

# Retrieval-Augmented Generation in a Web-Based Question Answering System for Fiqh Books

Ridho Ahadi<sup>1</sup>, Nazruddin Safaat Harahap<sup>2\*</sup>, Muhammad Fikry<sup>3</sup>, Fitra Kurnia<sup>4</sup>

<sup>1,2,3,4</sup> Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, 28293, Indonesia

## Informasi Artikel

Diterima : 23 Mei 2025  
Revisi : 31 Mei 2025  
Publikasi : 20 Juni 2025

## Kata Kunci:

Sistem Tanya Jawab  
Artificial Intelligence  
Natural Language Processing  
Retrieval-Augmented Generation  
BERTScore

## ABSTRAK

Mayoritas masyarakat Indonesia yang beragama Islam cenderung tidak mengetahui madzhab yang mereka anut, padahal pemahaman madzhab sangat penting untuk menjalankan ibadah secara benar dan sah. Salah satu madzhab yang berkembang di Indonesia, terutama melalui ulama Hadramaut, adalah madzhab Imam Asy-Syafi'i. Kurangnya akses terhadap literatur fikih otentik serta metode pembelajaran interaktif menyebabkan masyarakat kesulitan memahami fikih secara mendalam. Perkembangan teknologi kecerdasan buatan (AI), memberikan dampak signifikan dalam berbagai aspek, termasuk pendidikan dan keagamaan. Salah satu implementasi AI yang berkembang adalah chatbot, sistem interaktif berbasis percakapan yang mampu memahami dan merespons pertanyaan secara alami. Dalam konteks pembelajaran keislaman, khususnya fikih, penggunaan chatbot AI menghadirkan peluang baru untuk pembelajaran yang lebih personal dan interaktif. Fikih sebagai cabang ilmu hukum Islam menuntut pemahaman mendalam terhadap sumber klasik dan aplikasinya dalam kehidupan modern. Penelitian ini memiliki tujuan untuk mengembangkan sistem tanya jawab fikih berbasis AI dan Natural Language Processing menggunakan Large Language Model (LLM), framework LangChain, serta metode Retrieval-Augmented Generation (RAG). Sistem ini dirancang memberikan jawaban relevan berdasarkan konteks fikih. Evaluasi menggunakan metrik BERTScore menghasilkan precision 86,50%, recall 84,76%, dan F1-score 85,52%, hasil ini menunjukkan akurasi tinggi dalam menjawab pertanyaan fikih.

## ABSTRACT

The majority of Indonesian Muslims tend not to know the madzhab they adhere to, even though understanding madzhab is very important for carrying out worship correctly and legally. One of the madzhab that developed in Indonesia, especially through Hadramaut scholars, is the madzhab of Imam Asy-Syafi'i. The lack of access to authentic fiqh literature and interactive learning methods makes it difficult for people to understand fiqh in depth. The development of artificial intelligence (AI) technology has had a significant impact on various aspects, including education and religion. One of the developing AI implementations is chatbot, an interactive conversation-based system that is able to understand and respond to questions naturally. In the context of Islam, especially fiqh, the use of AI chatbots presents new opportunities for more personal and interactive learning. Fiqh as a branch of Islamic law requires a deep understanding of classical sources and their applications in modern life. This study aims to develop a fiqh question and answer system based on AI and Natural Language Processing using the Large Language Model (LLM), the LangChain framework, and the Retrieval-Augmented Generation (RAG) method. This system is designed to provide relevant answers based on the context of fiqh. Evaluation using the BERTScore metric produced a precision of 86.50%, a recall of 84.76%, and an F1-score of 85.52%, these results indicate high accuracy in answering fiqh questions.

**\*Penulis Koresponden**Email: [nazruddin.safaat@uin-suska.ac.id](mailto:nazruddin.safaat@uin-suska.ac.id)

Cara sitasi IEEE::

R. Ahadi, N. S. Harahap, M. Fikry dan F. Kurnia, "Retrieval-Augmented Generation in a Web-Based Question Answering System for Fiqh Books," *Journal of Artificial Intelligence and Software Engineering (J-AISE)*, vol. 5, no. 2, p. 626-635, Juni 2025. doi:10.30811/jaise.v5i2.7005

**1. PENDAHULUAN**

Mayoritas masyarakat Indonesia beragama Islam, tetapi mereka tidak mengetahui madzhab yang mereka anut. Bagi kebanyakan muslim, satu-satunya hal yang mereka ketahui tentang Islam adalah didasarkan pada Al-Qur'an dan hadis, dan tidak bermazhab. Secara historis, Islam masuk ke Indonesia dengan berbagai cara, dan salah satu cara Islam menyebar di Indonesia adalah melalui Muslim Hadramaut yang menganut madzhab Imam Syafi'i. Untuk memperdalam ilmu pengetahuan, seseorang dapat memperoleh pengetahuan melalui lembaga pendidikan nonformal dan formal. Pengetahuan tentang fikih bertujuan untuk meningkatkan ibadah seseorang, sehingga mereka menjadi lebih memahami dan lebih percaya pada ibadah mereka [1].

Pada masa sekarang ini, perkembangan teknologi semakin cepat dan pesat, hampir memasuki semua bidang dalam kehidupan [2]. Berkembangnya teknologi informasi dan komunikasi juga dikenal sebagai teknologi digital, adalah tanda dari revolusi industri 4.0. Perangkat digital baik hardware maupun software, dirancang untuk membuat manusia lebih mudah melakukan aktivitas dan tugas. Semua pekerjaan manusia menjadi lebih cepat dan mudah, sehingga manusia akhirnya lebih suka sesuatu yang instan atau cepat. Perangkat-perangkat ini menjadi lebih futuristik dan pintar berkat kemajuan teknologi ini, mereka bahkan dikembangkan untuk menyerupai kecerdasan manusia [3].

Kecerdasan Buatan atau Artificial Intelligence (AI) semakin mengintegrasikan kehidupan sehari-hari kita dengan penciptaan dan analisis perangkat lunak dan perangkat keras cerdas, yang disebut dengan agen cerdas. Agen cerdas dapat melakukan banyak hal, seperti pekerjaan buruh dan operasi canggih. Chatbots adalah salah satu contoh AI dan interaksi manusia dan komputer atau Human-Computer Interaction (HCI) yang paling umum [4]. Dalam beberapa tahun terakhir, perkembangan kecerdasan buatan (AI) telah membawa transformasi signifikan dalam berbagai aspek kehidupan, termasuk di bidang pendidikan. Salah satu inovasi yang menonjol adalah chatbot berbasis AI, yaitu sistem yang dirancang untuk berinteraksi dengan manusia melalui percakapan alami, dan kini menjadi elemen penting dalam proses pembelajaran modern. Dalam konteks pembelajaran keagamaan seperti fikih (hukum Islam), penggunaan chatbot AI membuka peluang baru yang potensial, namun sekaligus memunculkan persoalan-persoalan etis. Fikih sendiri merupakan disiplin ilmu dalam Islam yang membahas ketentuan hukum berdasarkan sumber ajaran Islam. Proses pembelajaran fikih menuntut pemahaman yang mendalam terhadap literatur klasik, interpretasi, dan penerapannya dalam konteks kontemporer. Kehadiran chatbot AI memungkinkan pengalaman belajar fikih yang lebih interaktif dan personal sesuai kebutuhan masing-masing individu [5].

Chatbot merupakan layanan obrolan yang digunakan oleh robot atau tokoh virtual untuk berinteraksi secara otomatis dengan pengguna. Chatbot dapat berfungsi sebagai pengganti manusia untuk menjawab pertanyaan pengguna [6]. Chatbot hanya bisa berbicara dalam beberapa bahasa, meskipun chatbot adalah sistem sederhana yang membantu manusia berbicara dengan mereka menggunakan "aturan matching pattern". Chatbot dapat digunakan dalam berbagai cara, mereka dapat dibuat secara mandiri atau menggunakan aplikasi chatting yang menawarkan layanan untuk meningkatkannya. Chatbot mampu memahami dan menjawab percakapan dengan menggunakan metode pengolahan bahasa alami, yang sering dikenal sebagai Natural Language Processing [7]. Chatbot menggunakan teknologi pemrosesan bahasa alami dan kecerdasan buatan untuk berinteraksi dan memberikan jawaban yang tepat pada pertanyaan. Chatbot menggunakan platform dari OpenAI yaitu ChatGPT. ChatGPT memungkinkan pengguna berinteraksi dengan robot berbasis teks. Platform ini menggunakan GPT-3 yang merupakan singkatan dari Generative Pre-trained Transformer, untuk memahami respons yang diinginkan manusia dan menghasilkan respons yang relevan. Hal ini memungkinkan pengguna menemukan jawaban atas pertanyaan mereka dengan cepat dan relevan [8].

LangChain merupakan sebuah framework yang dapat berfungsi dengan mudah dengan berbagai sumber data dan aplikasi. Struktur modular LangChain memungkinkan pengembangan aplikasi AI yang disesuaikan dengan cepat dan dapat disesuaikan untuk memenuhi berbagai kebutuhan. Dalam penelitian ini, Large Language Models (LLM) digunakan untuk menjawab pertanyaan tentang dokumen tertentu. LLM dapat digunakan untuk membangun aplikasi yang menjawab pertanyaan tentang isi dokumen, sehingga pengguna dapat lebih memahami isi dokumen [9].

Large Language Models (LLM) adalah jenis kecerdasan buatan canggih yang memproses dan menghasilkan teks dengan menggunakan pembelajaran mendalam atau Deep Learning, terutama arsitektur transformator [10]. Llama merupakan model bahasa besar berbasis transformers yang dikembangkan oleh Meta untuk berbagai tugas NLP. Model LLaMA lebih kecil dan dilatih menggunakan dataset yang tersedia untuk umum, tetapi mampu bersaing dengan model-model modern seperti GPT-3, Chinchilla, dan PaLM [11]. Retrieval-Augmented Generation (RAG) merupakan metode untuk meningkatkan kinerja dari Large Language Model (LLM) pada tugas-tugas dibidang tertentu, khususnya di Indonesia, dengan menyediakan informasi eksternal pada model [12].

Penelitian “Aplikasi Tanya Jawab Tentang Fiqih Bersuci Berbasis Web” bertujuan untuk meningkatkan akses terhadap informasi mengenai fiqih bersuci (yurisprudensi Islam tentang kebersihan) melalui pengembangan Question Answering System (QAS) berbasis web. Hasil penelitian menunjukkan bahwa pengujian aplikasi mendapatkan pemahaman dan pemrosesan bahasa manusia yang cukup baik. Dengan BERTScore, akurasi jawaban sebesar 69%, recall sebesar 56%, dan F1-score sebesar 83% [13]. Penelitian Penerapan LangChain Retriever Dengan Model Chat Openai Dalam Pengembangan Sistem Chatbot Hadis Berbasis Telegram tujuan penelitian ini adalah untuk membuat sistem kueri yang menggunakan Large Language Model (LLM) dan kerangka kerja LangChain. Tujuannya adalah untuk meningkatkan efektivitas dan efisiensi pengambilan informasi dalam berbagai aplikasi, terutama dalam bidang teknologi dan teknik. Framework Scoring Evaluator, yang menawarkan cara terstruktur untuk menilai relevansi dan keakuratan tanggapan chatbot, digunakan untuk mengevaluasi implementasi LangChain. Hasilnya menunjukkan bahwa LangChain sangat efektif dalam aplikasi ini, dengan skor rata-rata 0,97 dan pengujian untuk mengevaluasi kualitas jawaban dari lima ahli hadis dengan persentase akurasi sebesar 90.48 %. Hasil pengujian Scoring Evaluator menunjukkan bahwa jawaban rata-rata benar-benar akurat dan sejalan dengan referensi hadis [14].

Penelitian Enhancing Textbook Question Answering Task with Large Language Models and Retrieval Augmented Generation. Penelitian ini menggunakan Large Language Model (LLM) tingkat lanjut, terutama Llama-2, yang telah menunjukkan potensi besar untuk meningkatkan akurasi jawaban pada tugas Textbook Question Answering. Arsitektur yang diusulkan menunjukkan peningkatan kinerja yang signifikan. Ini menunjukkan peningkatan akurasi 4,12% pada set validasi dan 9,84% pada set tes untuk pertanyaan pilihan ganda non-diagram [15]. Penelitian dengan judul Aplikasi Question Answer Sebagai Media Pembelajaran Interaktif Untuk Mata Pelajaran Akuntansi. Pada penelitian ini bertujuan untuk membuat aplikasi Question Answer interaktif untuk membantu siswa belajar akuntansi. Aplikasi ini dirancang untuk meningkatkan partisipasi dan pemahaman siswa dalam pelajaran. Studi menunjukkan bahwa peserta secara signifikan lebih memahami konsep akuntansi ketika menggunakan aplikasi. Skor pasca-tes rata-rata 82,5, dibandingkan dengan skor pra-tes rata-rata 75,71, menunjukkan peningkatan yang signifikan dalam pengetahuan [16].

Penelitian mengenai penerapan Retrieval-Augmented Generation (RAG) menggunakan LangChain dalam pengembangan sistem tanya jawab hadis berbasis web bertujuan untuk merancang sistem yang dapat menjawab pertanyaan seputar hadis melalui platform web. Sistem ini memanfaatkan framework LangChain yang terintegrasi dengan model bahasa besar GPT-4-1106-preview dari OpenAI. Tujuan utamanya adalah memudahkan pengguna dalam memperoleh jawaban yang relevan berdasarkan sembilan kitab hadis. Ruang lingkup penelitian dibatasi hanya pada pertanyaan yang berkaitan dengan hadis. Hasil penelitian menunjukkan bahwa dengan menyediakan sumber hadis yang sesuai, model mampu menghasilkan jawaban yang akurat sesuai konteks dan data yang diberikan. Evaluasi sistem dilakukan dengan dua pendekatan, yaitu penilaian kualitas jawaban secara manual dan perhitungan skor BERTScore terhadap sepuluh pertanyaan hadis. Hasil evaluasi menunjukkan BERTScore dengan rata-rata nilai f1 sebesar 0,7962, menandakan bahwa jawaban sistem cukup sebanding dengan referensi. Sementara itu, penilaian kualitas jawaban oleh responden menunjukkan tingkat akurasi sebesar 89,4%, di mana mayoritas menyatakan “Sangat Setuju” terhadap kualitas jawaban yang diberikan [17].

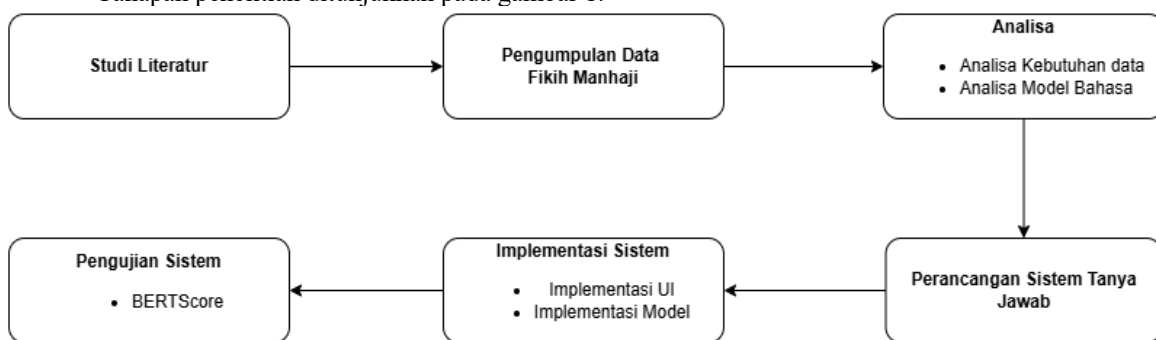
Penelitian Implementation Retrieval-Augmented Generation and Vector Databases for Chatbots in the Context of Public Services Agencies. Fokus penelitian ini adalah untuk menentukan seberapa baik berbagai Large Language Model berfungsi dalam menjawab pertanyaan tentang peraturan layanan publik. Untuk mengukur efektivitas respons yang dihasilkan oleh model, terutama digunakan metrik kesamaan kosinus. Semua metrik yang digunakan untuk menilai kinerja dalam penelitian ini adalah skor cosine similarity. Namun, penelitian masa depan perlu memasukkan metrik tambahan seperti akurasi dan presisi. Model gpt-4 OpenAI menunjukkan kinerja terbaik di antara model yang diuji, dengan skor cosine similarity rata-rata 0,404. Model

ini terkenal karena waktu responsnya yang cepat dan jawabannya membutuhkan kurang dari lima detik, kemudian dipilih untuk diterapkan [18].

Berdasarkan pemaparan diatas, penelitian ini bertujuan untuk mengembangkan sistem tanya jawab kitab Fikih Manhaji Fikih Imam Asy-Syafi'i dengan menggunakan metode Retrieval-Augmented Generation (RAG) untuk menghasilkan jawaban yang relevan terhadap permasalahan fikih sesuai kebutuhan masyarakat. Meskipun RAG telah banyak digunakan dalam berbagai bidang, tetapi masih sangat terbatas untuk penelitian yang secara spesifik mengimplementasikan metode ini dengan model Large Language Model (LLM) lokal berbasis LangChain untuk konteks pembelajaran fikih bermadzhab secara interaktif dalam bahasa Indonesia. Oleh karena itu, penelitian ini difokuskan untuk menjawab permasalahan bagaimana merancang sistem tanya jawab fikih manhaji menggunakan LLM dan metode RAG secara akurat dan relevan. Kontribusi utama dari penelitian ini adalah implementasi sistem tanya jawab berbasis web yang menjalankan model LLM lokal Llama 3.2:3B, serta evaluasi performa sistem menggunakan metrik BERTScore untuk mengukur keakuratan jawaban berdasarkan referensi dari kitab Fikih Manhaji Fikih Lengkap Imam Asy-Syafi'i.

## 2. METODE

Tahapan penelitian ditunjukkan pada gambar 1.



Gambar 1. Diagram Metode Penelitian

### 2.1. Studi Literatur

Studi literatur dilakukan peneliti dalam pengumpulan data dengan melakukan pencarian, pembelajaran serta pemahaman dari berbagai sumber literatur seperti jurnal, paper, dan website yang terkait dengan judul peneliti gunakan, seperti pada literatur tentang Sistem tanya jawab, Natural Language Processing, LangChain, Retrieval-Augmented Generation (RAG) dan LLM serta objek data penelitian yaitu kitab fikih manhaji.

### 2.2. Pengumpulan Data

Peneliti mengumpulkan data kitab fikih manhaji. Kitab fikih manhaji berbicara tentang fikih madzhab Imam Syafi'i, termasuk metode dan hukumnya. Kitab fikih dalam bahasa Arab ini tidak memiliki harakat. Penggunaan kitab Al-Fiqh Al-Manhaji 'Ala Madzhabil Imam Asy-Syafi'i memiliki beberapa keuntungan. Pertama adalah kitab ini bermadzhab Imam Syafi'i. Kedua, itu mencakup poin masalah lengkap dengan dalil dari Al-Quran, Hadist dan pendapat para sahabat. Ketiga, kitab ini sangat sistematis, yang berarti bahwa kitab fikih ini ditulis dengan terstruktur. Keempat, terletak pada sisi layout-nya, yang berarti rapi dan baik. Kitab fikih manhaji merupakan terjemahan dari kitab aslinya yang berjudul Al-Fiqh Al-Manhaji 'Ala Mazhabil Imam Asy-Syafi'i karya Dr. Mustofa Al-Khin, Dr. Mustofa Al-Bugho dan Ali Asy-Syarbaji [19]. Data fikih manhaji ini masih berbentuk fisik yang kemudian di jadikan sebagai format (.txt) untuk disimpan ke vectorstore database. Adapun data fikih manhaji fikih lengkap imam Asy-Syafi'i sebanyak 5 jilid ditunjukkan pada tabel 1.

Tabel 1. Daftar kitab Fikih Manhaji Fikih Lengkap Imam Asy-Syafi'i

No	Judul	Total Halaman
1	Jilid 1: Taharah	136
2	Jilid 2: Shalat	189
3	Jilid 3: Aturan Jenazah dan Puasa	166
4	Jilid 4: Zakat dan Wakaf	144
5	Jilid 5: Haji dan Umrah	147

Berdasarkan tabel 1, terdapat 5 Jilid buku fikih manhaji dengan berbagai pembahasan terkait mazhab imam Ays-Syafi'i.

## 2.3. Analisa

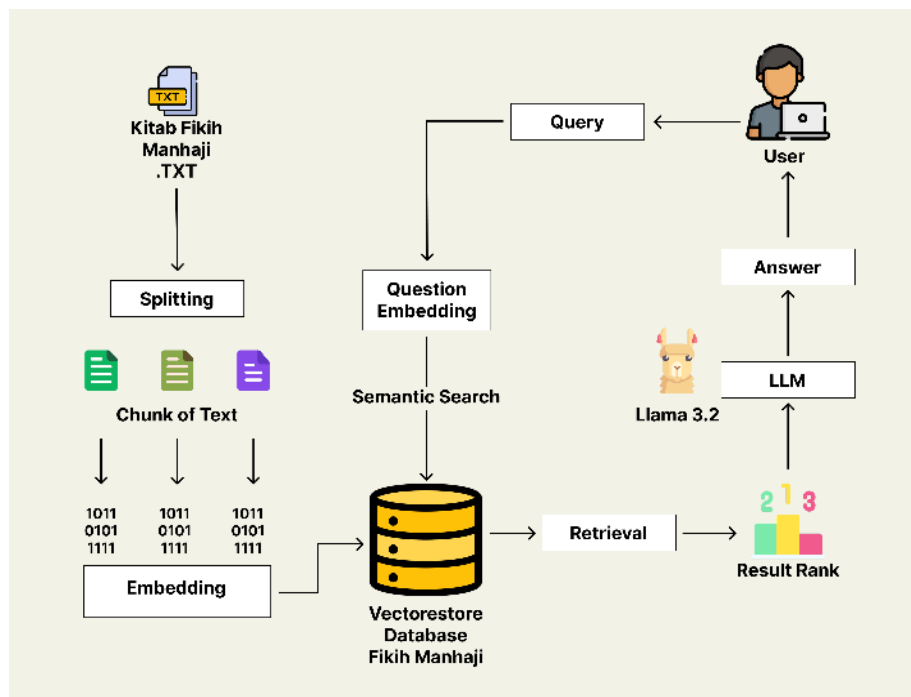
### 2.3.1. Analisa Kebutuhan Data

Data untuk penelitian ini dikumpulkan dari kitab fikih manhaji fikih lengkap imam Asy-Syafi'i. Data ini terdiri dari teks fikih manhaji sebagai data penelitian. Kemudian, data fikih manhaji disajikan dalam bentuk format (.txt), selanjutnya data tersebut di bersihkan agar tidak terjadi kesalahan pada saat di simpan ke database vektor. Setelah proses konversi dari buku fisik ke format digital (.txt), dilakukan proses validasi manual terhadap hasil OCR untuk memastikan bahwa setiap ayat, dalil, dan penjelasan dalam teks tidak mengalami kesalahan pembacaan. Proses ini mencakup pemeriksaan ejaan, keutuhan paragraf, serta kesesuaian makna dari setiap potongan teks dengan referensi aslinya. Validasi ini penting untuk menjaga integritas semantik sebelum proses embedding dilakukan.

### 2.3.2. Analisa Model Bahasa

Analisa model bahasa adalah proses menganalisa model bahasa yang sesuai digunakan untuk mengimplementasikan Sistem Tanya Jawab yang berfungsi dengan baik. Peneliti menggunakan Large Language Model (LLM) dan dibantu dengan metode Retrieval-Augmented Generation (RAG) untuk memprediksi kata-kata dan dapat menghasilkan teks yang baru secara otomatis serta mendapatkan jawaban yang relevan. LLM yang digunakan peneliti adalah model Llama-3.2:3b dari Meta.

## 2.4. Perancangan



Gambar 2. Arsitektur Sistem Tanya Jawab Fikih

Perancangan arsitektur sistem tanya jawab fikih manhaji ditunjukkan pada gambar 2. Tahapan pertama dalam pengembangan sistem tanya jawab fikih manhaji dimulai dari proses preprocessing data, yaitu mengubah data dalam bentuk buku fisik menjadi format digital (.txt). Setelah itu, dilakukan proses splitting teks, pada tahap ini, data yang telah diproses dibagi menjadi unit-unit kecil yang lebih terstruktur dan mudah dikelola, yang selanjutnya akan digunakan sebagai masukan untuk Large Language Model (LLM). Proses segmentasi ini dilakukan dengan memanfaatkan modul pemrosesan teks dari LangChain, yang dirancang untuk menghasilkan potongan teks dengan ukuran yang sesuai dan efisien untuk dikonsumsi oleh LLM, sehingga dapat menjaga konteks dan relevansi informasi [20]. yaitu membagi dokumen menjadi beberapa potongan teks (chunk of text) dengan ukuran 500 karakter. Kemudian proses embedding, embedding berfungsi untuk merepresentasikan unit teks dalam bentuk numerik, yang mencerminkan makna semantik dari teks tersebut. Ketika diterapkan pada teks-teks yang memiliki kesamaan isi, embedding akan menghasilkan vektor-vektor yang berada dalam kedekatan ruang vektor yang sama. Hal ini memungkinkan pengukuran kesamaan semantik antar teks secara lebih sistematis. Proses ini sangat krusial dalam memilih bagian-bagian teks yang relevan untuk dimasukkan ke dalam model bahasa saat menjawab pertanyaan. Representasi vektor ini disimpan dalam

basis data vektor. Ketika pengguna mengajukan pertanyaan, teks kueri tersebut diubah menjadi embedding, kemudian dibandingkan dengan embedding dalam basis data vektor menggunakan perhitungan kesamaan, seperti dot product, untuk menemukan bagian teks yang paling relevan secara semantik[21]. Sistem ini menerapkan metode Retrieval-Augmented Generation (RAG) untuk menjawab pertanyaan pengguna, Retrieval-Augmented Generation (RAG) memiliki kemampuan untuk mengekstraksi informasi dari basis pengetahuan eksternal. Dengan memanfaatkan kombinasi antara memori parametrik (yang tertanam dalam parameter model) dan non-parametrik (seperti basis data vektor eksternal), RAG cenderung mengurangi kecenderungan model dalam menghasilkan informasi yang tidak akurat (halusinasi), sekaligus meningkatkan interpretabilitas dalam berbagai tugas pemrosesan bahasa alami, seperti question answering dan rangkuman teks[22]. Dengan pendekatan RAG ini, sistem diharapkan mampu memberikan jawaban yang tidak hanya berdasarkan pengetahuan internal model, tetapi juga diperkuat oleh dokumen referensi fikih manhaji yang tersedia dalam database.

## 2.5. Implementasi

Implementasi Sistem Tanya Jawab pada penelitian ini menggunakan framework Streamlit pada sisi tampilan antarmuka. Streamlit merupakan framework yang dirancang khusus untuk membangun aplikasi web interaktif dengan Python. Penelitian ini menggunakan Ollama sebagai platform open-source untuk menjalankan Large Language Model secara lokal. Peneliti menggunakan model LLM dari Meta, yaitu Llama-3.2:3B.

## 2.6. Pengujian

Pengujian dilakukan untuk memastikan Sistem Tanya Jawab sudah layak digunakan dan memberikan informasi yang akurat kepada pengguna. Peneliti melakukan proses pengujian sistem dengan menggunakan BERTScore. Dalam bidang Natural Language Processing (NLP) yang lebih luas, telah terjadi kemajuan besar. Proses pengolahan bahasa alami mencakup interaksi antara komputer dan bahasa manusia. Kemajuan baru-baru ini dalam pembelajaran mendalam, khususnya dengan model seperti transformer, yaitu bidirectional encoder representations from transformers (BERT), telah membawa perubahan besar dalam aplikasi NLP seperti analisis sentimen, penerjemahan mesin, dan tanya jawab [23]. BERT dibuat untuk memahami kalimat yang ambigu dengan menggunakan konteks teks sekitarnya, menggunakan transformer untuk mendapatkan pemahaman yang lebih lengkap. Dengan menggunakan mekanisme self-attention, transformator belajar dan menyesuaikan pemahaman mereka tentang hubungan kontekstual antar kata. Dalam model pre-training, BERT menggunakan encoder dalam transformator untuk menyelesaikan berbagai tugas pemrosesan bahasa alami, termasuk Sentiment Analysis (SA), Question Answering (QA), dan Text Summarization (TS). Secara praktis, BERT melewati dua tahap dalam prosesnya: pelatihan awal untuk memahami bahasa dan penyesuaian untuk tugas tertentu. Menghitung BERTScore, persamaan ini digunakan [24].

Pengujian dilakukan menggunakan 10 pertanyaan yang disusun berdasarkan lima jilid kitab Fikih Manhaji, masing-masing diambil 2 pertanyaan per jilid. Pertanyaan-pertanyaan ini disusun untuk mencakup topik inti dari masing-masing jilid, sehingga mewakili variasi tema dalam dataset.

### 1) Precision (P)

Precision adalah tingkat kesesuaian prediksi ( $x_i$ ) jawaban model dengan data referensi ( $y_j$ ).

$$P = \frac{1}{|x|} \sum_{x_i \in x} \max_{y_j \in y} \text{sim}(x_i, y_j) \quad (1)$$

### 2) Recall (R)

Seberapa efektif model memprediksi kelas positif secara akurat ditentukan oleh recall.

$$R = \frac{1}{|y|} \sum_{y_j \in y} \max_{x_i \in x} \text{sim}(y_j, x_i) \quad (2)$$

### 3) F1-Score (F1)

F1-Score merupakan nilai rata-rata yang membandingkan precision dan recall.

$$F1 = \frac{2PR}{P+R} \quad (3)$$

## 3. HASIL DAN PEMBAHASAN

Pada bab ini peneliti melakukan beberapa tahapan, yaitu hasil perancangan sistem, hasil implementasi sistem yang telah dirancang dan membahas pengujian terhadap sistem serta dilanjutkan dengan pembahasan.

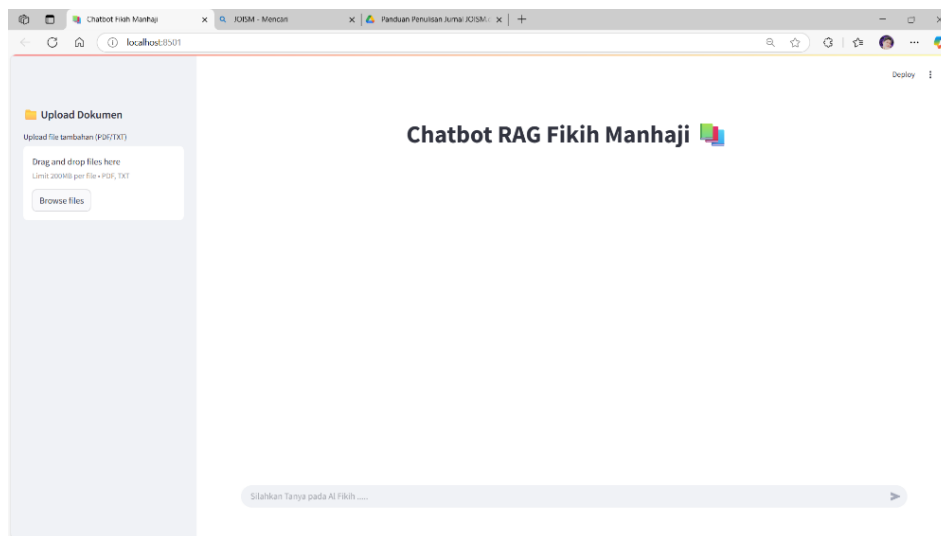
### 3.1. Hasil Perancangan

Pada proses ini peneliti menggunakan model LLM open source dari Meta yaitu “Llama-3.2:3B” dengan menjalankannya secara lokal menggunakan Ollama dan peneliti juga menggunakan framework LangChain untuk pengembangan sistem tanya jawab fikih manhaji. Model embedding yang digunakan peneliti adalah “intfloat/multilingual-e5-base” yang disediakan oleh HuggingFace secara gratis. Model embedding ini mendukung banyak bahasa. Kemudian menggunakan library ChromaDB untuk menyimpan data dan sebagai basis pengetahuan yang akan digunakan oleh LLM. Selanjutnya, hasil perancangan masuk ke proses implementasi.

## 3.2. Hasil Implementasi Sistem

### 3.2.1. Tampilan Utama

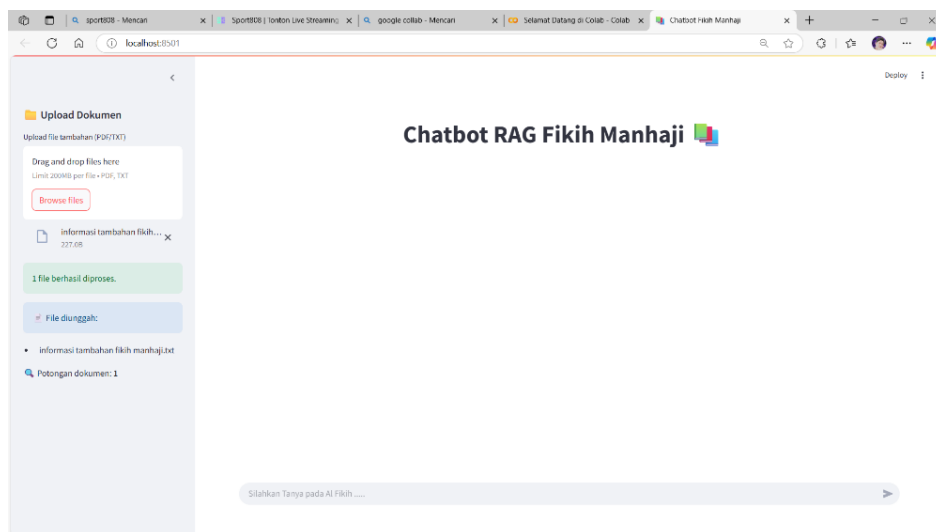
Tampilan ini merupakan tampilan utama dari sistem tanya jawab fikih manhaji. Pengguna dapat melakukan upload file dan memberikan pertanyaan dengan konteks fikih kepada sistem, seperti pada gambar 3.



Gambar 3. Tampilan Utama Sistem Tanya Jawab Fikih

### 3.2.2. Tampilan Upload File

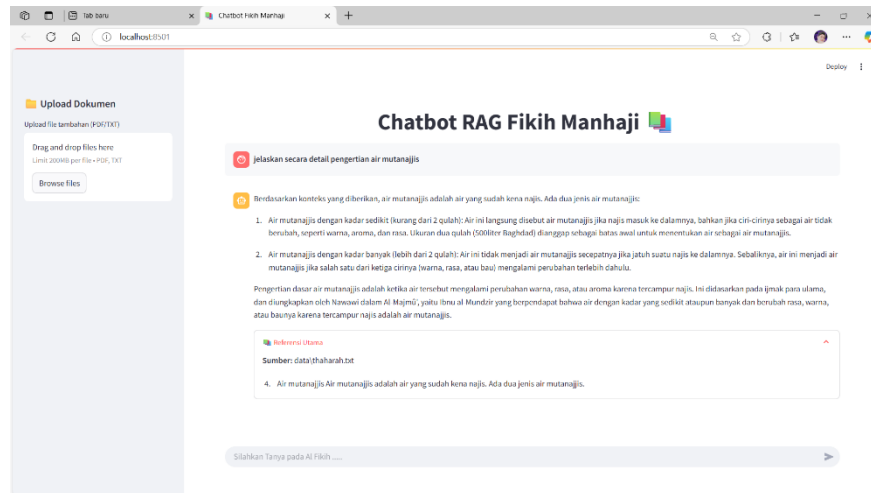
Halaman ini merupakan tampilan sistem tanya jawab pada saat pengguna melakukan upload file. Untuk mempermudah pengguna mengetahui file yang di upload berhasil, sistem akan memberikan keterangan bahwasannya file tersebut akan diproses dengan embedding. Setelah file berhasil di embedding, kemudian sistem memunculkan keterangan file berhasil di upload atau diunggah. Pengguna juga dapat melihat berapa total potongan file setelah di embedding.



Gambar 4. Tampilan Upload File Sistem Tanya Jawab Fikih

### 3.2.3. Tampilan Jawaban

Halaman ini merupakan halaman yang menampilkan jawaban dari pertanyaan yang diajukan oleh pengguna dan sistem menampilkan referensi yang sesuai dengan jawaban.



Gambar 5. Tampilan Jawaban Sistem Tanya Jawab Fiqih

### 3.3. Hasil Pengujian

Tabel 2. Pengujian Sistem Tanya Jawab Fiqih menggunakan BERTScore

No	Question	Precision	Recall	F1-Score
1	Jelaskan secara detail pengertian air mutanajjis!	0,7368	0,6444	0,6875
2	Sebutkan syarat-syarat tayamum!	0,9004	0,9249	0,9125
3	Apa hukum orang yang meninggalkan shalat?	0,8581	0,8027	0,8295
4	Apa hukum shalat id dan kapan waktu shalat id?	0,8138	0,9086	0,8586
5	Sebutkan pengertian syahid!	0,9385	0,9637	0,9509
6	Sebutkan rukun utama puasa!	0,9004	0,8745	0,8873
7	Apa saja binatang ternak yang wajib dizakati?	0,8585	0,8930	0,8754
8	Bagaimana hukum wakaf kepada non-Muslim?	0,8444	0,8060	0,8248
9	Kapan haji dan umrah pertama kali diwajibkan?	0,9623	0,9173	0,9392
10	Bagaimana hukum ibadah umrah menurut mazhab syafi'i dan sebutkan dalilnya?	0,8371	0,7407	0,7860
	Jumlah	8,6503	8,4758	8,5517
	Rata-rata	0,8650	0,8476	0,8552

Peneliti melakukan pengujian terhadap tingkat keakuratan jawaban yang dihasilkan oleh sistem dengan jawaban dari data fikih manhaji. Pengujian menggunakan BERTScore dengan 10 sampel pertanyaan yang diambil dari masing-masing jilid buku sebanyak 2 pertanyaan. Berdasarkan table 2, didapatkan nilai rata-rata dari precision atau tingkat keakuratan jawaban sistem sebesar 0,8650 atau 86,50%, recall tingkat kelengkapan jawaban sistem sebesar 0,8476 atau 84,76% dan F1-score atau nilai rata-rata perbandingan precision dan recall sebesar 0,8552 atau 85,52%. Hasil ini menunjukkan bahwa sistem mampu memberikan jawaban yang cukup akurat terhadap pertanyaan yang diajukan.

### 3.4. Pembahasan

Penelitian ini bertujuan untuk membuat sistem tanya jawab fikih manhaji fikih lengkap Imam Asy-Syafi'i berbasis web. Sistem dibangun dengan menggunakan Large Language Model (LLM) dari Meta, yaitu model Llama3-2:3B. Penelitian ini juga menggunakan framework LangChain dan Ollama untuk mendapatkan akses model secara lokal dan tidak menggunakan API Key sehingga mencegah batasan penggunaan. Penerapan metode Retrieval-Augmented Generation (RAG) dapat membantu mengurangi tingkat halusinasi yang di hasilkan oleh LLM, Sehingga sistem tanya jawab atau chatbot fikih manhaji dapat menjawab pertanyaan berdasarkan data yang dimiliki. Tahapan yang dilakukan peneliti adalah mengumpulkan data-data fikih dari buku Fiqih Manhaji: Fiqih Lengkap Imam Asy-Syafi'i, buku ini masih bersifat fisik sehingga penulis melakukan pemrosesan data dari buku tersebut menjadi format (.txt). Langkah selanjutnya adalah melakukan proses splitting atau membagi dokumen menjadi beberapa bagian yang disebut dengan chunk of text, kemudian melakukan proses embedding agar data bisa ditempatkan ke dalam vectorstore database.

Hasil implementasi tampilan antarmuka dalam bentuk web dengan Streamlit yaitu, framework yang dirancang khusus untuk membangun aplikasi web interaktif dengan Python. Hal ini mempermudah pengguna

melakukan interaksi tanya jawab dengan chatbot. Pengujian dilakukan dengan menggunakan BERTScore untuk mendapatkan kesesuaian jawaban yang diberikan sistem dengan jawaban berdasarkan data fikih manhaji. Hasil pengujian yang dilakukan mendapatkan tingkat akurasi yang baik sehingga sistem tanya jawab dapat memahami dan menjawab pertanyaan tentang fikih manhaji. Hasil pengujian nilai precision yang cukup tinggi (86,50%) menunjukkan bahwa sistem mampu menghasilkan jawaban yang sangat sesuai dengan referensi yang tersedia. Hal ini dimungkinkan karena kombinasi antara segmentasi teks yang baik, embedding multibahasa yang akurat, serta penggunaan metode RAG yang efektif dalam memilih konteks yang relevan. Namun, precision bisa menurun pada pertanyaan yang mengandung ambiguitas atau membutuhkan penalaran tambahan yang tidak eksplisit tertulis di teks, yang menyebabkan model memberikan respons di luar konteks atau kurang tepat.

Penelitian ini berbeda dengan penelitian sebelumnya, dengan menerapkan metode RAG, chatbot dapat menjawab pertanyaan berdasarkan referensi atau data yang telah disediakan, sehingga tingkat halusinasinya dapat diperkecil. Dengan Menggunakan model lokal Llama3-2:3B dapat mengatasi masalah terhadap keterbatasan akses pada API Key. Penelitian ini bertujuan untuk menerapkan metode RAG pada LLM agar dapat meminimalisir tingkat halusinasi pada model, sehingga informasi yang diberikan lebih akurat kepada pengguna dan mudah digunakan.

#### 4. KESIMPULAN

Penelitian yang telah dilakukan berhasil menerapkan sistem tanya jawab fikih manhaji berbasis web dengan menggunakan model Llama3-2:3B dari metaAI dengan penerapan metode Retrieval-Augmented Generation (RAG). Penelitian ini mempermudah pengguna untuk mendapatkan informasi terkait fikih manhaji dengan akurasi jawaban yang tinggi melalui web. Hasil penelitian dengan menggunakan BERTScore untuk mencari tingkat kesesuaian mendapatkan nilai precision sebesar 86,50% recall sebesar 84,76% dan F1-score sebesar 85,52%. Penelitian ini memiliki beberapa keterbatasan, yaitu sistem hanya diuji dengan 10 pertanyaan yang berasal dari 5 jilid kitab fikih manhaji, sehingga cakupan evaluasinya masih terbatas. Model dijalankan secara lokal sehingga performa dapat menurun pada perangkat yang memiliki spesifikasi rendah. Saran yang dapat diberikan penulis untuk pengembangan sistem tanya jawab fikih manhaji di masa mendatang yaitu: pada sistem tanya jawab fikih manhaji dengan metode RAG, dapat menghasilkan jawaban yang lebih akurat dengan menggunakan model LLM yang memiliki pengetahuan lebih baik dan model embedding yang mendukung banyak Bahasa terutama untuk kajian Islam.

#### REFERENSI

- [1] S. Ariga, "Pelaksanaan Pembelajaran Fiqih Dengan Kitab Al-Fiqh Al-Manhaj 'Ala Madzhabil Imam Asy-Syafi'i," *MUDABBIR (Journal Research and Education Studies)*, vol. 2, no. 1, Jan. 2022, doi: <https://doi.org/10.56832/mudabbir.v2i1.236>.
- [2] Eldi and H. Syaputra, "Implementasi Chatbot Untuk Mendukung Sistem Informasi Pada Rumah Sakit Muhammadiyah Palembang," *Jurnal Nasional Ilmu Komputer*, vol. 1, no. 3, pp. 2746–1343, Aug. 2020, doi: <https://doi.org/10.47747/jurnalnik.v1i3.160>.
- [3] N. Hikmawati, S. M. Imam, and Jamilah, "Konsep Dan Implementasi Kecerdasan Buatan (Artificial Intelligence) Dalam Manajemen Kurikulum SD/MI," *ABUYA: Jurnal Pendidikan Dasar*, vol. 1, 2023, doi: <https://doi.org/10.52185>.
- [4] E. Adamopoulou and L. Moussiades, "An Overview of Chatbot Technology," *IFIP Adv Inf Commun Technol*, vol. 584 IFIP, pp. 373–383, 2020, doi: [10.1007/978-3-030-49186-4\\_31](https://doi.org/10.1007/978-3-030-49186-4_31).
- [5] 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 (Dordr)*, vol. 3, no. Vol. 3 (2024), pp. 177–186, Mar. 2024.
- [6] D. I. Mulyana, D. Lestari, F. Ramdhani, M. J. Ruliansyah, and R. Beay, "Implementasi Chatbot Telegram Dalam Meningkatkan Partisipasi Kegiatan Warga," *Jurnal Pengabdian kepada Masyarakat Nusantara (JPkMN)*, vol. 4, pp. 866–874, Jul. 2023.
- [7] F. Ishlakhuddin, A. Basir, and Nurlaela, "Rancang Bangun Sistem Tanya-jawab Berbasis Aturan STMIK Muhammadiyah Paguyangan Brebes dengan Menggunakan Telegram Chatbot," *Jurnal Informatika: Jurnal pengembangan IT (JPIT)*, vol. 5, no. 3, Sep. 2020, doi: <https://doi.org/10.30591/jpit.v5i3.2900>.
- [8] G. F. Avisyah, I. J. Putra, and S. S. Hidayat, "Open Artificial Intelligence Analysis using ChatGPT Integrated with Telegram Bot," *Jurnal ELTIKOM*, vol. 7, no. 1, pp. 60–66, Jun. 2023, doi: [10.31961/eltikom.v7i1.724](https://doi.org/10.31961/eltikom.v7i1.724).
- [9] Nurhapiza, N. S. Harahap, M. Fikry, and M. Affandes, "Penerapan Chatbot pada Aplikasi Web Tanya Jawab Tentang Fiqih Jual Beli Islam Menggunakan LangChain," *Journal of Computer System and Informatics (JoSYC)*, vol. 5, pp. 548–557, May 2024, doi: [10.47065/josyc.v5i3.5148](https://doi.org/10.47065/josyc.v5i3.5148).
- [10] D. Firdaus, I. Sumardi, and Y. Kulsum, "Integrating Retrieval-Augmented Generation with Large Language Model Mistral 7b for Indonesian Medical Herb," *JISKA (Jurnal Informatika Sunan Kalijaga)*, vol. 9, no. 3, pp. 230–243, Sep. 2024, doi: [10.14421/jiska.2024.9.3.230-243](https://doi.org/10.14421/jiska.2024.9.3.230-243).
- [11] S. Elysia and Herianto, "Chatbot Berbasis Retrieval Augmented Generation (RAG) untuk Peningkatan Layanan Informasi Sekolah," *Journal TIFDA (Technology Information and Data Analytic)*, vol. 1, no. 2, pp. 52–58, Dec. 2024, doi: [10.70491/tifda.v1i2.52](https://doi.org/10.70491/tifda.v1i2.52).
- [12] H. Tohir, N. Merlina, and M. Haris, "Utilizing Retrieval-Augmented Generation In Large Language Models To Enhance Indonesian Language NLP," *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 10, no. 2, pp. 352–360, Nov. 2024, doi: [10.33480/jitk.v10i2.5916](https://doi.org/10.33480/jitk.v10i2.5916).
- [13] E. Afriani, N. Safaat H, M. Fikry, and M. Affandes, "Aplikasi Tanya Jawab Tentang Fiqih Bersuci Berbasis Web," *ZONAsi: Jurnal Sistem Informasi*, vol. 6, no. 2, pp. 380–390, May 2024, doi: <https://doi.org/10.31849/zn.v6i2.19970>.

- [14] N. A. M. Herwanza, N. S. Harahap, F. Yanto, and F. Insani, "Penerapan Langchain Retriever dengan Model Chat Openai dalam Pengembangan Sistem Chatbot Hadis Berbasis Telegram," *JTIM : Jurnal Teknologi Informasi dan Multimedia*, vol. 6, no. 1, pp. 70–83, May 2024, doi: 10.35746/jtim.v6i1.514.
- [15] H. A. Alawwad, A. Alhothali, U. Naseem, A. Alkhathlan, and A. Jamal, "Enhancing Textbook Question Answering Task with Large Language Models and Retrieval Augmented Generation," Feb. 2024, [Online]. Available: <http://arxiv.org/abs/2402.05128>
- [16] D. Apriliani, S. F. Handayani, T. N. Anugrahaeni, A. Miftahudin, L. Nurarifiah, and I. T. Saputra, "Aplikasi Question Answer Sebagai Media Pembelajaran Interaktif Untuk Mata Pelajaran Akuntansi," *JMM (Jurnal Masyarakat Mandiri)*, vol. 7, no. 2, p. 2003, Apr. 2023, doi: 10.31764/jmm.v7i2.13867.
- [17] M. Irfan, N. S. Harahap, Novriyanto, and S. Sanjaya, "Penerapan Retrieval Augmented Generation Menggunakan Langchain dalam," *ZONAsi: Jurnal Sistem Informasi*, vol. 6, no. 2, pp. 370–379, May 2024, doi: <https://doi.org/10.31849/zn.v6i2.19940>.
- [18] I. Pujiono, I. M. Agtyaputra, and Y. Ruldeviyani, "Implementing Retrieval-Augmented Generation And Vector Databases For Chatbots In Public Services Agencies Context," *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 10, no. 1, pp. 216–223, Aug. 2024, doi: 10.33480/jitk.v10i1.5572.
- [19] M. Al-Bugha, M. Al-Khann, and A. Asy-Syurbaji, *Fikih Manhaji: Fikih Lengkap Imam Asy-Syafi'i*, Jilid 1: Taharah. Yogyakarta: Pro-U Media, 2018.
- [20] J. Jeong, D. Gil, D. Kim, and J. Jeong, "Current Research and Future Directions for Off-Site Construction through LangChain with a Large Language Model," *Buildings*, vol. 14, no. 8, Aug. 2024, doi: 10.3390/buildings14082374.
- [21] O. Topsakal and T. C. Akinci, "Creating Large Language Model Applications Utilizing LangChain: A Primer on Developing LLM Apps Fast," *International Conference on Applied Engineering and Natural Sciences*, vol. 1, no. 1, pp. 1050–1056, Jul. 2023, doi: 10.59287/icaens.1127.
- [22] S. Siriwardhana, R. Weerasekera, E. Wen, T. Kaluarachchi, R. † Rajib, and S. Nanayakkara, "Improving the Domain Adaptation of Retrieval Augmented Generation (RAG) Models for Open Domain Question Answering", doi: 10.1162/tacl.
- [23] N. S. Harahap, A. Saad, and N. H. Ubaidullah, "Comprehensive Bibliometric Literature Review of Chatbot Research: Trends, Frameworks, and Emerging Applications," *International Journal of Advanced Computer Science and Applications*, vol. 16, no. 1, pp. 886–896, 2025, doi: 10.14569/IJACSA.2025.0160185.
- [24] S. Rahayu, N. S. Harahap, S. Agustian, and P. Pizaini, "Penerapan Teknologi LangChain pada Question Answering System Fikih Empat Madzhab," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 3, pp. 974–983, Jun. 2024, doi: 10.57152/malcom.v4i3.1397.