



Question Answering System on Telegram Chatbot Using Large Language Models (LLM) and Langchain (Case Study: Health Law)

Question Answering System pada Chatbot Telegram Menggunakan Large Language Models (LLM) dan Langchain (Studi Kasus UU Kesehatan)

**Anggun Tri Utami Br. Lubis^{1*}, Nazruddin Safaat Harahap², Surya Agustian³,
Muhammad Irsyad⁴, Iis Afrianty⁵**

^{1,2,3,4,5}Program Studi Teknik Informatika, Fakultas Sains dan Teknologi,
Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia

E-Mail: ¹angguntriutamibrubis@gmail.com, ²nazruddin.safaat@uin-suska.ac.id,
³surya.agustian@uin-suska.ac.id, ⁴irsyadtech@uin-suska.ac.id, ⁵iis.afrianty@uin-suska.ac.id

Received Apr 25th 2024; Revised May 16th 2024; Accepted May 21th 2024
Corresponding Author: Anggun Tri Utami Br. Lubis

Abstract

In the health sector, the regulations implemented are known as health laws, which aim to protect the interests of patients and improve the standards of medical practice. In 2023, Indonesia will implement Law No. 17 of 2023 concerning Health, covering patient rights, service standards and community participation. This Omnibus Law is expected to solve health problems and protect service providers. This research aims to develop a chatbot-based Question Answering System (QAS) that is integrated with Telegram. The methods used are Langchain and Large Language Models (LLM). Langchain is used to facilitate chatbot building, while LLM is a type of AI model that uses a machine learning approach to generate text similar to human language. The data source used as a knowledge base is Law No. 17 of 2023 concerning health. The chatbot that was built has succeeded in providing answers to users with test results using BERTScore getting average precision, recall and f1-score values of 76%, 80%, 78% respectively. Meanwhile for ROUGE-1 it is 60%, 45%, 50%, for ROUGE-2 it is 34%, 25%, 28%, and for ROUGE-L it is 45%, 34%, 38%.

Keyword: Chatbot, Health Law Legal Document, Langchain, LLM, Question Answering System

Abstrak

Di bidang kesehatan, peraturan yang diterapkan dikenal sebagai hukum kesehatan, yang bertujuan untuk melindungi kepentingan pasien dan meningkatkan standar praktik medis. Pada tahun 2023, Indonesia menerapkan UU No 17 Tahun 2023 tentang Kesehatan, mencakup hak pasien, standar layanan, dan partisipasi masyarakat. Omnibus Law ini diharapkan menyelesaikan masalah kesehatan dan melindungi penyedia layanan. Penelitian ini bertujuan untuk mengembangkan *Question Answering System (QAS)* berbasis *chatbot* yang terintegrasi dengan Telegram. Metode yang digunakan adalah *Langchain* dan *Large Language Models (LLM)*. *Langchain* digunakan untuk memfasilitasi pembangunan *chatbot*, sementara LLM adalah jenis model AI yang menggunakan pendekatan pembelajaran mesin untuk menghasilkan teks yang serupa dengan bahasa manusia. Sumber data yang digunakan sebagai basis pengetahuan adalah UU No 17 tahun 2023 tentang kesehatan. *Chatbot* yang dibangun telah berhasil memberikan jawaban kepada pengguna dengan hasil pengujian menggunakan BERTScore mendapatkan rata-rata nilai *precision*, *recall*, *f1-score* masing-masing sebesar 76%, 80%, 78%. Sedangkan untuk ROUGE-1 sebesar 60%, 45%, 50%, untuk ROUGE-2 sebesar 34%, 25%, 28%, dan untuk ROUGE-L sebesar 45%, 34%, 38%.

Kata Kunci: *Chatbot, Langchain, LLM, Sistem Tanya Jawab, Undang-undang Kesehatan*

1. PENDAHULUAN

Pandangan masyarakat yang beragam terhadap hukum menciptakan persepsi tentang fungsi hukum sebagai alat pengatur perilaku. Hukum kesehatan adalah aturan tertulis mengenai hubungan pemberi layanan kesehatan dengan masyarakat, bertujuan melindungi kepentingan pasien dan meningkatkan standar praktik

medis serta kualitas pelayanan kesehatan[1]. Secara umum, semua individu memerlukan layanan hukum dan kesehatan. Peran kesehatan adalah untuk menjaga dan menyembuhkan, sementara hukum berfungsi untuk melindungi individu agar terhindar dari kejadian yang tidak diinginkan[2].

Pada tahun 2023, pemerintah Indonesia mengesahkan Undang-Undang Nomor 17 Tahun 2023 tentang Kesehatan sebagai dasar hukum untuk mengatur layanan kesehatan di negara tersebut. Undang-undang tersebut mencakup berbagai aspek, termasuk hak dan kewajiban pasien, standar pelayanan kesehatan, dan partisipasi masyarakat dalam mencapai tingkat kesehatan yang optimal[3]. Diharapkan dengan diberlakukannya Omnibus Law UU No 17 Tahun 2023 tentang Kesehatan ini, akan memberikan solusi bagi sejumlah masalah di bidang kesehatan dan memberikan perlindungan hukum bagi para penyedia layanan kesehatan[4].

Oleh karena itu diperlukan pengetahuan masyarakat tentang aturan kesehatan baru yang berlaku di Indonesia agar masyarakat mengetahui hak dan kewajiban yang diperoleh dari layanan kesehatan. maka diperlukan sebuah sistem untuk mempermudah masyarakat mencari informasi dan interaktif, yaitu memberikan jawaban yang sesuai dengan permintaan pengguna. Dalam upaya untuk mempermudah pencarian informasi, dibangun sebuah sistem tanya jawab atau *Question Answering System (QAS)*. Pada sistem QAS, pengguna dapat bertanya dengan cepat dan singkat. QAS juga sering dilengkapi dengan kalimat pendukung sebagai referensi jawaban yang diberikan[5]. QAS merupakan pengembangan dari *Natural Language Processing (NLP)*, yang merupakan sebuah teknologi yang memfasilitasi interaksi antara manusia dan komputer melalui bahasa manusia[6]. Berbeda dengan mesin pencari, QAS memberikan jawaban langsung terhadap pertanyaan tanpa perlu membuka halaman web terkait. Pengguna dapat berkomunikasi dengan sistem seperti dalam percakapan. Hal ini membuatnya lebih intuitif dan mudah digunakan[7].

Beberapa penelitian sebelumnya oleh Keivalya Pandya dan kawan-kawan [8] yang membangun sebuah *chatbot* berbasis web tentang *Birla Vishakarma Mahavidyalaya (BVM) Engineering College* dengan mengintegrasikan LLM oleh *Langchain*, *chatbot* ini digunakan sebagai *customer service* dan mampu memberikan dukungan yang efisien, personal, dan responsif kepada pengguna. Selain itu pada penelitian yang dilakukan oleh Joseph Benjamin Ilagan dan kawan-kawan [9], mengembangkan *chatbot* berbasis Telegram menggunakan LLM dan GPT-3.5, *Chatbot* ini digunakan untuk membantu mahasiswa yang ingin belajar dan menyempurnakan ide startup mereka. Penelitian oleh Pranav Nataraj Devaraj dan kawan-kawan [10] membangun sebuah *chatbot* berbasis Android seputar hukum di India. *Chatbot* ini juga menggunakan *Langchain* yang memanfaatkan LLM dan memproses pertanyaan dalam konteks dokumen yang relevan dan memberikan respon yang sesuai.

Penelitian ini membangun sebuah *chatbot* yang nantinya akan diintegrasikan dengan Telegram. Algoritma atau metode yang digunakan adalah *Large Language Models (LLM)* dan *Langchain*. Salah satu fitur utama dari LLM adalah pembelajaran kontekstual, yaitu model dilatih untuk menghasilkan teks berdasarkan konteks atau prompt yang diberikan. Hal ini memungkinkan LLM untuk menghasilkan tanggapan yang lebih koheren dan relevan, sehingga cocok untuk aplikasi interaktif dan percakapan[11] Sehingga LLM nantinya akan digunakan untuk memberikan jawaban yang sesuai atas pertanyaan yang diajukan[12]. Seperti ChatGPT yang merupakan sebuah LLM berbasis kecerdasan buatan, telah memberikan kontribusi positif dalam berbagai bidang[13].

Langchain adalah kerangka kerja untuk mengembangkan aplikasi dengan memanfaatkan LLM, dan tujuannya adalah memberikan kemudahan bagi pengembang untuk menggunakan sumber data lain dan berinteraksi dengan aplikasi lainnya[14], [15]. *Langchain* mengelompokkan teks besar menjadi segmen yang lebih kecil dan lebih mudah dikelola. Selain itu *Langchain* hanya bisa digunakan untuk mengelola data berupa teks, *Langchain* tidak mendukung data berupa gambar.

Sumber pengetahuan *chatbot* menggunakan data UU No 17 Tahun 2023 tentang Kesehatan yang akan dijadikan sebagai basis pengetahuan *chatbot* nantinya. Dan pengujian sistem menggunakan BERTScore dan Rouge Score untuk melihat kualitas jawaban yang diberikan sistem.

2. METODOLOGI PENELITIAN

Tahapan penelitian yang dilakukan dapat dilihat pada gambar 1.

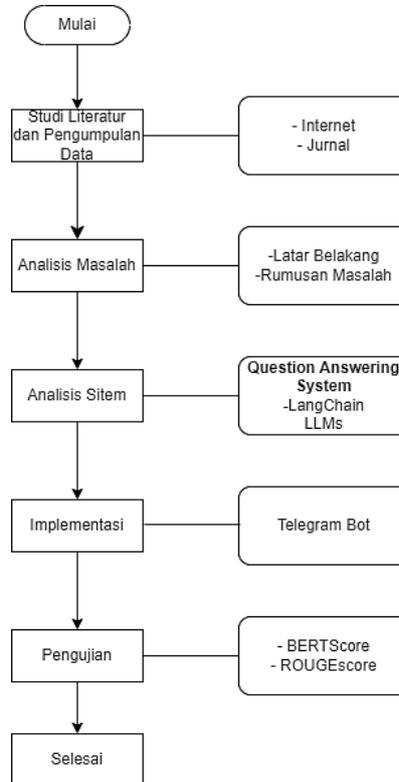
2.1. Studi Literatur dan Pengumpulan Data

Studi literatur yang juga dikenal sebagai tinjauan literatur, merupakan bentuk penelitian yang fokus pada penyelidikan dan eksplorasi literatur yang telah ada dalam suatu bidang pengetahuan. Pada tahap ini peneliti melakukan studi literatur untuk membaca jurnal atau sumber terkait penelitian melalui Google Scholar, IEEE Explorer dan repository kampus. Pengumpulan data ini dilakukan guna membantu peneliti memperoleh informasi guna mendukung penelitian untuk dijadikan referensi penelitian. Sumber data yang digunakan adalah data UU No 17 Tahun 2023 tentang Kesehatan dari laman peraturan.bpk.go.id yang berbentuk file PDF.

2.2. Analisis Masalah

Menganalisis suatu permasalahan adalah suatu langkah untuk memahami situasi atau keadaan yang dianggap sebagai permasalahan dengan tujuan untuk mengenali akar penyebab dari permasalahan tersebut dan

memahami dampaknya, sehingga dapat menghasilkan berbagai alternatif solusi yang efektif. Menggunakan data yang telah terkumpul, menyiapkan latar belakang dan merumuskan permasalahan dalam penelitian. Pada tahap ini, juga mengkonversi data yang semula dalam format PDF menjadi format TXT. Ini dilakukan karena *chatbot* tidak dapat merespons pertanyaan menggunakan file PDF. Selain itu juga dilakukan penghapusan bagian *header* dan *footer* dalam file serta memperbaiki beberapa kesalahan penulisan agar *chatbot* tidak keliru dalam mengambil informasi ketika memberikan jawaban.

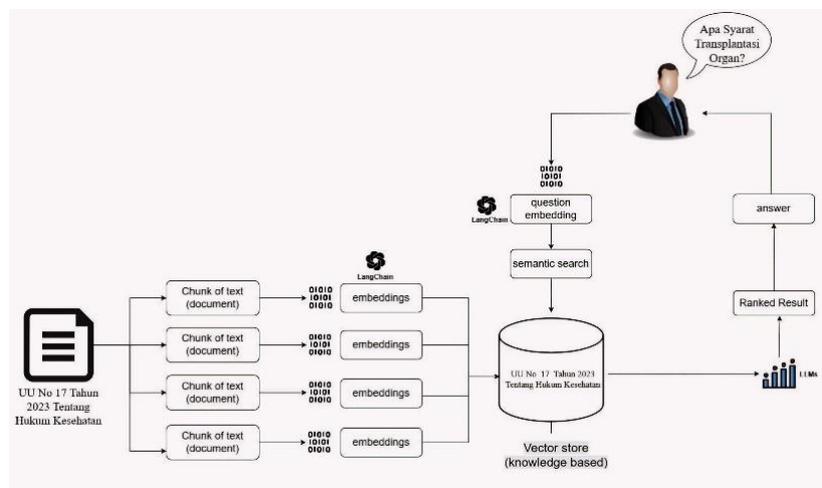


Gambar 1. Tahapan Penelitian

2.3. Analisis Sistem

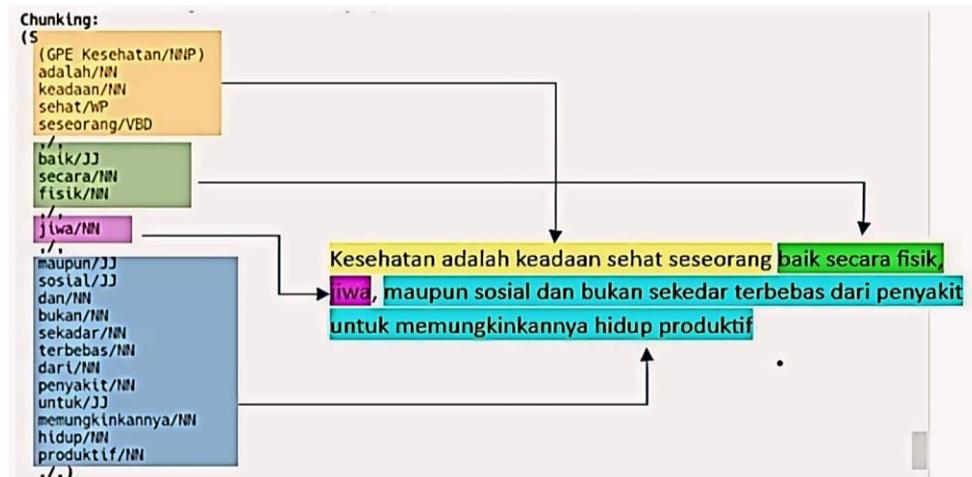
Analisis sistem adalah proses yang sistematis untuk memahami, mendokumentasikan, dan mengevaluasi struktur, fungsi, dan kinerja sistem yang ada atau yang direncanakan. Analisis sistem bertujuan untuk meningkatkan efisiensi, produktivitas, dan kualitas sistem, serta memastikan bahwa sistem tersebut sesuai dengan tujuan. Dalam hal ini memperhatikan *Langchain* dan LLM dalam membangun *Question Answering System*.

Perancangan *Question Answering System* dapat dilihat pada gambar 2.



Gambar 2. Alur Perancangan *Question Answering System*

Dalam penelitian ini menggunakan UU No 17 Tahun 2023 tentang Kesehatan sebagai *knowledge based* sistem. *Knowledge based* yang ditanamkan dalam *chatbot* dirancang untuk mencakup informasi yang paling diperlukan pengguna. Tahap awal dalam perancangan QAS ini, data undang-undang akan melalui tahap *preprocessing* data, yaitu tahap *chunking* dan *embeddings*. Proses *chunking* adalah teknik yang digunakan untuk memecah informasi besar menjadi bagian-bagian lebih kecil (token) agar lebih mudah dikelola. Berikut adalah bentuk kalimat setelah melalui proses *chunking* seperti Gambar 3.



Gambar 3. Hasil Chunking

Dapat dilihat pada Gambar 3, bahwasannya setiap kata dalam kalimat tersebut dipisahkan sesuai dengan jenis kata sebagai penandaan teks. Pada proses *chunking*, tag-tag seperti NN, JJ, WP, VBD, GPE, dan NNP digunakan untuk mengidentifikasi bagian-bagian penting dari teks, seperti frasa kata benda, frasa kata sifat, atau frasa kata kerja, yang dapat membantu dalam pemahaman struktur dan makna teks secara lebih baik.

Token-token tersebut akan melewati proses *embeddings* yaitu proses mengubah setiap kata atau frasa dari kosakata dipetakan ke dalam bentuk vektor hingga terjalin hubungan semantik antara kata-kata[16]. Representasi vektor ini membantu mesin untuk menangkap makna, struktur, dan konteks dari teks. Memanfaatkan OpenAI untuk melakukan proses *embeddings*. Selanjutnya vektor-vektor tersebut akan disimpan ke dalam *vector database* untuk menyimpan data vektor tersebut menggunakan FAISS sebagai *vectorstore* sistem yang memungkinkan melakukan pencarian secara efisien berdasarkan kesamaan vektor.

Pada tahap ini, data undang-undang telah menjadi data yang siap dipakai. Saat pengguna mengajukan pertanyaan kepada sistem, pertanyaan tersebut akan melalui proses *chunk* dan *embeddings*. Setelah melewati proses tersebut, saat pengguna mengajukan pertanyaan, sistem akan mencari kemiripan kata atau teks (*semantic search*) pada data yang berisi jawaban atas pertanyaan yang diajukan yang memiliki arti serupa pada database dan menampilkan beberapa alternatif jawaban. Selanjutnya dengan model *auto-generatif* dari LLM akan mengolah setiap jawaban dan melakukan pembobotan (*scoring*) pada setiap kemungkinan jawaban yang dihasilkan. Selanjutnya model akan memilih jawaban terbaik berdasarkan skor yang diberikan kepada setiap kemungkinan jawaban. Jawaban terbaik kemudian akan disajikan kepada pengguna sebagai respons dari *chatbot*. Pasangan pertanyaan dan jawaban ini nantinya akan disimpan dalam memori *Langchain* guna mengatasi jika ada pertanyaan yang sama diajukan kembali.

2.4. Implementasi

Implementasi adalah tahap dalam siklus pengembangan atau perubahan yaitu rencana atau solusi yang telah dirancang dijalankan dalam praktek. Ini melibatkan penerapan ide, strategi, atau sistem yang telah dirancang dalam lingkungan nyata atau operasional. Pada tahap ini menggunakan Telegram sebagai wadah pembangunan *chatbot*.

1. Pengembangan Chatbot

Pengembangan *chatbot* dilakukan menggunakan Telegram. Telegram menawarkan sejumlah keuntungan dan fitur yang membuatnya menjadi platform yang menarik untuk pengembangan dan implementasi *chatbot*. Selain itu Telegram menyediakan API terbuka yang memungkinkan membuat dan mengintegrasikan bot dengan mudah. Untuk mengintegrasikan bot dengan Telegram dibutuhkan token yang bisa didapatkan melalui BotFather. BotFather adalah bot resmi yang disediakan oleh Telegram untuk membuat dan mengelola bot Telegram.

2. Pengembangan API

Mendukung pengembangan chatbot, membutuhkan Application Programming Interface (API) yang merupakan seperangkat metode perangkat lunak yang memungkinkan dua aplikasi berkomunikasi satu sama lain melalui seperangkat aturan yang ditentukan oleh penyedia[18]. Menggunakan OpenAI untuk mengakses layanan API. API key ini nantinya digunakan untuk mengautentikasi permintaan ke server OpenAI untuk memanggil model bahasa LLM, yaitu Generative Pre-trained Transformer (GPT). Pada penelitian ini menggunakan GPT-3.5 Turbo.

2.5. Pengujian

Pengujian ini bertujuan untuk menentukan kelayakan penggunaan sistem yang telah dikembangkan. Pengujian ini dilakukan untuk melihat kecocokan antara jawaban sistem dengan referensi data yang ada menggunakan BERTScore dan ROUGE *score* untuk mengukur kualitas jawaban *chatbot*. Tujuannya adalah untuk menentukan kelayakan penggunaan sistem yang telah dibuat.

1. BERTScore

BERTScore adalah metrik evaluasi otomatis untuk generasi teks yang menghitung skor kemiripan untuk setiap token dalam kalimat prediksi dengan setiap token dalam kalimat referensi. BERTScore menghitung kesamaan kosinus antara *embedding* token dari dua kalimat untuk mengukur kesamaan di antara keduanya[17]. BERTScore menghitung *precision*, *recall*, dan *f1-score* yang didefinisikan sebagai kesamaan kosinus antara *embedding* kontekstual yang dinormalisasi[18]. Metrik ini menggunakan model bahasa BERT (*Bidirectional Encoder Representations from Transformers*) untuk mengekstrak representasi teks yang mendalam dan kemudian menghitung kesamaan antara representasi teks tersebut untuk mengevaluasi kalimat kandidat dengan mengacu pada kalimat referensi[19], [20]. BERTScore memberikan skor antara 0 dan 1, yaitu nilai 1 menunjukkan kesesuaian sempurna antara dua teks, sedangkan nilai 0 menunjukkan ketidaksesuaian total. Metrik ini telah terbukti sangat efektif dalam menangkap kesesuaian semantik antara teks dalam berbagai tugas evaluasi, seperti pemindaian mesin terjemahan, evaluasi ringkasan, dan pengevaluasi kualitas generasi bahasa alami. BERT memetakan setiap kata dalam sebuah kalimat dengan semua kata lainnya dalam kalimat tersebut untuk menginterpretasikan makna kontekstual dari masing-masing kata[21]. Semakin tinggi skor BERTScore, semakin sesuai kesesuaian semantik konten antara dua entitas tersebut[22]. Berikut adalah persamaan untuk menghitung BERTScore[17]:

a. Precision

Precision diperoleh dengan cara mencocokkan (*pairwise cosine similarity*) token-token dalam jawaban sistem (kandidat) dengan token-token referensi. Berikut adalah rumus perhitungan *precision*, ditunjukkan pada persamaan 1.

$$P_{BERT} = \frac{1}{|\hat{x}|} \sum_{\hat{x}_j \in \hat{x}} \max x_i^\top \hat{x}_j \quad (1)$$

b. Recall

Recall diperoleh dengan cara mencocokkan (*pairwise cosine similarity*) token-token dalam kalimat referensi dengan token-token kalimat kandidat. Berikut adalah rumus perhitungan *recall*, ditunjukkan pada persamaan 2.

$$R_{BERT} = \frac{1}{|x|} \sum_{x_i \in x} \max x_i^\top \hat{x}_j \quad (2)$$

c. F1-Score

F1-score merupakan kombinasi *precision* dan *recall*. Berikut adalah rumus perhitungan *f1-score*, ditunjukkan pada persamaan 3.

$$F_{BERT} = 2 \frac{P_{BERT} \cdot R_{BERT}}{P_{BERT} + R_{BERT}} \quad (3)$$

2. ROUGE Score

Recall-Oriented Undersity for Gisting Evaluation (ROUGE) adalah kumpulan metrik yang dirancang untuk mengevaluasi ringkasan otomatis dari teks-teks panjang yang terdiri dari beberapa kalimat atau

paragraf. Meskipun awalnya ditujukan untuk mengevaluasi ringkasan tunggal atau multi-dokumen, ROUGE juga telah digunakan untuk mengevaluasi tugas-tugas generasi teks pendek, seperti terjemahan mesin, penjelasan gambar, dan pembangkitan pertanyaan[23]. ROUGE bekerja dengan mengukur tumpang tindih n-gram antara teks yang dihasilkan dan teks asli (ditulis oleh manusia). Skor yang lebih besar mengacu pada kemiripan teks yang lebih tinggi antara kedua jawaban[24].

Dalam penelitian ini, ROUGE-N dan ROUGE-L akan digunakan sebagai metrik untuk mengevaluasi kinerja sistem. ROUGE-N membandingkan jumlah n-gram yang cocok antara ringkasan sistem dan ringkasan manual, dengan variasi nilai n mulai dari 1 hingga 4. Misalnya, saat $n=1$, ROUGE-N menghitung unigram, dan saat $n=2$, menghitung bigram. Sedangkan ROUGE-L mencocokkan susunan kata terpanjang yang sama antara kedua ringkasan, dikenal sebagai longest common subsequence (LCS). Persamaan untuk menghitung ROUGE-N dan ROUGE-L dapat ditemukan di referensi yang bersangkutan[25], [26]. Berikut adalah persamaan menghitung ROUGE-N dan ROUGE-L[26]:

a. *Precision*

Precision adalah metode untuk mengukur jumlah yang diprediksi relevan dengan cara menghitung jumlah kata yang sama, baik unigram, bigram, atau LCS dibagi keseluruhan kata pada ringkasan sistem.

$$ROUGE - 1 = \frac{\text{jumlah unigram kata sama}}{\text{keseluruhan kata di ringkasan sistem}} \quad (4)$$

$$ROUGE - 2 = \frac{\text{jumlah bigram kata sama}}{\text{keseluruhan kata di ringkasan sistem}} \quad (5)$$

$$ROUGE - L = \frac{LCS(\text{Longest Common Subsequent})}{\text{keseluruhan kata di ringkasan sistem}} \quad (6)$$

b. *Recall*

Recall adalah metode untuk mengukur jumlah prediksi yang relevan dengan cara menghitung jumlah kata yang sama, baik unigram, bigram, atau LCS dibagi keseluruhan kata pada ringkasan manusia.

$$ROUGE - 1 = \frac{\text{jumlah unigram kata sama}}{\text{keseluruhan kata di ringkasan manusia}} \quad (7)$$

$$ROUGE - 2 = \frac{\text{jumlah bigram kata sama}}{\text{keseluruhan kata di ringkasan manusia}} \quad (8)$$

$$ROUGE - L = \frac{LCS(\text{Longest Common Subsequent})}{\text{keseluruhan kata di ringkasan manusia}} \quad (9)$$

c. *F1-Score*

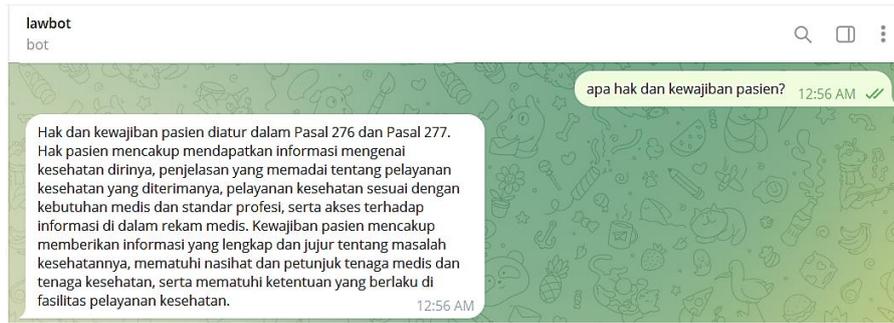
F1-score atau *F-measure* adalah metode untuk mengukur nilai rata-rata harmonik (*harmonic mean*) antara *recall* dan *precision*.

$$F1 \text{ score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (10)$$

3. HASIL DAN PEMBAHASAN

Setelah model telah siap, implementasi *chatbot* diintegrasikan dengan Telegram. Hasil dari implementasi *Question Answering System* (QAS) menggunakan *chatbot* Telegram seperti pada gambar 4.

Chatbot membentuk hubungan langsung antara pengguna dan sistem. Dengan memanfaatkan LLM seperti yang ditunjukkan pada gambar 2, *chatbot* dapat berinteraksi dengan cara yang dipersonalisasi dan responsif. *LangChain* mengoptimalkan pengetahuan khusus tentang undang-undang sehingga dapat secara efektif menangani berbagai pertanyaan dan masalah pengguna. Kemampuannya dalam memahami konteks memungkinkan dukungan yang dipersonalisasi. Jawaban otomatis yang didukung oleh LLM memberikan respons yang mudah dimengerti.



Gambar 4. Tampilan Halaman Interaksi *Chatbot*

3.1. Pengujian

Pengujian dilakukan dengan menggunakan 10 sampel pertanyaan seperti pada tabel 1 dengan menghitung nilai *precision*, *recall*, dan *f1-score* setiap pertanyaan.

Tabel 1. Sampel Pertanyaan

No	Pertanyaan
1	Apa hak dan kewajiban sebagai pasien?
2	Saya adalah mahasiswa kesehatan dari luar negeri, apakah saya boleh bekerja di Indonesia?
3	Apa syarat melakukan transplantasi organ?
4	Apa hak bayi dan anak?
5	Apa sanksi orang yang tidak mau di karantina?
6	Apa hak yang diperoleh oleh tenaga kesehatan yang bekerja di tempat daerah tertinggal?
7	Apa sanksi jika mengedarkan narkotika secara ilegal?
8	Apakah tenaga Kesehatan boleh membantu aborsi?
9	Apa hukuman untuk tenaga Kesehatan yang lamban memberikan penanganan kepada pasien?
10	Apakah asuransi Kesehatan dapat membantu dalam mengakses layanan Kesehatan?

Berikut adalah contoh jawaban referensi dan jawaban prediksi:

Jawaban referensi Soal no 1:

“Berdasarkan pasal 276 dan 277, pasien memiliki hak untuk mendapatkan informasi, penjelasan yang memadai, dan pelayanan Kesehatan sesuai dengan kebutuhan medis, standar profesi, dan bermutu, pasien juga berhak untuk mendapatkan akses terhadap informasi tentang rekam medisnya dan mendapatkan hak lain sesuai dengan ketentuan peraturan perundang-undangan, selain itu pasien juga berhak menolak atau menyetujui tindakan medis kecuali untuk mencegah dan menanggulangi wabah atau KLB. Pasien juga memiliki kewajiban untuk memberikan informasi secara lengkap dan jujur, mematuhi nasihat dan ketentuan yang berlaku dari tenaga medis atau tenaga Kesehatan, serta memberikan imbalan jasa atas pelayanan yang diterima.”

Jawaban prediksi Soal no 1:

“Hak dan kewajiban pasien diatur dalam Pasal 276 dan Pasal 277. Hak pasien mencakup mendapatkan informasi mengenai kesehatan dirinya, penjelasan yang memadai tentang pelayanan kesehatan yang diterimanya, pelayanan kesehatan sesuai dengan kebutuhan medis dan standar profesi, serta akses terhadap informasi di dalam rekam medis. Kewajiban pasien mencakup memberikan informasi yang lengkap dan jujur tentang masalah kesehatannya, mematuhi nasihat dan petunjuk tenaga medis dan tenaga kesehatan, serta mematuhi ketentuan yang berlaku di fasilitas pelayanan kesehatan.”

Jawaban referensi Soal no 2:

“Ketentuan untuk warga negara asing lulusan luar negeri yang ingin melaksanakan praktik Kesehatan di Indonesia diatur pada pasal 248-257. Tenaga Medis dan Tenaga Kesehatan warga negara asing lulusan luar negeri yang dapat melaksanakan praktik di Indonesia hanya berlaku untuk Tenaga Medis spesialis dan subspecialis serta Tenaga Kesehatan tingkat kompetensi tertentu setelah mengikuti evaluasi kompetensi. Evaluasi kompetensi dilakukan oleh Menteri dengan melibatkan menteri yang menyelenggarakan urusan pemerintahan di bidang pendidikan, Konsil, dan Kolegium. Evaluasi kompetensi meliputi penilaian kelengkapan administratif dan penilaian kemampuan praktik, yang dilakukan setelah penilaian kelengkapan administratif. Penilaian kemampuan praktik meliputi penyetaraan kompetensi dan uji kompetensi. Penyetaraan kompetensi bertujuan untuk memastikan kesesuaian dengan standar kompetensi Tenaga Medis dan Tenaga Kesehatan di Indonesia. Hasil uji

kompetensi dapat menentukan apakah tenaga medis tersebut kompeten atau belum, yang akan menentukan apakah mereka harus mengikuti adaptasi pada Fasilitas Pelayanan Kesehatan atau harus kembali ke negara asal sesuai dengan peraturan perundang-undangan. Tenaga Medis spesialis dan subspecialis serta Tenaga Kesehatan tingkat kompetensi tertentu warga negara asing lulusan luar negeri yang akan melakukan praktik di Indonesia harus memiliki STR dan SIP. Persyaratan lainnya serta ketentuan lebih lanjut mengenai pendayagunaan mereka diatur dengan Peraturan Pemerintah.”

Jawaban prediksi Soal no 2:

“Tidak bisa dipastikan karena tergantung pada beberapa faktor seperti apakah Anda sudah lulus dari penyelenggara pendidikan di luar negeri yang sudah direkognisi dan telah memiliki pengalaman praktik paling sedikit 2 tahun di luar negeri atau merupakan ahli dalam bidang unggulan tertentu dalam Pelayanan Kesehatan yang dibuktikan dengan sertifikasi kompetensi. Selain itu, Anda juga perlu memiliki STR dan SIP sesuai dengan ketentuan yang berlaku. Lebih lanjut mengenai pendayagunaan Tenaga Medis dan Tenaga Kesehatan warga negara asing lulusan luar negeri di Indonesia dapat dilihat pada Pasal 251 dan Pasal 249 dalam Undang-Undang yang diatur dengan Peraturan Pemerintah.”

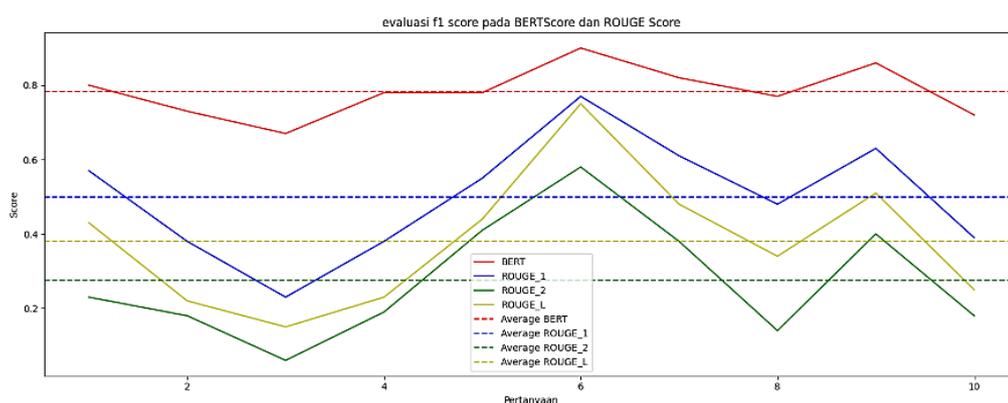
Tabel 2 merupakan hasil evaluasi *precision*, *recall*, dan *f1-score* masing masing pertanyaan menggunakan BERTScore.

Tabel 2. Hasil Evaluasi BertScore dan ROUGE Score

Pertanyaan	ROUGE-1(%)			ROUGE-2(%)			ROUGE-3(%)			BERTScore(%)		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
1	65%	51%	57%	27%	21%	23%	49%	38%	43%	79%	82%	80%
2	57%	29%	38%	27%	14%	18%	33%	17%	22%	71%	74%	73%
3	22%	24%	23%	6%	7%	6%	13%	16%	15%	67%	68%	67%
4	78%	25%	38%	42%	13%	20%	48%	15%	23%	73%	84%	78%
5	75%	43%	55%	58%	32%	42%	60%	34%	44%	72%	84%	78%
6	76%	78%	77%	58%	59%	58%	74%	76%	75%	90%	91%	90%
7	74%	52%	61%	47%	33%	38%	58%	41%	48%	78%	87%	82%
8	50%	46%	48%	15%	14%	14%	35%	32%	34%	76%	78%	77%
9	62%	64%	63%	40%	40%	40%	51%	52%	51%	86%	86%	86%
10	43%	36%	39%	21%	17%	19%	27%	23%	25%	72%	71%	72%
Rata-Rata	60%	45%	50%	34%	25%	28%	45%	34%	38%	76%	80%	78%

Berdasarkan evaluasi pada tabel 2, BERTScore menunjukkan keunggulan dibandingkan ROUGE score karena ROUGE cenderung lebih fokus pada panjang kesamaan setiap kata dalam peringkasan dokumen, sehingga tidak sensitif terhadap perubahan makna kata.

Grafik hasil evaluasi BERTScore dan ROUGE score ditunjukkan pada gambar 5.



Gambar 5. Hasil Evaluasi BERTScore dan ROUGE Score

Gambar 5 menunjukkan bahwa terdapat korelasi antara pergerakan grafik BERTScore dan ROUGE score untuk setiap pertanyaan. Ketika grafik BERTScore menurun, hasil ROUGE score juga cenderung menurun, dan sebaliknya.

4. KESIMPULAN

Penelitian ini telah berhasil memanfaatkan teknologi *Langchain* dan *Large Language Models (LLM)* untuk mengembangkan sebuah sistem *Question Answering* yang terintegrasi dengan *chatbot*. Dalam sistem ini,

chatbot dapat berinteraksi dengan pengguna dan memberikan jawaban atas setiap pertanyaan yang diajukan. Namun, untuk menggunakan file PDF UU No. 17 Tahun 2023 tentang Kesehatan sebagai sumber data, harus mengubahnya menjadi format TXT agar *chatbot* dapat memberikan jawaban yang tepat kepada pengguna. Data yang telah diproses menggunakan LLM melalui proses *chunking* dan *embeddings* telah menjadi basis pengetahuan sistem ini, yang digunakan sebagai referensi dalam memberikan jawaban kepada pengguna.

Kelemahan dari *Chatbot* ini hanya dapat merespons pertanyaan yang diajukan oleh pengguna dan tidak mampu merespons kata sapaan, ucapan terima kasih, atau permintaan maaf. Setelah dilakukan pengujian menggunakan BERTScore dan ROUGE Score, hasil pengujian menggunakan BERTScore menunjukkan rata-rata nilai *precision*, *recall*, dan *f1-score* sebesar 76%, 80%, dan 78% secara berturut-turut. Sedangkan untuk ROUGE-1, nilai *precision*, *recall*, dan *f1-score* adalah 60%, 45%, dan 50%. Untuk ROUGE-2, nilainya adalah 34%, 25%, dan 28%, dan untuk ROUGE-L adalah 45%, 34%, dan 38%.

Saran untuk penelitian selanjutnya, dapat menggunakan *chatbot* yang lebih interaktif. Dengan meningkatkan interaktivitas, *chatbot* akan lebih efektif dalam memahami dan menjawab pertanyaan, memberikan umpan balik yang relevan, dan berinteraksi lebih alami dengan pengguna. Hal ini akan meningkatkan pengalaman pengguna dalam menggunakan *chatbot*.

REFERENCES

- [1] E. Susanna and Y. Kamaruzzaman, "Penegakan Hukum Pelayanan Publik di Bidang Kesehatan pada Rumah Sakit Umum di Kota Banda Aceh," *Journal of Law and Government Science*, vol. 9, no. 1, pp. 1–10, 2023.
- [2] Yandriza and D. Arma, "Pertanggungjawaban Pidana Rumah Sakit atas Kelalaian Perawat yang Mengakibatkan Tertukarnya Bayi yang Baru Dilahirkan dalam Perspektif Hukum Kesehatan di Indonesia," *Unes Law Review*, vol. 6, no. 1, pp. 2462–2656, 2023, doi: 10.31933/unesrev.v6i1.
- [3] G. Widjaja, "Pelayanan Kesehatan Bagi Pasien Menurut UU No. 17 Tahun 2023 Tentang Kesehatan," *INNOVATIVE: Journal Of Social Science Research*, vol. 3, no. 5, 2023, doi: 10.31004/innovative.v3i6.6066.
- [4] S. I. Kesuma, "Sosialisasi Tentang Ulasan Undang-Undang No. 17 Tahun 2023 Tentang Kesehatan," *Jurnal Ilmu Hukum dan Tata Negara*, vol. 1, no. 4, pp. 143–156, 2023, doi: 10.55606/birokrasi.v1i4.731.
- [5] A. Setiawan, O. N. Pratiwi, and ..., "Question Answering System Dalam Bentuk Chatbot Pada Platform Line Untuk Mata Pelajaran Sejarah Sma/ma Dengan Menggunakan Algoritma Levenshtein ...," *eProceedings ...*, vol. 8, no. 5, pp. 9794–9802, 2021, [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/15756%0Ah> <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/15756/15469>
- [6] G. F. Avisyah, I. J. Putra, and S. S. Hidayat, "Open Artificial Intelligence Analysis using ChatGPT Integrated with Telegram Bot," *Jurnal ELTIKOM: Jurnal Teknik Elektro, Teknologi Informasi dan Komputer*, vol. 7, no. 1, pp. 60–66, Jun. 2023, doi: 10.31961/eltikom.v7i1.724.
- [7] A. Marsadualan, Harmastuti, and J. Triyono, "Rancang Bangun Aplikasi Tanya Jawab Mengenai IST Akprind Yogyakarta Berbasis Mobile Menggunakan Algoritma Boyer Moore," in *SNISTIK: Seminar Nasional Inovasi Sains Teknologi Informasi Komputer*, 2023, pp. 226–238.
- [8] K. Pandya and M. Holia, "Automating Customer Service using LangChain Building custom open-source GPT Chatbot for organizations," in *International Conference on "Women in Science & Technology: Creating Sustainable Career"*, 2023. doi: 10.48550/arXiv.2310.05421.
- [9] J. B. Ilagan and J. R. Ilagan, "A Prototype of a Chatbot for Evaluating and Refining Student Startup Ideas Using a Large Language Model," 2023.
- [10] P. N. Devaraj, R. T. P. V, A. Gangrade, and M. K. R, "Development of a Legal Document AI-Chatbot," 2023, doi: 10.48550/arXiv.2311.12719.
- [11] Y. Chang *et al.*, "A Survey on Evaluation of Large Language Models," *Journal of the ACM*, vol. 37, no. 4, Jul. 2023, [Online]. Available: <http://arxiv.org/abs/2307.03109>
- [12] A. Pesaru, T. S. Gill, and A. R. Tangella, "AI assistant for document management Using Lang Chain and Pinecone," *International Research Journal of Modernization in Engineering Technology and Science*, vol. 05, no. 06, pp. 3980–3983, Jun. 2023, doi: 10.56726/irjmets42630.
- [13] N. Hidayah, "Tantangan dan Strategi Pengajar dalam Menyikapi Perkembangan Teknologi Artificial Intelligence (AI), Study Kasus pada Mahasiswa Keperawatan dalam Penyelesaian Tugas Akhir," in *Prosiding Seminar Nasional Pascasarjana*, 2023, pp. 552–555. [Online]. Available: <http://pps.unnes.ac.id/pp2/prodi/prosiding-pascasarjana-unnes>
- [14] J. Ioannidis, J. Harper, M. S. Quah, and D. Hunter, "Gracernote.ai: Legal Generative AI for Regulatory Compliance," in *Third International Workshop on Artificial Intelligence and Intelligent Assistance for Legal Professionals in the Digital Workplace*, 2023. doi: 10.2139/ssrn.4494272.

-
- [15] 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.
- [16] T. Iskandar Zulkarnain Maulana Putra, A. Farhan Bukhori, dan Ilmu Pengetahuan Alam, and U. Gadjah Mada, "Model Klasifikasi Berbasis Multiclass Classification dengan Kombinasi Indobert Embedding dan Long Short-Term Memory untuk Tweet Berbahasa Indonesia (Classification Model Based on Multiclass Classification with a Combination of Indobert Embedding and Long Short-Term Memory for Indonesian-language Tweets)," *Jurnal Ilmu Siber dan Teknologi Digital (JISTED)*, vol. 1, no. 1, pp. 1–28, 2022, doi: 10.35912/jisted.v1i1.1509.
- [17] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, and Y. Artzi, "BERTScore: Evaluating Text Generation with BERT," in *ICLR*, 2020, pp. 1–43. doi: 10.48550/arXiv.1904.09675.
- [18] I. J. Unanue, J. Parnell, and M. Piccardi, "BERTTune: Fine-Tuning Neural Machine Translation with BERTScore," Jun. 2021, doi: 10.48550/arXiv.2106.02208.
- [19] M. Hanna and O. Bojar, "A Fine-Grained Analysis of BERTScore," in *Sixth Conference on Machine Translation (WMT)*, 2021, pp. 507–517.
- [20] T. Sun, J. He, X. Qiu, and X. Huang, "BERTScore is Unfair: On Social Bias in Language Model-Based Metrics for Text Generation," Oct. 2022, doi: 10.48550/arXiv.2210.07626.
- [21] F. V. P. Samosir, H. Toba, and M. Ayub, "BESKlus : BERT Extractive Summarization with K-Means Clustering in Scientific Paper," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 8, no. 1, Apr. 2022, doi: 10.28932/jutisi.v8i1.4474.
- [22] I. Athiyah Rahma and L. Hulliyyatus Suadaa, "Penerapan Text Augmentation untuk Mengatasi Data yang Tidak Seimbang pada Klasifikasi Teks Berbahasa Indonesia Studi Kasus: Deteksi Judul Clickbait dan Komentar Hate Speech pada Berita Online," *Jurnal Teknologi Informasi dan Ilmu Komputer(JTIK)*, vol. 10, no. 6, pp. 1329–1340, 2023, doi: 10.25126/jtiik.2023107325.
- [23] A. Celikyilmaz, E. Clark, and J. Gao, "Evaluation of Text Generation: A Survey," Jun. 2020, doi: 10.48550/arXiv.2006.14799.
- [24] M. Moradi, M. Dashti, and M. Samwald, "Summarization of biomedical articles using domain-specific word embeddings and graph ranking," *J Biomed Inform*, vol. 107, Jul. 2020, doi: 10.1016/j.jbi.2020.103452.
- [25] Halimah, Surya Agustian, and Siti Ramadhani, "Peringkasan teks otomatis (automated text summarization) pada artikel berbahasa indonesia menggunakan algoritma lexrank," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 3, no. 3, pp. 371–381, Dec. 2022, doi: 10.37859/coscitech.v3i3.4300.
- [26] Z. Idhafi, S. Agustian, F. Yanto, and N. Safaat H, "Peringkasan teks otomatis pada artikel berbahasa indonesia menggunakan metode maximum marginal relevance," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 4, no. 3, pp. 609–618, Dec. 2023, doi: 10.37859/coscitech.v4i3.6311.