di *ChromaDB* menggunakan metrik *cosine similarity*. Rumus *cosine similarity* yang diterapkan adalah sebagai berikut:

$$\cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (2)$$

Dimana *A* adalah *vektor* *embedding* pertanyaan dan *B* adalah *vektor embedding* dokumen. Sistem mengambil *top-k* = 3 dokumen dengan nilai kemiripan tertinggi yang melebihi ambang batas (*similarity threshold* = 0,8). Strategi *top-k retrieval* dengan *threshold* tertentu terbukti efektif dalam mengurangi konteks yang tidak relevan dan menekan potensi *hallucination* pada sistem RAG [11]. Mekanisme ini memastikan hanya dokumen yang benar-benar relevan yang digunakan sebagai *context*.

b. Generation Jawaban

Dokumen relevan yang diperoleh dari tahap *retrieval* digabungkan dengan pertanyaan pengguna dalam bentuk *augmented prompt*, kemudian dikirim ke *Large Language Model Google Gemini 2.5 Flash* untuk proses *generation*. Parameter *temperature* ditetapkan sebesar 0 guna menghasilkan keluaran yang bersifat deterministik (*deterministic output*) dan meminimalkan variasi kreatif (*creative variance*). Penelitian sebelumnya menunjukkan bahwa pengaturan *temperature* rendah berkontribusi signifikan terhadap konsistensi dan stabilitas jawaban *LLM* pada tugas yang menuntut ketepatan informasi.

2.3 Prompt Engineering dan Pembatasan Halusinasi

Prompt engineering merupakan tahap krusial dalam proses *generation* untuk memastikan *Large Language Model (LLM)* menghasilkan jawaban yang secara ketat mengikuti korpus tertutup kitab *Safinatun Najah*. Prompt dirancang menggunakan instruksi eksplisit (*explicit instruction prompting*) agar *LLM* hanya memanfaatkan informasi dari dokumen relevan yang di-*retrieve*, tanpa menambahkan pengetahuan eksternal (*external knowledge*), sehingga mampu meminimalkan terjadinya *hallucination* pada sistem berbasis *Retrieval-Augmented Generation (RAG)* [12], [11].

Selain desain prompt, pengaturan parameter generasi juga berperan penting dalam mengendalikan keluaran *LLM*. Pada penelitian ini, *LLM Google Gemini 2.5 Flash* dikonfigurasi dengan nilai *temperature* = 0 untuk menghasilkan respons yang bersifat deterministik (*deterministic output*) dan meminimalkan variasi

kreatif (*creative variance*). Penelitian sebelumnya menunjukkan bahwa nilai *temperature* yang rendah secara signifikan meningkatkan konsistensi dan stabilitas jawaban LLM, khususnya pada tugas yang menuntut ketepatan informasi dan penalaran faktual [13].

Kombinasi antara *prompt engineering* yang terkontrol, pengaturan parameter generasi yang konservatif, serta penggunaan korpus tertutup (*closed-domain corpus*) memungkinkan sistem chatbot menghasilkan jawaban yang akurat, terverifikasi, dan dapat dipertanggungjawabkan. Pendekatan ini sejalan dengan prinsip kehati-hatian (*prudential principle*) dalam penyampaian hukum fiqih, di mana keakuratan dalil dan kesesuaian sumber rujukan menjadi aspek yang sangat fundamental [14].

2.4 Antarmuka Pengguna

Implementasi sistem dilakukan menggunakan bahasa pemrograman *Python* versi 3.11 sebagai bahasa utama. *Framework* LangChain digunakan untuk *orchestrasi pipeline Retrieval-Augmented Generation (RAG)*, sementara *ChromaDB* berfungsi sebagai *vector database* untuk menyimpan dan mencari *embedding* dokumen. Akses ke *Large Language Model Google Gemini 2.5 Flash* dilakukan melalui *API* resmi untuk proses *generation*.

Akses ke *Large Language Model Google Gemini 2.5 Flash* dilakukan melalui *API* resmi untuk proses *generation*. Arsitektur *RAG* yang memanfaatkan *framework* orkestrasi dan *vector database* telah diteliti secara luas dalam literatur terkait, di mana penggunaan *vector database* seperti *Chroma* mendukung pencarian *semantic similarity* yang cepat dan akurat untuk konteks domain-spesifik sebelum pemanggilan *LLM* secara langsung [15].

Antarmuka pengguna dikembangkan berbasis *web* menggunakan *Streamlit*, yang memungkinkan *deployment* cepat dan akses responsif melalui *browser* desktop maupun *mobile*. Desain antarmuka mengedepankan kesederhanaan dengan elemen utama berupa:

1. Halaman *login* sederhana untuk *otentikasi* pengguna.
2. Kolom *input* pertanyaan dengan placeholder “Tanya fiqih mazhab Syafi’i...”.
3. Tampilan riwayat percakapan dalam format *chat bubble*.
4. Indikator *loading* selama proses *retrieval* dan *generation*.
5. *Output* jawaban terstruktur dengan *expander* untuk teks Arab asli dan referensi bab.

Implementasi ini memastikan sistem mudah digunakan oleh santri dengan literasi digital terbatas, sekaligus mendukung akses kapan saja melalui perangkat *mobile* di lingkungan pesantren.

2.5 Spesifikasi Teknis Chatbot Fiqih

Tabel 1. Komponen Chatbot	
Komponen Sistem	Teknologi / Konfigurasi
Bahasa Pemrograman	Python 3.11
Pendekatan	Retrieval-Augmented Generation (RAG)
Model Embedding	intfloat/multilingual-e5-large
Vector Database	ChromaDB
Similarity Metric	Cosine Similarity
Nilai Top-k Retrieval	3 dokumen
Similarity Threshold	0,8
Large Language Model	Google Gemini 2.5 Flash
Nilai Temperature	0
Framework Orkestrasi	LangChain
Antarmuka Pengguna	Streamlit
Sumber Data	Kitab Safinatun Najah (korpus tertutup)

Tabel 1. Komponen *Chatbot* menyajikan spesifikasi teknis sistem *chatbot* fiqih yang dikembangkan dalam penelitian ini. Konfigurasi sistem dirancang untuk mendukung proses *retrieval* dan *generation* secara terkontrol dengan memanfaatkan korpus tertutup kitab Safinatun Najah. Pemilihan model *embedding* multibahasa dan *Large Language Model* deterministik bertujuan menjaga konsistensi jawaban serta meminimalkan risiko *hallucination* dalam konteks pembelajaran fiqih.

3. HASIL DAN PEMBAHASAN

3.1 Tampilan Aplikasi *Chatbot* Fiqih



Gambar 1. Tampilan Aplikasi *Chatbot*

Gambar 1 menunjukkan tampilan antarmuka aplikasi *chatbot* fiqih yang dikembangkan pada penelitian ini. Aplikasi bertajuk *Chatbot* Fiqih Santri menggunakan korpus tertutup kitab *Safinatun Najah* dan dirancang sebagai media pendukung pembelajaran fiqih bagi santri Pesantren Al-Ma'ruf.

Pada tampilan utama, pengguna dapat mengajukan pertanyaan fiqih sesuai mazhab *Syafi'i* melalui kolom input yang tersedia. Sistem kemudian memproses pertanyaan tersebut menggunakan metode *retrieval-augmented generation (RAG)* untuk menghasilkan jawaban yang relevan berdasarkan dokumen kitab. Antarmuka dibuat sederhana dan fokus pada kemudahan interaksi agar dapat digunakan secara mandiri oleh santri.

Penyajian aplikasi secara *online* memungkinkan sistem diakses secara fleksibel tanpa instalasi tambahan, sehingga mendukung pemanfaatan *chatbot* sebagai sarana pembelajaran fiqih berbasis teknologi di lingkungan pesantren.

3.2 Hasil Uji Fungsional Sistem *Chatbot* Fiqih

Pengujian sistem *chatbot* fiqih dilakukan secara fungsional dengan memberikan pertanyaan terkait materi fiqih ibadah, khususnya mengenai waktu yang disunahkan untuk mengangkat kedua tangan dalam shalat berdasarkan kitab *Safinatun Najah*. Pengujian ini bertujuan untuk memastikan bahwa sistem mampu menampilkan jawaban fiqih yang sesuai dengan sumber rujukan serta menyajikan teks Arab asli sebagai penguat validitas jawaban.

Salah satu *query* yang diuji adalah pertanyaan mengenai waktu yang disunahkan untuk mengangkat kedua tangan dalam shalat. Berdasarkan hasil pengujian, sistem mampu menghasilkan jawaban yang menjelaskan bahwa terdapat empat waktu disunahkan untuk mengangkat kedua tangan, yaitu saat *takbiratul ihram*, ketika hendak *ruku'*, saat *i'tidal* (bangkit dari *ruku'*), dan ketika berdiri dari *tasyahud* awal. Penjelasan tersebut sesuai dengan keterangan yang terdapat dalam kitab *Safinatun Najah*.

Selain memberikan penjelasan dalam bahasa Indonesia, sistem juga menampilkan teks Arab asli dari kitab *Safinatun Najah* yang menjelaskan keempat waktu tersebut. Penyajian ini bertujuan untuk menjaga kesesuaian jawaban dengan sumber fiqih yang dijadikan rujukan utama.

3.3 Hasil Pengujian Sistem

Hasil pengujian sistem *chatbot* fiqih diperoleh melalui pengujian fungsional dengan memberikan pertanyaan fiqih yang bersumber dari materi kitab *Safinatun Najah*. Pengujian ini bertujuan untuk menilai kemampuan sistem dalam menghasilkan jawaban yang relevan, terstruktur, serta sesuai dengan sumber rujukan kitab yang digunakan sebagai korpus tertutup.

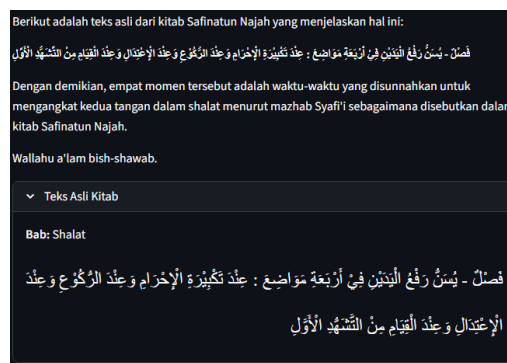
Salah satu pertanyaan yang diuji adalah terkait waktu yang disunahkan untuk mengangkat kedua tangan dalam shalat. Berdasarkan hasil pengujian, sistem mampu menyajikan jawaban yang menjelaskan empat waktu disunahkan mengangkat kedua tangan, yaitu saat *takbiratul ihram*, ketika hendak *ruku'*, saat *i'tidal*, serta ketika berdiri dari *tasyahud* awal. Contoh hasil keluaran sistem berdasarkan pertanyaan pengguna disajikan pada Tabel 2.

Tabel 2. Contoh Hasil Jawaban Sistem *Chatbot* Fiqih

Pertanyaan	Jawaban <i>chatbot</i>	Bab
Jelaskan Waktu yang disunahkan untuk mengangkat kedua tangan didalam shalat?	Sistem menjelaskan empat waktu disunahkan mengangkat kedua tangan, yaitu saat <i>takbiratul ihram</i> , ketika hendak <i>ruku'</i> , saat <i>i'tidal</i> , dan ketika berdiri dari <i>tasyahud awal</i> , disertai teks Arab asli dari kitab. فَصَلِّ - يُسَبِّحُ رَفَعُ الْيَدَيْنِ فِي أَرْبَعَةِ مَوَاضِعَ : عِنْدَ تَكْبِيرَةِ الْإِحْرَامِ وَعِنْدَ الرُّكُوعِ وَعِنْدَ الْإِعْتِدَالِ وَعِنْدَ الْقِيَامِ مِنَ التَّشَهُُّدِ الْأَوَّلِ	Kitab <i>Safinatun Najah</i> , Bab Shalat

Selain disajikan dalam bentuk tabel, hasil pengujian sistem juga dapat diamati melalui tampilan keluaran *chatbot* secara langsung. Gambar 2. menunjukkan contoh hasil keluaran sistem ketika pengguna mengajukan pertanyaan terkait waktu yang disunahkan untuk mengangkat kedua tangan dalam shalat.

Pada tampilan tersebut, sistem menampilkan jawaban dalam bahasa Indonesia yang disertai kutipan teks Arab asli dari kitab *Safinatun Najah* serta keterangan bab kitab sebagai sumber rujukan. Penyajian ini menunjukkan bahwa sistem tidak hanya menghasilkan jawaban deskriptif, tetapi juga mampu mempertahankan keterkaitan langsung dengan sumber fiqih yang digunakan, sehingga meningkatkan kepercayaan pengguna terhadap keabsahan jawaban yang diberikan.



Gambar 2. Hasil Sisitem *chatbot*

3.4 Pembahasan

Berdasarkan hasil uji fungsional dan contoh keluaran sistem yang telah ditampilkan, penerapan metode *retrieval-augmented generation (RAG)* pada sistem *chatbot* fiqih terbukti mampu menyajikan jawaban yang sesuai dengan konteks kitab *Safinatun Najah*. Sistem tidak hanya memberikan penjelasan deskriptif, tetapi juga menyertakan dalil tekstual berupa kutipan Arab asli, sehingga dapat meningkatkan kepercayaan pengguna terhadap keabsahan jawaban yang diberikan.

Keunggulan sistem terlihat pada kemampuannya mengintegrasikan proses *retrieval* dokumen dari korpus tertutup dengan proses *generation* jawaban secara kontekstual. Namun demikian, pengujian pada penelitian ini masih bersifat terbatas dan hanya mencakup materi tertentu. Oleh karena itu, hasil yang diperoleh belum dapat menggambarkan performa sistem secara menyeluruh untuk seluruh materi fiqih.

3.5 Implikasi terhadap Pembelajaran Fiqih di Pesantren

Penerapan *chatbot* fiqih berbasis *natural language processing (NLP)* dan *retrieval-augmented generation (RAG)* memiliki implikasi penting terhadap proses pembelajaran fiqih di lingkungan pesantren. Sistem ini memungkinkan santri untuk mengajukan pertanyaan fiqih secara mandiri dan memperoleh jawaban yang merujuk langsung pada sumber kitab, dalam hal ini *Safinatun Najah*.

Penggunaan korpus tertutup memastikan bahwa jawaban yang diberikan sistem tetap berada dalam koridor materi yang diajarkan di pesantren. Dengan demikian, *chatbot* berperan sebagai media pendukung pembelajaran yang dapat membantu santri dalam memahami dan mengulang materi fiqih dasar di luar jam belajar formal tanpa menggantikan peran ustadz sebagai pendidik utama.

3.6 Peran *Chatbot* sebagai Asisten Pembelajaran

Dalam konteks pendidikan pesantren, *chatbot* fiqih berfungsi sebagai *learning assistant* yang mendukung proses belajar santri secara fleksibel [14]. *Chatbot* dapat digunakan kapan saja untuk membantu santri memahami istilah fiqih, rukun ibadah, maupun penjelasan singkat terkait bab tertentu dalam kitab *Safinatun Najah*.

Keberadaan *chatbot* ini tidak dimaksudkan untuk menggantikan peran ustadz, melainkan sebagai sarana pendamping yang membantu mengurangi ketergantungan santri pada ustadz untuk pertanyaan-pertanyaan dasar. Dengan demikian, waktu pembelajaran tatap muka dapat dimanfaatkan secara lebih optimal untuk diskusi yang bersifat pendalaman materi.

3.7 Keterbatasan Sistem

Meskipun memiliki potensi yang signifikan, sistem *chatbot* fiqih yang dikembangkan dalam penelitian ini memiliki sejumlah keterbatasan. Pertama, cakupan materi *chatbot* terbatas pada kitab *Safinatun Najah*, sehingga pertanyaan di luar ruang lingkup kitab tersebut tidak dapat dijawab secara komprehensif. Keterbatasan ini merupakan konsekuensi dari penerapan *korpus* tertutup yang sengaja dipilih untuk menjaga keakuratan dan validitas jawaban.

Kedua, sistem belum dirancang untuk menangani perbedaan pendapat (*ikhtilaf*) antar mazhab fiqih, karena fokus penelitian ini adalah pada materi fiqih mazhab Syafi'i sebagaimana yang diajarkan di pesantren. Selain itu, kualitas jawaban sangat bergantung pada formulasi pertanyaan pengguna dan kelengkapan dokumen yang berhasil di-*retrieve* dari *korpus*.

3.8 Peluang Pengembangan Lanjutan

Pengembangan lebih lanjut dari sistem *chatbot* fiqih ini masih terbuka luas. Salah satu peluang pengembangan adalah penambahan *korpus* kitab fiqih lain yang relevan dengan kurikulum pesantren, seperti *Fathul Qarib* atau *Taqrib*, dengan tetap mempertahankan pendekatan *korpus* tertutup. Selain itu, sistem dapat dikembangkan untuk mendukung fitur pencarian berbasis bab atau tema guna memudahkan santri dalam menelusuri materi tertentu.

Dari sisi pedagogis, *chatbot* dapat diintegrasikan dengan sistem pembelajaran pesantren sebagai media pendamping resmi yang digunakan di bawah pengawasan ustadz. Dengan pengembangan yang berkelanjutan dan evaluasi yang sistematis, *chatbot* fiqih berbasis RAG berpotensi menjadi inovasi teknologi yang mendukung transformasi pembelajaran fiqih tanpa menghilangkan nilai-nilai tradisi pesantren.

4. KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan, dapat disimpulkan sebagai berikut:

1. Penelitian ini berhasil mengembangkan *chatbot* fiqih berbasis *retrieval-augmented generation* (RAG) dengan korpus tertutup kitab *Safinatun Najah* yang dapat digunakan sebagai media pendukung pembelajaran fiqih di lingkungan pesantren.
2. Sistem yang dikembangkan mampu menghasilkan jawaban fiqih yang terkontrol dan berbasis rujukan kitab, sehingga dapat meminimalkan potensi *hallucination* dalam pemanfaatan *large language model*.
3. Berdasarkan hasil uji fungsional terbatas, *chatbot* fiqih berpotensi membantu meningkatkan kemandirian santri dalam memahami materi fiqih dasar tanpa menggantikan peran ustadz sebagai pendidik utama.
4. Keterbatasan sistem terletak pada cakupan materi yang masih terbatas pada satu kitab serta belum mendukung perbedaan pandangan antar mazhab dalam pembahasan fiqih.

5. SARAN

Berdasarkan keterbatasan penelitian ini, beberapa saran untuk penelitian selanjutnya adalah sebagai berikut:

1. Penelitian lanjutan dapat menambahkan korpus kitab fiqih lain yang relevan dengan kurikulum pesantren guna memperluas cakupan materi yang dapat diakses oleh sistem.
2. Evaluasi kinerja sistem dapat dilakukan secara kuantitatif dengan melibatkan dataset uji yang lebih besar dan beragam agar diperoleh gambaran performa sistem yang lebih komprehensif.
3. Pengembangan fitur pencarian berbasis bab atau tema kitab dapat dipertimbangkan untuk meningkatkan kemudahan akses dan efektivitas pembelajaran fiqih bagi santri.

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] A. Gan *et al.*, “Retrieval Augmented Generation Evaluation in the Era of Large Language Models: A Comprehensive Survey,” vol. 0, no. 0, pp. 1–18, 2025, [Online]. Available: <http://arxiv.org/abs/2504.14891>
- [3] L. Wang, N. Yang, X. Huang, L. Yang, R. Majumder, and F. Wei, “Multilingual E5 Text Embeddings: A Technical Report,” 2024, [Online]. Available: <http://arxiv.org/abs/2402.05672>
- [4] 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.
- [5] 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.
- [6] W. Uriawan, A. O. Hamza, A. R. Nuralim, A. Purnama, A. J. Yunus, and A. A. S. Putri, “Implementing a Sharia Chatbot as a Consultation Medium for Questions About Islam,” 2025, [Online]. Available: <http://arxiv.org/abs/2512.16644>
- [7] G. Lang and T. Gurpinar, “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.
- [8] 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.
- [9] A. Rahman Ramadhan, “The AI Chatbot Phenomenon and Its Impact on Learning Fiqh (A Study of the Ethical Dimensions of Artificial Intelligence in Learning Islamic Jurisprudence),” *Sci. Educ.*, vol. 3, pp. 177–186, 2024.
- [10] L. M. Amugongo, P. Mascheroni, S. Brooks, S. Doering, and J. Seidel, “Retrieval augmented generation for large language models in healthcare: A systematic review,” *PLOS Digit. Heal.*, vol. 4, no. 6 JUNE, 2025, doi: 10.1371/journal.pdig.0000877.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] M. R. Febriansyah, G. Boy Hertantyo, and W. Wilonotomo, “Implementasi Chatbot Sebagai Virtual Assistant : Systematic Literature Review,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 4, pp. 6161–6168, 2025, doi: 10.36040/jati.v9i4.13876.
- [15] F. S. Tamenković, J. S. Tanojević, and M. M. Inović, “A SYSTEMATIC REVIEW OF VECTOR DATABASE USE IN RETRIEVAL-AUGMENTED GENERATION FOR LLM-BASED EDUCATIONAL PLATFORMS,” pp. 262–267, 2025.