

Penerapan Teknologi Langchain dan LLM pada Sistem Question Answering Berbasis Chatbot Telegram: *Literature Review*

Danang Kurniawan^{1a*}, Joko Triloka^{2b}

^{ab}Institut Informatika dan Bisnis Darmajaya

^adanang.2321210009@mail.darmajaya.ac.id

^bjoko.triloka@darmajaya.ac.id

Abstract

This study explores the application of LangChain technology and Large Language Models (LLM) in a Question Answering (QA) system based on Telegram chatbots, designed to provide accurate and relevant automatic answers on various topics. The main objective of this research is to utilize LangChain and LLM to enhance the interactivity and accuracy of the chatbot system in answering questions related to Islamic Business Fiqh, Tafsir Al-Jalalain, Tafsir Al-Azhar, and Health Law. This research investigates how these technologies can be integrated into a Telegram-based chatbot to process complex questions and provide contextually appropriate answers. The system processes data using techniques like chunking and embeddings to match questions with relevant answers effectively. Testing results indicate that the system delivers high accuracy in providing answers, as evaluated through BERTScore and ROUGE Score metrics. Despite this, challenges remain in handling long texts with multiple interpretations and ensuring answers are highly specific to the context of each question. User Acceptance Testing (UAT) showed positive results, with high user satisfaction, but also highlighted the need for improvement in contextual management and reducing information errors or hallucinations. This study recommends further development, including improving chatbot interactivity to respond better to various types of questions and enhancing the system's ability to provide more natural and contextually relevant answers. Overall, the study demonstrates the significant potential of LangChain and LLM technologies in advancing chatbot systems, making them effective solutions for diverse applications in religious, legal, and other fields.

Keywords : LangChain; Large Language Models; Telegram Chatbot; Question Answering; Chatbot Interactivity

Abstrak

Penelitian ini membahas penerapan teknologi LangChain dan Large Language Models (LLM) dalam sistem Question Answering (QA) berbasis chatbot Telegram, yang bertujuan memberikan jawaban otomatis yang cepat, akurat, dan relevan dalam berbagai topik. Fokus utama penelitian ini adalah memanfaatkan LangChain dan LLM untuk meningkatkan interaktivitas dan akurasi sistem dalam menjawab pertanyaan terkait Fiqh Jual Beli Islam, Tafsir Al-Jalalain, Tafsir Al-Azhar, dan Undang-Undang Kesehatan. Penelitian ini mengeksplorasi integrasi kedua teknologi tersebut dalam chatbot berbasis Telegram untuk memproses pertanyaan kompleks dan memberikan jawaban yang sesuai dengan konteks. Data dalam sistem diproses melalui teknik chunking dan embeddings untuk mencocokkan pertanyaan dengan jawaban yang relevan. Hasil pengujian menunjukkan bahwa sistem mampu memberikan jawaban dengan tingkat akurasi tinggi, yang dievaluasi menggunakan BERTScore dan ROUGE Score. Meskipun demikian, tantangan utama yang dihadapi adalah pengelolaan teks panjang dengan banyak interpretasi dan kemampuan memberikan jawaban yang lebih spesifik sesuai konteks. Evaluasi melalui User Acceptance Testing (UAT) menunjukkan hasil positif dengan tingkat kepuasan pengguna yang tinggi. Namun, masih ada ruang untuk perbaikan, terutama dalam hal pengelolaan konteks dan pengurangan kesalahan informasi atau hallucinations. Penelitian ini juga merekomendasikan pengembangan lebih lanjut dalam meningkatkan interaktivitas chatbot agar lebih responsif terhadap berbagai jenis pertanyaan serta meningkatkan kemampuan sistem dalam memberikan jawaban yang lebih natural dan relevan dengan konteks.

Keywords : LangChain; Large Language Models; Chatbot Telegram; Question Answering; Interaktivitas Chatbot

1. PENDAHULUAN

Pada era digital yang semakin maju, teknologi *Large Language Models* (LLMs) telah berkembang pesat dan semakin banyak diterapkan dalam berbagai sektor. Salah satu implementasi terbarunya adalah dalam pengembangan sistem *Question Answering* (QA) berbasis chatbot, yang memungkinkan pengguna untuk berinteraksi secara langsung dengan mesin untuk mendapatkan jawaban dari berbagai pertanyaan yang diajukan. Dalam konteks ini, platform Telegram menjadi pilihan utama, mengingat kemampuannya dalam menyediakan antarmuka interaktif

yang dapat diakses secara luas, serta kemudahan dalam integrasi dengan berbagai teknologi seperti Langchain dan LLM. Teknologi Langchain, sebagai framework yang menghubungkan LLM dengan berbagai alat pendukung lainnya, berfungsi untuk memperkaya kemampuan pemrosesan bahasa alami dan meningkatkan keakuratan jawaban yang diberikan oleh chatbot (Aji Bayu Permadi et al., 2024). *Question Answering* juga sangat diperlukan dalam bidang tafsir seperti yang dibahas oleh (Rizki et al., 2024) yang membahas pengembangan sistem tanya jawab berbasis chatbot untuk Tafsir Al-Jalalain menggunakan teknologi LangChain dan *Large Language Models* (LLM). Penelitian ini berfokus pada penerapan kecerdasan buatan untuk memberikan akses yang lebih cepat, efisien, dan akurat dalam memahami tafsir Al-Qur'an, khususnya Tafsir Al-Jalalain, melalui sistem tanya jawab berbasis Telegram.

Melalui pemanfaatan Langchain, sistem QA berbasis Telegram dapat memanfaatkan kekuatan LLM dalam mengolah teks dan menghasilkan jawaban yang relevan. Hal ini sejalan dengan berbagai penelitian terdahulu yang telah menunjukkan potensi besar dari kedua teknologi ini. Misalnya, penelitian yang dilakukan oleh (Aji Bayu Permadi et al., 2024) yang mengimplementasikan sistem QA pada Tafsir Al-Azhar dengan memanfaatkan Langchain dan LLM, yang menunjukkan bagaimana kedua teknologi ini dapat bekerja bersama untuk memberikan jawaban yang akurat dan berbobot dalam konteks agama. Begitu pula, penelitian oleh (Herwanza et al., 2024) yang memanfaatkan Langchain untuk membangun chatbot berbasis Telegram yang dapat menjawab pertanyaan seputar hadis, membuktikan bahwa integrasi OpenAI's GPT dengan Langchain memberikan hasil yang memuaskan dalam hal ketepatan dan keberagaman jawaban.

Selanjutnya (Vidivelli et al., 2024) mengusulkan sebuah metodologi baru dalam pengembangan chatbot kustom yang fokus pada efisiensi. Metodologi ini menggabungkan tiga inovasi utama: LangChain untuk penyesuaian chatbot dengan tujuan spesifik, Retrieval-Augmented Generation (RAG) yang memungkinkan chatbot mengakses data secara luas untuk memberikan respon yang lebih lengkap, dan model bahasa besar (LLMs) yang dioptimalkan dengan strategi seperti LoRA dan QLoRA untuk meningkatkan kinerja. Kombinasi ini menghasilkan chatbot yang lebih efektif dalam berinteraksi dengan pengguna, memberikan jawaban yang lebih tepat dan cepat, serta meningkatkan produktivitas pengembangan chatbot secara keseluruhan. Kemudian hal yang serupa juga di bahas pada penelitian (Firdaus et al., 2024) yang mengkaji penggunaan teknik Retrieval-Augmented Generation (RAG) untuk meningkatkan performa model bahasa besar (LLM) seperti Mistral 7b dalam konteks pengobatan herbal di Indonesia. Penelitian ini bertujuan untuk mengoptimalkan kemampuan model dalam menghasilkan teks yang lebih relevan dan berbobot dengan cara menggabungkan teknik pengambilan informasi (retrieval) dengan kemampuan generatif dari LLM. Dalam studi ini, model Mistral 7b yang memiliki 7 miliar parameter dimanfaatkan untuk meningkatkan efisiensi dalam memproses dan menghasilkan teks yang terkait dengan pengobatan herbal Indonesia. Teknik RAG digunakan untuk menambah wawasan dengan mengambil informasi dari berbagai sumber eksternal yang relevan, sehingga model dapat memberikan respon yang lebih lengkap dan akurat. Tidak berhenti disitu, (Pujiono et al., 2024) juga mengkaji penerapan teknologi *Retrieval-Augmented Generation* (RAG) dan *vector databases* dalam pengembangan chatbot untuk layanan publik. Penelitian ini bertujuan untuk mengukur kinerja chatbot yang dikembangkan dengan menggunakan metode RAG, yang menggabungkan pencarian informasi eksternal dengan kemampuan generasi teks, serta penggunaan *vector database* untuk meningkatkan efisiensi dalam mengelola dan mengambil data yang relevan.

Merujuk pada teknologi NLP dibahas pada penelitian (Fauzan et al., 2024) yang mengkaji penerapan teknologi pemrosesan bahasa alami (NLP) untuk merancang chatbot dalam sistem informasi penerimaan siswa baru di sebuah website. Fokus utama penelitian ini adalah untuk mengatasi tantangan dalam memberikan layanan yang responsif dan berkualitas kepada calon siswa, mengingat banyaknya permintaan informasi yang perlu diproses dengan cepat. Kemudian jika mengombinasikan pendekatan LLM dengan platform Google DialogFlow bisa juga dilakukan untuk sektor pertanian yang dalam pengembangannya penelitian ini berfokus pada pengembangan chatbot yang dapat memberikan informasi dan dukungan secara real-time kepada petani, mengingat masalah di sektor pertanian yang sering kali melibatkan pengetahuan mendalam, manajemen, dan akses informasi yang kompleks (Mohamad Alvian Fazlur Febrian Dhika, 2024).

Teknologi *Generative Pre-Trained Transformer 3.5* (GPT-3.5) juga memiliki keterkaitan dengan pengembangan chatbot dalam berbagai bidang, seperti yang dikembangkan oleh (Elvino Dwi Saputra, 2024) yang membahas penggunaan teknologi *Generative Pre-Trained Transformer 3.5* (GPT-3.5) dalam pengembangan sistem tanya jawab berbasis chatbot untuk menjawab pertanyaan seputar Al-Qur'an. Penelitian ini bertujuan untuk mengoptimalkan proses pencarian dan pemahaman informasi terkait Al-Qur'an dengan memanfaatkan kecanggihan kecerdasan buatan (AI). Kemudian (Saputra et al., 2024) dalam penelitiannya juga membahas pengembangan sistem tanya jawab untuk tafseer (penafsiran) Al-Qur'an menggunakan model bahasa besar (*Large Language Models* - LLM). Penelitian ini berfokus pada aplikasi kecerdasan buatan untuk memahami dan memberikan jawaban yang relevan terkait tafseer Ibnu Katsir, salah satu tafseer yang paling banyak digunakan dan dihormati di kalangan umat Islam.

Pada pembahasan (Alsafari et al., 2024) dalam penelitiannya yang mengkaji perkembangan teknologi chatbot dalam sektor pendidikan, dengan fokus pada pembuatan asisten pengajaran untuk mata kuliah Data Mining dan Text

Analytics di University of Leeds. Penelitian ini bertujuan untuk menyelidiki dan membandingkan pendekatan chatbot tradisional berbasis niat (intent-based) dengan metode canggih Retrieval-Augmented Generation (RAG), yang dapat meningkatkan efisiensi dan kemampuan adaptasi asisten pengajaran dalam konteks pendidikan tinggi. Dalam studi ini, pengembangan dimulai dengan pembuatan "Amazon Alexa teaching skill", dan hasil penelitian menunjukkan bagaimana teknologi AI yang lebih maju dapat mengubah peran chatbot sebagai asisten pengajaran, meningkatkan interaksi dengan siswa dan kemampuan chatbot untuk memberikan bantuan yang lebih kontekstual dan relevan.

Kemudian berangkat dari kesulitan dalam mengakses dan mengorganisir informasi tentang fiqh bersuci, (Afriani et al., 2024) melakukan penelitian yang mengkaji penerapan teknologi *Artificial Intelligence* (AI), khususnya sistem *Question Answering* (QAS), dalam meningkatkan akses dan pemahaman masyarakat terhadap fiqh bersuci. Fokus utama penelitian adalah pengembangan aplikasi berbasis web yang dapat memberikan jawaban otomatis terkait fiqh bersuci melalui sistem tanya jawab, sehingga memudahkan masyarakat dalam memahami ajaran fiqh dengan cara yang lebih cepat dan efisien.

Selain itu, (Nurhapiza et al., 2024) mengembangkan aplikasi QA berbasis Telegram untuk menjawab pertanyaan terkait fiqh jual beli Islam, memanfaatkan Langchain dan LLM untuk menyesuaikan jawaban dengan konteks hukum Islam yang relevan. Hal serupa juga terlihat pada penerapan Langchain dalam UU Kesehatan yang dilakukan oleh (Lubis et al., 2024), yang menunjukkan bahwa LLM dapat digunakan untuk menjawab pertanyaan yang berkaitan dengan hukum dan peraturan secara jelas dan tepat.

Berdasarkan berbagai penelitian yang telah ada, penelitian ini akan mengkaji lebih mendalam mengenai persentase dan tingkat keakuratan penerapan Langchain dan LLM dalam sistem *Question Answering* berbasis chatbot Telegram. Selain itu, penelitian ini bertujuan untuk mengevaluasi efektivitas teknologi ini dalam meningkatkan kinerja dan kualitas sistem dalam memberikan jawaban yang relevan dan akurat, serta menggali potensi besar dari kedua teknologi tersebut dalam berbagai aplikasi di berbagai bidang, seperti agama, hukum, fiqh, dan UU Kesehatan.

2. KERANGKA TEORI

2.1. *Large Language Models (LLM)*

Large Language Models (LLM) merujuk pada model bahasa besar yang dilatih menggunakan data dalam jumlah besar, yang dirancang untuk memahami, menghasilkan, dan merespons teks dalam konteks yang luas. **LLM** seperti GPT-3.5-turbo (yang digunakan dalam penelitian (Chen et al., 2024)) dikembangkan oleh OpenAI dan merupakan contoh dari model bahasa yang telah melalui pelatihan dengan berbagai jenis data, termasuk teks dari literatur ilmiah, buku, artikel, dan dokumen lainnya (Chen et al., 2024).

2.2. Sistem *Question Answering* (QA) Berbasis Chatbot

Sistem *Question Answering* (QA) berbasis chatbot adalah sebuah sistem yang dirancang untuk mempermudah masyarakat dalam mencari informasi dan memberikan jawaban yang sesuai dengan permintaan pengguna. Dalam sistem ini, pengguna dapat bertanya dengan cepat dan singkat, dan chatbot memberikan jawaban langsung tanpa perlu membuka halaman web terkait. Sistem ini merupakan pengembangan dari *Natural Language Processing* (NLP) yang memfasilitasi interaksi antara manusia dan komputer melalui bahasa manusia, sehingga membuatnya lebih intuitif dan mudah digunakan (Lubis et al., 2024).

2.3. Chatbot Telegram

Chatbot Telegram adalah sebuah sistem yang dibangun untuk berinteraksi dengan pengguna melalui platform Telegram. Pengembangan chatbot ini dilakukan dengan memanfaatkan API terbuka yang disediakan oleh Telegram, yang memungkinkan integrasi bot dengan mudah. Chatbot ini dapat digunakan untuk berbagai tujuan, termasuk memberikan informasi dan menjawab pertanyaan pengguna secara langsung (Lubis et al., 2024).

Sedangkan dalam penelitian (Rizki et al., 2024) chatbot telegram diartikan sebagai sistem berbasis *Question Answering* (QA) yang dirancang untuk memberikan jawaban atas pertanyaan terkait Tafsir Al-Jalalain. Chatbot ini menggunakan *Large Language Models* (LLM) dan Langchain untuk memproses pertanyaan yang diajukan oleh pengguna dan memberikan jawaban yang relevan secara efisien dan akurat.

2.4. *Teori Pemrosesan Bahasa Alami (NLP)*

Natural Language Processing (NLP) adalah suatu teknologi yang digunakan untuk memproses dan memahami bahasa alami, memungkinkan sistem untuk berinteraksi dengan pengguna dalam percakapan yang lebih alami dan mudah dipahami. Secara khusus, NLP digunakan dalam chatbot untuk mengubah percakapan yang sebelumnya bersifat mekanis menjadi percakapan yang lebih manusiawi dan mudah dimengerti. Dengan menggunakan metode NLP, chatbot ini mampu memahami intent dari pesan yang dikirim oleh pengguna dan memberikan respons yang relevan. Ini membantu meningkatkan kecepatan dan efisiensi dalam memberikan jawaban tanpa memerlukan staf untuk merespons secara langsung (Fauzan et al., 2024).

Metode **NLP** yang diterapkan pada chatbot di sini juga berfungsi untuk meningkatkan kualitas layanan dengan mengurangi keterlambatan dalam memberikan jawaban atas pertanyaan umum, sehingga mempercepat proses pendaftaran dan meningkatkan pengalaman pengguna di sistem pendaftaran siswa baru. Hasil dari pengujian sistem menunjukkan bahwa chatbot ini memiliki akurasi yang baik dalam merespons pertanyaan, dengan tingkat keberhasilan mencapai 87,5% (Fauzan et al., 2024).

2.5. Langchain

Langchain adalah framework yang digunakan untuk membangun dan mengelola alur percakapan dalam chatbot. Langchain mengintegrasikan berbagai metode dan teknik untuk memperlancar interaksi antara pengguna dan sistem, memanfaatkan Large Language Models (LLM) untuk mengelola bahasa alami. Secara khusus, Langchain digunakan dalam konteks pembangunan sistem Question Answering (QA) berbasis chatbot. Dalam aplikasi ini, Langchain berfungsi untuk mengelola alur percakapan, menghubungkan input dari pengguna (pertanyaan) dengan LLM yang telah disesuaikan dengan data tertentu (misalnya, data Tafsir Al-Azhari) untuk memberikan jawaban yang relevan dan sesuai. Langchain memungkinkan chatbot untuk menangani pertanyaan yang lebih kompleks dengan mengoptimalkan penggunaan LLM, serta memastikan percakapan yang lebih alami dan terstruktur dengan baik (Aji Bayu Permadi et al., 2024).

Dengan menggunakan Langchain, sistem dapat lebih efisien dalam mengelola percakapan dan memungkinkan chatbot untuk memberikan respons yang lebih tepat dan cepat, tanpa memerlukan interaksi manual dari staf. Langchain juga membantu dalam menangani berbagai jenis pertanyaan dan memastikan bahwa chatbot dapat terus belajar dan beradaptasi dengan informasi baru yang mungkin ditambahkan ke dalam basis data (Aji Bayu Permadi et al., 2024).

3. METODOLOGI

Metode penelitian ini menggunakan pendekatan literature review untuk mengevaluasi dan menganalisis berbagai penerapan teknologi Langchain dan LLM dalam sistem Question Answering berbasis chatbot Telegram. Dalam penelitian ini, kami melakukan review terhadap lima paper yang relevan dengan topik ini untuk mengeksplorasi berbagai aspek yang terkait dengan keakuratan dan efektivitas teknologi dalam meningkatkan kinerja sistem. Penelitian-penelitian yang direview mencakup berbagai implementasi Langchain dan LLM dalam konteks aplikasi berbasis Telegram yang berbeda, seperti Tafsir Al-jalalain (Rizki et al., 2024), Tafsir Al-Azhar (Aji Bayu Permadi et al., 2024), hadis (Herwanza et al., 2024), fiqih jual beli Islam (Nurhapiza et al., 2024), dan UU Kesehatan (Lubis et al., 2024). Berikut adalah rincian langkah-langkah metodologi yang digunakan:

3.1 Pendekatan Penelitian

Penelitian ini menggunakan metode tinjauan literatur, yang memungkinkan peneliti untuk menggali informasi dari berbagai sumber ilmiah yang telah ada. Dengan pendekatan ini, peneliti dapat memanfaatkan hasil dan temuan penelitian sebelumnya untuk memahami bagaimana teknologi LangChain dan LLM telah diterapkan di berbagai domain. Fokus utama adalah untuk menganalisis efektivitas, akurasi, dan tantangan teknologi ini dalam pengembangan chatbot interaktif berbasis Telegram. Pendekatan ini dipilih karena memberikan gambaran menyeluruh tanpa memerlukan eksperimen langsung, sehingga cocok untuk mengevaluasi teknologi yang sedang berkembang pesat.

3.2 Pengumpulan Data

Proses pengumpulan data dilakukan dengan seleksi ketat terhadap artikel ilmiah dari jurnal terindeks. Lima artikel utama yang dipilih memiliki relevansi tinggi dengan topik penelitian, mencakup aplikasi LangChain dan LLM di domain seperti Tafsir Al-Jalalain, Tafsir Al-Azhar, Fiqih Jual Beli Islam, dan UU Kesehatan. Kriteria pemilihan meliputi:

- a. Fokus pada pengembangan sistem QA berbasis Telegram.
- b. Penyajian data kuantitatif seperti akurasi dan relevansi jawaban.
- c. Inovasi atau pendekatan unik yang digunakan dalam setiap penelitian.

Artikel-artikel ini dianalisis untuk mengidentifikasi bagaimana LangChain dan LLM memproses data kompleks, serta sejauh mana teknologi ini meningkatkan kemampuan chatbot dalam menjawab pertanyaan berbasis konteks.

3.3 Analisis Data

Analisis data menggunakan metode content analysis untuk menggali informasi penting dari artikel yang direview. Teknik ini melibatkan identifikasi pola, tema utama, dan indikator keberhasilan dari setiap implementasi teknologi. Aspek yang dianalisis meliputi:

- a. Tingkat akurasi: Menggunakan metrik seperti BERTScore dan ROUGE Score untuk menilai kualitas jawaban chatbot.
- b. Konteks jawaban: Menilai seberapa relevan jawaban dengan pertanyaan pengguna.
- c. Kendala teknis: Mengidentifikasi tantangan seperti manajemen teks panjang dan pengurangan hallucination.

Proses ini membantu memberikan wawasan yang mendalam tentang kelebihan dan keterbatasan teknologi.

3.4 Instrumen Evaluasi

Untuk mengevaluasi efektivitas teknologi, beberapa instrumen digunakan, yaitu:

- a. BERTScore: Digunakan untuk mengevaluasi kesamaan antara jawaban yang dihasilkan chatbot dengan referensi jawaban.
- b. ROUGE Score: Mengukur kesesuaian antara jawaban yang diberikan dengan teks sumber.
- c. User Acceptance Testing (UAT): Metode ini mengevaluasi kepuasan pengguna berdasarkan pengalaman interaksi dengan chatbot.
- d. DeepEval Framework: Digunakan untuk menganalisis hallucination score dan contextual relevancy, yaitu aspek-aspek yang menunjukkan ketepatan dan relevansi jawaban yang dihasilkan chatbot.

Hasil dari setiap instrumen ini memberikan gambaran komprehensif tentang kualitas dan efektivitas sistem.

3.5 Sampel Penelitian

Penelitian ini mencakup lima domain aplikasi utama sebagai studi kasus, yaitu:

1. Tafsir Al-Jalalain: Menguji kemampuan chatbot menjawab pertanyaan kompleks terkait tafsir.
2. Tafsir Al-Azhar: Fokus pada analisis kemampuan chatbot dalam mengelola konteks.
3. Hadis: Mengukur akurasi jawaban chatbot dalam konteks keagamaan.
4. Fiqih Jual Beli Islam: Mengkaji relevansi jawaban dengan prinsip hukum Islam.
5. UU Kesehatan: Mengevaluasi sistem dalam menangani pertanyaan berbasis hukum dan regulasi.

Setiap domain memberikan perspektif unik tentang potensi teknologi LangChain dan LLM di berbagai bidang.

3.6 Prosedur Penelitian

Prosedur penelitian dilakukan melalui langkah-langkah berikut:

1. Identifikasi artikel: Artikel yang relevan dipilih dari jurnal dan konferensi yang diakui.
2. Klasifikasi temuan: Hasil penelitian diklasifikasikan berdasarkan domain aplikasi, metrik evaluasi, dan tantangan implementasi.
3. Analisis perbandingan: Temuan dibandingkan untuk menilai kelebihan dan kelemahan teknologi di setiap domain.

Prosedur ini memastikan bahwa setiap langkah dilakukan secara sistematis untuk memperoleh hasil yang dapat diandalkan.

3.7 Analisis dan Interpretasi Hasil

Hasil dari analisis data kemudian diinterpretasikan untuk menjawab pertanyaan penelitian utama, yaitu:

- a. Seberapa akurat teknologi LangChain dan LLM dalam sistem QA berbasis Telegram?
- b. Apa saja tantangan utama dalam pengembangan sistem ini?
- c. Bagaimana teknologi ini dapat dioptimalkan di masa depan?

Interpretasi dilakukan dengan memperhatikan konteks setiap penelitian dan kesesuaian temuan dengan tujuan penelitian ini.

4. HASIL DAN PEMBAHASAN

Pada bagian ini, kami akan membahas hasil-hasil penelitian yang telah dilakukan dalam lima paper yang dijadikan sebagai studi literatur dalam pengembangan sistem Question Answering (QA) berbasis chatbot menggunakan Langchain dan Large Language Models (LLM). Teknologi ini semakin menjadi pusat perhatian dalam pengembangan sistem interaktif yang cerdas, khususnya dalam memberikan jawaban berbasis teks yang kompleks. Seiring dengan kemajuan Natural Language Processing (NLP) dan integrasi model bahasa besar, chatbot kini tidak hanya sekadar alat komunikasi, tetapi juga solusi cerdas untuk menjawab pertanyaan dengan tingkat akurasi yang tinggi dalam berbagai domain.

Kelima penelitian yang dibahas berfokus pada berbagai penerapan sistem QA berbasis chatbot Telegram, dengan topik yang melibatkan Fiqh Jual Beli Islam, Tafsir Al-Jalalain, UU Kesehatan, serta Tafsir Al-Azhar, yang masing-masing memanfaatkan kemampuan Langchain dan LLM untuk memberikan jawaban yang sesuai dengan konteks. Masing-masing penelitian menunjukkan bahwa teknologi ini mampu menangani dan memproses informasi yang sangat kompleks dengan lebih efektif, sehingga menghasilkan jawaban yang lebih relevan dan mendalam. Namun, meskipun tingkat akurasi yang tercapai cukup tinggi, tantangan dalam pengembangan sistem ini tetap ada, terutama dalam mengelola teks yang panjang dan penuh interpretasi. Hasil dari penelitian ini tidak hanya mengungkapkan kelebihan dari penggunaan Langchain dan LLM, tetapi juga menyoroti berbagai tantangan dan peluang untuk pengembangan lebih lanjut dalam memaksimalkan potensi teknologi ini.

Dalam pembahasan berikut, kami akan menyajikan analisis terperinci mengenai setiap paper yang telah diulas, mengungkapkan temuan-temuan utama yang ditemukan, serta mengeksplorasi hasil perhitungan yang mendalam mengenai tingkat akurasi dan efektivitas masing-masing sistem. Dengan demikian, pembaca dapat lebih memahami bagaimana Langchain dan LLM berperan dalam meningkatkan kualitas interaksi manusia dengan sistem informasi berbasis chatbot.

Pertama, penelitian (Rizki et al., 2024) melakukan implementasi question answering berbasis chatbot telegram pada Tafsir Al-Jalalain menggunakan Langchain dan LLM, Penelitian ini bertujuan untuk mengevaluasi penerapan sistem Question Answering (QA) berbasis chatbot Telegram, yang menggunakan teknologi Langchain dan Large Language Models (LLM), untuk memberikan jawaban yang akurat dan cepat mengenai Tafsir Al-Jalalain. Sebagai bagian dari evaluasi, kuesioner dibagikan kepada mahasiswa yang menjawab serangkaian pertanyaan terkait dengan kualitas dan keefektifan sistem. Dalam penelitian ini, jumlah pertanyaan yang diuji adalah 10, dengan 7 responden yang berpartisipasi. Untuk mengukur keberhasilan sistem, digunakan skala Likert yang terdiri dari 5 kategori responden, yaitu: Sangat Setuju (5), Setuju (4), Ragu-Ragu (3), Tidak Setuju (2), dan Sangat Tidak Setuju (1). Setiap responden diminta untuk memberikan penilaian terhadap kualitas jawaban yang diberikan oleh chatbot. Skor hasil yang diperoleh dari setiap kategori diuraikan sebagai berikut:

- a. Sangat Setuju (5): 24 responden dengan skor 120
- b. Setuju (4): 38 responden dengan skor 152
- c. Ragu-Ragu (3): 7 responden dengan skor 21
- d. Tidak Setuju (2): 1 responden dengan skor 2
- e. Sangat Tidak Setuju (1): 0 responden dengan skor 0

Total Skor yang diperoleh adalah 295. Selanjutnya, perhitungan dilakukan untuk menentukan nilai X, yang dihitung dengan menggunakan skor tertinggi dan jumlah pertanyaan serta jumlah responden sebagai berikut:

$$\begin{aligned} X &= \text{skor tertinggi} * (\text{jumlah pertanyaan} * \text{jumlah responden}) \\ X &= 5 * (10 * 7) \\ &= 350 \end{aligned}$$

Dengan Total Skor 295 dan nilai X 350, perhitungan persentase User Acceptance Test (UAT) dilakukan menggunakan rumus berikut:

$$\begin{aligned}\text{Persentase} &= \frac{\text{Total Skor}}{X} * 100 \% \\ &= \frac{295}{350} * 100 \% \\ &= 84.29 \%\end{aligned}$$

Berdasarkan hasil perhitungan yang dilakukan, tingkat penerimaan pengguna (user acceptance) terhadap sistem Question Answering berbasis chatbot Telegram mencapai 84.29%. Angka ini menunjukkan bahwa sebagian besar responden memberikan penilaian yang positif terhadap kualitas jawaban yang diberikan oleh sistem. Ini mencerminkan bahwa teknologi Langchain dan LLM yang diterapkan dalam chatbot mampu memberikan jawaban yang relevan dan akurat, serta memenuhi harapan pengguna dalam menjawab pertanyaan terkait dengan Tafsir Al-Jalalain.

Kedua, penelitian (Aji Bayu Permadi et al., 2024) melakukan implementasi question answering system tafsir alazhar menggunakan langchain dan large language model berbasis chatbot telegram. Penelitian ini bertujuan untuk mengembangkan sistem Question Answering (QA) berbasis chatbot Telegram yang mengintegrasikan Langchain dan Large Language Models (LLM) dalam menjawab pertanyaan terkait Tafsir Al-Azhar, sebuah tafsir yang sangat penting dalam studi Islam. Dalam penelitian ini, sistem diuji menggunakan dua metode evaluasi utama: User Acceptance Testing (UAT) dan DeepEval framework, untuk mengukur sejauh mana sistem ini mampu memberikan jawaban yang akurat dan relevan.

Hasil dari User Acceptance Testing (UAT) menunjukkan bahwa sistem ini mampu mencapai akurasi 83.71%. Ini berarti sebagian besar responden merasa puas dengan jawaban yang diberikan oleh chatbot. Responden yang terdiri dari mahasiswa jurusan studi Islam menganggap jawaban yang diberikan relevan dengan pertanyaan yang diajukan, meskipun ada beberapa area yang memerlukan penyempurnaan lebih lanjut. Selain itu, sistem ini juga diuji menggunakan DeepEval framework, yang mengukur kualitas jawaban berdasarkan beberapa metrik, termasuk hallucination score, contextual precision, dan contextual relevancy. Hasilnya menunjukkan bahwa sistem memiliki hallucination score sebesar 41%, yang berarti hampir setengah dari jawaban yang diberikan tidak sepenuhnya sesuai atau relevan dengan teks asli Tafsir Al-Azhar. Hal ini menunjukkan bahwa meskipun sistem memberikan jawaban yang tepat dalam banyak kasus, ada beberapa pertanyaan yang menyebabkan chatbot memberikan informasi yang tidak akurat atau tidak relevan. Namun, sistem menunjukkan kinerja yang cukup baik dalam hal contextual precision, dengan skor 90%, yang menandakan bahwa sebagian besar jawaban yang diberikan sesuai dengan konteks yang diminta oleh pertanyaan. Skor contextual relevancy tercatat 79%, yang menunjukkan bahwa meskipun sebagian besar jawaban relevan, masih ada beberapa yang tidak sesuai dengan konteks atau terlalu umum. Secara keseluruhan, meskipun terdapat beberapa tantangan, terutama terkait dengan hallucination dan contextual relevancy, hasil penelitian ini menunjukkan bahwa sistem QA berbasis chatbot Telegram yang mengintegrasikan Langchain dan LLM dapat memberikan jawaban yang cukup akurat dan relevan terkait dengan Tafsir Al-Azhar. Sistem ini berhasil menunjukkan potensi besar dalam memudahkan pencarian informasi dari teks-teks yang kompleks, seperti tafsir, dan menyediakan jawaban yang lebih mudah diakses oleh pengguna.

Meskipun demikian, penelitian lebih lanjut diperlukan untuk mengurangi hallucination score dan meningkatkan contextual relevancy, terutama dalam menangani pertanyaan-pertanyaan yang lebih mendalam atau kompleks. Dengan pengembangan lebih lanjut, sistem ini memiliki potensi untuk menjadi alat yang sangat berguna dalam bidang studi Islam, mempermudah pengguna dalam mengakses penafsiran dan pemahaman yang lebih baik mengenai ayat-ayat Al-Quran.

Ketiga, penelitian (Herwanza et al., 2024) melakukan Penerapan Langchain Retriever Dengan Model Chat Openai Dalam Pengembangan Sistem Chatbot Hadis Berbasis Telegram. Penelitian ini bertujuan untuk mengembangkan sistem chatbot Hadis berbasis Telegram yang mengintegrasikan Langchain Retriever dan model Chat OpenAI untuk memberikan jawaban yang relevan dan akurat mengenai Hadis Nabi Muhammad (SAW). Untuk mengevaluasi kinerja sistem ini, dilakukan pengujian dengan dua parameter utama: prediksi akurasi dan kualitas jawaban yang dinilai oleh para ahli hadis. Evaluasi akurasi prediksi dilakukan dengan menggunakan parameter `normalize_by = 10`, yang bertujuan untuk menstandarisasi skor prediksi. Berdasarkan pengujian yang ditampilkan pada Tabel 2, sistem berhasil memberikan penilaian yang tepat dalam 9 dari 10 kasus. Pada satu kasus, prediksi menerima evaluasi skor 0.7, yang menunjukkan bahwa jawaban yang diberikan masih sejalan dengan referensi

namun mengandung kesalahan kecil atau kelalaian. Sementara itu, pada 9 kasus lainnya, sistem memberikan skor 10, yang berarti jawaban yang diberikan benar-benar akurat dan sepenuhnya sesuai dengan referensi. Hasil evaluasi ini menghasilkan nilai rata-rata evaluasi keseluruhan sebesar 0.97, yang menunjukkan bahwa kinerja sistem berada dalam rentang nilai yang sangat baik, menunjukkan prediksi yang sangat akurat pada sebagian besar kasus. Evaluasi kualitas jawaban dilakukan dengan melibatkan 5 ahli hadis yang memberikan penilaian terhadap 10 pertanyaan serta jawaban yang dihasilkan oleh sistem. Penilaian menggunakan Skala Likert, dengan kategori penilaian sebagai berikut:

- a. Sangat Setuju (SS) bernilai 5
- b. Setuju (S) bernilai 4
- c. Netral (N) bernilai 3
- d. Tidak Setuju (TS) bernilai 2
- e. Sangat Tidak Setuju (STS) bernilai 1

Hasil penilaian dari ahli hadis menunjukkan distribusi skor sebagai berikut:

- a. Sangat Setuju (5): 28 jawaban
- b. Setuju (4): 20 jawaban
- c. Netral (3): 1 jawaban
- d. Tidak Setuju (2): 1 jawaban
- e. Sangat Tidak Setuju (1): 0 jawaban

Dengan data ini, perhitungan persentase akurasi dilakukan menggunakan rumus sebagai berikut:

$$\begin{aligned}\text{Persentase Akurasi} &= \frac{\text{Total score}}{N} \times 100 \% \\ &= \frac{((5 \times 28) + (4 \times 20) + (3 \times 1) + (2 \times 1) + (1 \times 0))}{5 \times 10 \times 5} \times 100 \% \\ &= 90 \%\end{aligned}$$

Berdasarkan perhitungan tersebut, persentase akurasi yang diperoleh adalah 90%, yang menunjukkan bahwa sebagian besar responden sangat setuju dengan jawaban yang dihasilkan oleh sistem. Dalam Skala Likert, hasil ini berada pada rentang indeks 80% - 100%, yang berarti para ahli hadis sangat setuju terhadap kualitas jawaban yang dihasilkan oleh sistem chatbot. Evaluasi ini menunjukkan bahwa 90% akurasi yang tercatat berada dalam kategori Sangat Setuju menurut Skala Likert yang digunakan dalam penelitian ini. Skor ini mencerminkan penerimaan yang sangat positif dari para ahli terhadap kemampuan sistem dalam memberikan jawaban yang tepat, relevan, dan sesuai dengan referensi Hadis.

Keempat, penelitian (Nurhapiza et al., 2024) membahas Penerapan Chatbot pada Aplikasi Web Tanya Jawab Tentang Fiqih Jual Beli Islam Menggunakan LangChain. Pengujian terhadap aplikasi web tanya jawab berbasis chatbot yang dikembangkan untuk fiqih jual beli Islam dilakukan oleh dua kelompok pengguna yang berbeda: mahasiswa dari UIN Sultan Syarif Kasim Riau dan seorang ustaz yang memiliki pemahaman mendalam tentang fiqih muamalah. Pengujian ini bertujuan untuk menilai sejauh mana aplikasi mampu memberikan jawaban yang akurat dan relevan mengenai fiqih jual beli Islam. Dalam pengujian ini, sepuluh pertanyaan diajukan kepada sistem chatbot sebagai bagian dari evaluasi. Pertanyaan-pertanyaan tersebut dipilih untuk mencakup berbagai aspek fiqih jual beli Islam, dengan tujuan menguji kemampuan chatbot dalam memberikan jawaban yang sesuai dengan ajaran Islam. Setelah menguji aplikasi, hasilnya menunjukkan bahwa aplikasi web tersebut memberikan jawaban yang sangat sesuai dengan harapan. Evaluasi menunjukkan tingkat kesesuaian jawaban sebesar 88,8%, yang termasuk dalam kategori "sangat sesuai". Ini berarti bahwa mayoritas jawaban yang diberikan oleh chatbot benar dan sesuai dengan prinsip-prinsip fiqih jual beli Islam yang diajarkan dalam sumber-sumber fiqih yang diacu.

Secara keseluruhan, hasil pengujian ini mengindikasikan bahwa aplikasi web tanya jawab ini memiliki tingkat akurasi yang tinggi dan dapat diandalkan sebagai alat bantu bagi umat Islam yang ingin memahami lebih dalam mengenai fiqih jual beli Islam.

Kelima, penelitian (Lubis et al., 2024) membahas Question Answering System pada Chatbot Telegram Menggunakan Large Language Models (LLM) dan Langchain pada UU Kesehatan. Penelitian ini berhasil mengembangkan sistem Question Answering berbasis chatbot yang terintegrasi dengan Telegram, menggunakan

teknologi LangChain dan Large Language Models (LLM). Sistem ini dirancang untuk membantu pengguna mendapatkan jawaban atas pertanyaan terkait Undang-Undang No. 17 Tahun 2023 tentang Kesehatan. Namun, karena UU Kesehatan awalnya dalam format PDF, file tersebut perlu diubah menjadi format TXT agar dapat diproses dengan baik oleh sistem chatbot. Proses pengolahan data dilakukan dengan chunking dan embeddings, yang memungkinkan sistem untuk memahami dan mencari informasi yang relevan dengan pertanyaan yang diajukan oleh pengguna. Setelah proses ini, data UU No. 17 Tahun 2023 digunakan sebagai basis pengetahuan bagi chatbot untuk memberikan jawaban yang tepat. Meski demikian, ada beberapa keterbatasan pada sistem ini. Chatbot hanya dapat menjawab pertanyaan langsung dan tidak dapat merespons dengan baik terhadap kata sapaan, ucapan terima kasih, atau permintaan maaf. Hal ini tentu menjadi kekurangan dalam hal interaksi yang lebih natural antara pengguna dan sistem, yang seharusnya bisa menjadi lebih interaktif.

Dalam pengujian sistem, dua metrik utama digunakan untuk mengukur kualitas jawaban chatbot, yaitu BERTScore dan ROUGE Score. Hasil pengujian dengan BERTScore menunjukkan nilai yang cukup baik dengan:

- a. Precision: 76%
- b. Recall: 80%
- c. F1-Score: 78%

Sedangkan untuk ROUGE Score, hasilnya adalah:

- a. ROUGE-1: Precision 60%, Recall 45%, F1-Score 50%
- b. ROUGE-2: Precision 34%, Recall 25%, F1-Score 28%
- c. ROUGE-L: Precision 45%, Recall 34%, F1-Score 38%

Hasil pengujian ini menunjukkan bahwa meskipun chatbot ini mampu memberikan jawaban yang cukup relevan dan sesuai dengan data yang ada, masih ada ruang untuk perbaikan, terutama dalam meningkatkan tumpang tindih antara jawaban yang dihasilkan dengan referensi data asli, yang tercermin dalam nilai ROUGE Score yang lebih rendah.

5. KESIMPULAN

Penelitian ini telah mengkaji penerapan teknologi Langchain dan Large Language Models (LLM) dalam pengembangan sistem Question Answering (QA) berbasis chatbot Telegram. Sistem ini dirancang untuk memberikan jawaban yang akurat dan relevan atas berbagai pertanyaan, yang mencakup topik-topik kompleks seperti Fiqh Jual Beli Islam, Tafsir Al-Jalalain, Tafsir Al-Azhar, dan Undang-Undang Kesehatan. Teknologi Langchain dan LLM terbukti sangat efektif dalam mengelola dan memproses informasi yang sangat kompleks, memberikan jawaban yang lebih tepat dan relevan, serta meningkatkan kualitas interaksi antara pengguna dan sistem informasi berbasis chatbot. Beberapa penelitian yang dijadikan acuan dalam pengembangan sistem ini menunjukkan hasil yang sangat positif. Penelitian tentang Tafsir Al-Jalalain mencatatkan tingkat penerimaan pengguna sebesar 84.29%, yang mengindikasikan bahwa mayoritas responden puas dengan kualitas jawaban yang diberikan oleh chatbot. Begitu juga dengan Tafsir Al-Azhar, di mana chatbot berhasil memberikan jawaban yang relevan dengan tingkat akurasi 83.71%, meskipun terdapat beberapa tantangan terkait kesalahan dalam konteks dan relevansi jawaban. Sistem chatbot Hadis menunjukkan hasil yang sangat baik dengan tingkat akurasi 90%, yang membuktikan bahwa Langchain dan LLM dapat diandalkan dalam menjawab pertanyaan-pertanyaan keagamaan yang kompleks dan mendalam. Selain itu, pada aplikasi untuk Fiqh Jual Beli Islam, chatbot memberikan jawaban yang sangat sesuai dengan 88.8% tingkat kesesuaian jawaban, menunjukkan kehandalan sistem dalam menyediakan informasi yang tepat terkait dengan prinsip-prinsip fiqih. Pengujian pada UU Kesehatan menunjukkan hasil yang memadai dengan nilai Precision 76%, Recall 80%, dan F1-Score 78%. Meskipun hasilnya baik, ada ruang untuk meningkatkan kualitas jawaban, terutama dalam meningkatkan kesesuaian antara jawaban chatbot dengan referensi data asli. Meskipun hasilnya sudah cukup memuaskan, tantangan utama dalam pengembangan sistem ini adalah bagaimana mengelola teks panjang yang seringkali penuh dengan interpretasi. Hal ini dapat menyebabkan jawaban yang diberikan oleh chatbot tidak selalu sepenuhnya akurat atau relevan. Oleh karena itu, perlu adanya upaya lebih untuk meningkatkan pengelolaan konteks dan akurasi sistem, terutama dalam menghadapi pertanyaan-pertanyaan yang lebih mendalam atau kompleks.

Secara keseluruhan, meskipun masih ada beberapa kekurangan yang perlu diperbaiki, penerapan teknologi Langchain dan LLM dalam chatbot berbasis Telegram menunjukkan potensi yang besar. Dengan pengembangan lebih lanjut, sistem ini dapat lebih interaktif, lebih mampu memberikan jawaban yang lebih relevan, dan dapat

meningkatkan pengalaman pengguna dalam mencari informasi. Ke depan, pengembangan sistem ini diharapkan dapat mengatasi tantangan-tantangan yang ada dan memberikan solusi yang lebih efektif dalam menjawab pertanyaan berbasis teks yang kompleks.

DAFTAR PUSTAKA

- Afriani, E., H, N. S., Fikry, M., & Affandes, M. (2024). *Aplikasi Tanya Jawab Tentang Fiqih Bersuci Berbasis Web*. 6(2), 380–390.
- Aji Bayu Permadi, Nazruddin Safaat H, Lestari Handayani, & Yusra. (2024). Implementasi Question Answering System Tafsir Al-Azhar Menggunakan Langchain Dan Large Language Model Berbasis Chatbot Telegram. *Jurnal Teknoif Teknik Informatika Institut Teknologi Padang*, 12(1), 62–69. <https://doi.org/10.21063/jtif.2024.v12.1.62-69>
- Alsafari, B., Atwell, E., Walker, A., & Callaghan, M. (2024). Towards effective teaching assistants: From intent-based chatbots to LLM-powered teaching assistants. *Natural Language Processing Journal*, 8(April), 100101. <https://doi.org/10.1016/j.nlp.2024.100101>
- Chen, Y., Liebau, U., Guruprasad, S. M., Trofimenko, I., & Minke, C. (2024). *Advancing Life Cycle Assessment of Sustainable Green Hydrogen Production Using Domain-Specific Fine-Tuning by Large Language Models Augmentation*. 2494–2514.
- Elvino Dwi Saputra, N. S. H. J. Y. (2024). Question Answering Al-Qur'an Menggunakan GPT 3,5 Chatbot Telegram. *Jutisi: Jurnal Ilmiah Teknik Informatika Dan Sistem Informasi*, 13(No.1), 550–563.
- Fauzan, M. F., Imanda, R., & Hasbi, M. A. (2024). *Designing an Chatbot with NLP Technology in a Website-Based New Student Admission Information System*. 8(2), 358–366.
- Firdaus, D., Sumardi, I., & Kulsum, Y. (2024). *Integrating Retrieval-Augmented Generation with Large Language Model Mistral 7b for Indonesian Medical Herb*. 9(3), 230–243.
- Herwanza, N. A. M., Harahap, N. S., Yanto, F., & Insani, F. (2024). Penerapan Langchain Retriever dengan Model Chat Openai dalam Pengembangan Sistem Chatbot Hadis Berbasis Telegram. *JTIM: Jurnal Teknologi Informasi Dan Multimedia*, 6(1), 70–83. <https://doi.org/10.35746/jtim.v6i1.514>
- Lubis, A. T. U. B., Harahap, N. S., Agustian, S., Irsyad, M., & Afrianty, I. (2024). Question Answering System pada Chatbot Telegram Menggunakan Large Language Models (LLM) dan Langchain (Studi Kasus UU Kesehatan). *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(3), 955–964. <https://doi.org/10.57152/malcom.v4i3.1378>
- Mohamad Alvian Fazlur Febrian Dhika, S. F. K. (2024). 2) 1,2. 10(1), 25–32.
- Nurhapiza, Harahap, N. S., Fikry, M., & Affandes, M. (2024). Penerapan Chatbot pada Aplikasi Web Tanya Jawab Tentang Fiqih. *Journal of Computer System and Informatics (JoSYC)*, 5(3), 548–557. <https://doi.org/10.47065/josyc.v5i3.5148>
- Pujiono, I., Agtyaputra, I. M., & Ruldeviyani, Y. (2024). Implementing Retrieval-Augmented Generation and Vector Databases for Chatbots in Public Services Agencies Context. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, 10(1), 216–223. <https://doi.org/10.33480/jitk.v10i1.5572>
- Rizki, F., Sutiyo, A., Harahap, N. S., Agustian, S., & Candra, R. M. (2024). KLIK: Kajian Ilmiah Informatika dan Komputer Implementasi Question Answering Berbasis Chatbot Telegram Pada Tafsir Al-Jalalain Menggunakan Langchain dan LLM. *Media Online*, 4(5), 2464–2472. <https://doi.org/10.30865/klik.v4i5.1784>
- Saputra, E. D., Harahap, N. S., Jasril, J., & Yusra, Y. (2024). Question Answering Al-Qur'an Menggunakan Generative Pre-Trained Transformer 3.5 Berbasis Chatbot Telegram. *Jutisi: Jurnal Ilmiah Teknik Informatika Dan Sistem Informasi*, 13(1), 550. <https://doi.org/10.35889/jutisi.v13i1.1879>
- Vidivelli, S., Ramachandran, M., & Dharunbalaji, A. (2024). Efficiency-Driven Custom Chatbot Development: Unleashing LangChain, RAG, and Performance-Optimized LLM Fusion. *Computers, Materials and Continua*, 80(2), 2423–2442. <https://doi.org/10.32604/cmc.2024.054360>