



## Implementasi Retrieval Augmented Generation untuk Layanan Informasi Kampus dengan Chatbot Berbasis AI

Yoni Tribber<sup>1\*</sup>, Kusnadi<sup>2</sup>, Marsani Asfi<sup>3</sup>

<sup>1,2,3</sup>Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Catur Insan Cendekia  
[tribberyoni5@gmail.com](mailto:tribberyoni5@gmail.com)

### Abstract

*Information technology plays a crucial role in enhancing knowledge and awareness, particularly in educational institutions with intense competition among universities. Students and the general public still face difficulties in obtaining the necessary information due to the numerous available platforms and the lack of effective search features. This research aims to develop an AI-based chatbot, specifically using Large Language Models (LLM), that can offer efficient and responsive solutions for information services in educational institutions. By integrating the chatbot into the Telegram application, it is expected to facilitate users in quickly and accurately obtaining information. This chatbot is built with a Retrieval Augmented Generation (RAG) approach that enhances LLM capabilities and helps in obtaining fast and accurate information. The chatbot is developed using software such as Langchain to manage the RAG process, Python as the programming language, ChromaDB as the vector database, and Gemini AI to support the LLM model. Performance evaluation is conducted using RAGAS to ensure the quality and accuracy of responses. Testing with 10 sample questions showed positive results, a context precision score of 93%, faithfulness 100%, answer relevancy 96%, context recall 100%, answer correctness 71%, and answer similarity 93%. These results indicate that an RAG-based chatbot can be an effective tool to improve information accessibility in educational institutions.*

**Keywords:** Chatbot, Large Language Model, Retrieval Augmented Generation, Telegram, Langchain, Python, ChromaDB, Gemini AI.

### Abstrak

Teknologi informasi memiliki peran yang sangat penting untuk meningkatkan wawasan dan pengetahuan, terutama pada institusi pendidikan dengan persaingan perguruan tinggi yang ketat. Mahasiswa dan masyarakat umum masih mengalami kesulitan dalam mendapatkan informasi yang dibutuhkan karena banyaknya platform yang tersedia dan kurangnya fitur pencarian yang efektif. Penelitian ini bertujuan untuk membangun *chatbot* berbasis AI, khususnya dengan *Large Language Models* (LLM), yang dapat menawarkan solusi efisien dan responsif untuk layanan informasi di institusi pendidikan. Dengan mengintegrasikan *chatbot* ke dalam aplikasi Telegram, diharapkan dapat memudahkan pengguna dalam mendapatkan informasi secara cepat dan akurat. *Chatbot* ini dibangun dengan pendekatan *Retrieval Augmented Generation* (RAG) yang dapat meningkatkan kemampuan LLM dan membantu mendapatkan informasi yang cepat dan tepat. *Chatbot* ini dikembangkan menggunakan perangkat lunak seperti Langchain untuk mengelola proses RAG, Python sebagai bahasa pemrograman, ChromaDB sebagai basis data vektor, dan Gemini AI untuk mendukung model LLM. Evaluasi kinerja dilakukan menggunakan RAGAS untuk memastikan kualitas dan keakuratan jawaban. Pengujian ini menggunakan 10 sampel pertanyaan menunjukkan hasil yang positif, dengan mendapatkan skor *context precision* 93%, *faithfulness* 100%, *answer relevancy* 96%, *context recall* 100%, *answer correctness* 71%, dan *answer similarity* 93%. Hasil ini menunjukkan bahwa *chatbot* berbasis RAG dapat menjadi alat yang efektif untuk meningkatkan aksesibilitas informasi di institusi pendidikan.

**Kata kunci:** Chatbot, Large Language Model, Retrieval Augmented Generation, Telegram, Langchain, Python, ChromaDB, Gemini AI.

### 1. Pendahuluan

Teknologi tak terpisahkan dari informasi karena informasi memiliki peran yang sangat penting dalam

meningkatkan wawasan dan pengetahuan bagi semua individu [1]. Dalam konteks pendidikan, teknologi informasi dapat menyediakan akses yang lebih cepat dan mudah ke sumber daya akademik, sehingga

memperbaiki cara institusi menyampaikan layanan dan informasi kepada mahasiswa. Mengingat pertumbuhan teknologi informasi yang pesat dan persaingan antar perguruan tinggi yang semakin ketat, kualitas pelayanan akademik menjadi sangat penting [2].

Universitas Catur Insan Cendekia merupakan salah satu perguruan tinggi swasta di Cirebon. Berdasarkan observasi, mahasiswa dan masyarakat umum masih mengalami kesulitan dalam mendapatkan informasi yang dibutuhkan. Kesulitan ini disebabkan oleh banyaknya platform yang tersedia, yang sering kali menyebabkan kebingungan dan kesulitan dalam menemukan informasi yang tepat. Dengan banyaknya platform, informasi sering tersebar di berbagai tempat tanpa adanya integrasi yang baik, sehingga pengguna harus mencari di beberapa sumber yang berbeda. Kurangnya fitur pencarian yang efektif dan ketiadaan fitur pencarian yang memadai pada portal informasi institusi juga memperburuk situasi. Tingkat kepuasan terhadap penyampaian informasi pun rendah, menunjukkan kebutuhan akan solusi yang lebih efisien dan *user-friendly*.

*Chatbot* kini menjadi solusi menarik di tengah tantangan institusi pendidikan untuk memberikan layanan informasi secara efisien dan responsif. Dalam beberapa tahun terakhir, bidang kecerdasan buatan (AI) dan *machine learning*, terutama melalui *Large Language Models* (LLM) seperti Chat-GPT dan Gemini AI, telah berkembang pesat dengan minat publik yang signifikan [3]. LLM adalah model bahasa besar yang menggunakan AI untuk memahami dan menghasilkan teks dalam berbagai konteks. *Chatbot*, sebagai asisten digital berbasis AI, dapat menangani berbagai masalah, termasuk di bidang kesehatan dan pendidikan, dengan menyampaikan informasi yang relevan [4].

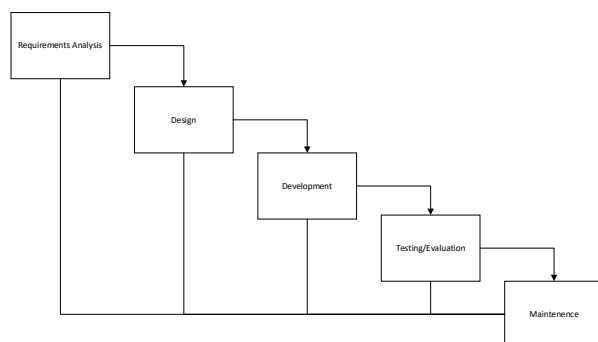
*Large Language Models* (LLM) terbaru sering mengalami ketidakakuratan atau halusinasi saat diminta menghasilkan informasi yang memerlukan pengetahuan baru atau *domain-specific*. Untuk mengatasi masalah ini, penting untuk menghubungkan pengetahuan umum LLM dengan konteks tambahan agar dapat menghasilkan jawaban yang lebih akurat. Salah satu solusi yang efektif adalah metode *Retrieval-Augmented Generation* (RAG), yang mengurangi ketergantungan pada *fine-tuning* yang memakan waktu dan biaya [5]. Metode RAG menggabungkan pengambilan data eksternal (*retrieval*) dengan proses generatif (*text-generation*), sehingga dapat meningkatkan akurasi dan relevansi respons dengan memanfaatkan informasi dari sumber eksternal. Ini memungkinkan model untuk memberikan jawaban yang lebih spesifik dan faktual [6]. Dengan menggunakan metode RAG dapat menghasilkan bahasa yang lebih spesifik, beragam, dan faktual dibandingkan dengan *baseline seq2seq parametrik* [7]. Keunggulan ini dicapai dengan memanfaatkan pengetahuan dari model bahasa besar (*parametrik*) serta informasi yang diperoleh dari sumber eksternal (*non-parametrik*). Dengan cara ini, RAG

berusaha untuk menghasilkan teks yang lebih akurat dan informatif [8].

Dari penjelasan di atas, tujuan penelitian ini adalah untuk mengembangkan sebuah *chatbot* yang berfungsi sebagai asisten digital bagi masyarakat umum dan mahasiswa di Universitas Catur Insan Cendekia, sehingga mereka dapat mengakses informasi akademik maupun non-akademik dengan mudah dan cepat. *Chatbot* ini akan memanfaatkan metode *Retrieval-Augmented Generation* (RAG), yang menggabungkan pengambilan data eksternal dalam proses generatif, dengan harapan dapat memberikan respons yang lebih akurat dan relevan. Selain itu, *chatbot* ini akan diintegrasikan ke dalam aplikasi Telegram, memungkinkan mahasiswa untuk mengakses informasi dengan lebih mudah melalui platform yang sudah familiar.

## 2. Metode Penelitian

Metode pengembangan yang digunakan adalah metode *waterfall* seperti terlihat pada Gambar 1. Metode ini menawarkan pendekatan sistematis dan berurutan untuk pengembangan perangkat lunak yang dimulai dengan spesifikasi kebutuhan pelanggan dan berlanjut melalui perencanaan/desain, pemodelan/pengembangan, pengujian, dan implementasi, hingga dukungan berkelanjutan (*maintenance*) [9]. Metode ini dipilih karena sangat cocok untuk proyek yang memiliki persyaratan yang jelas sejak awal, seperti layanan informasi kampus yang sudah memiliki kebutuhan fungsional dan non-fungsional yang ditetapkan.

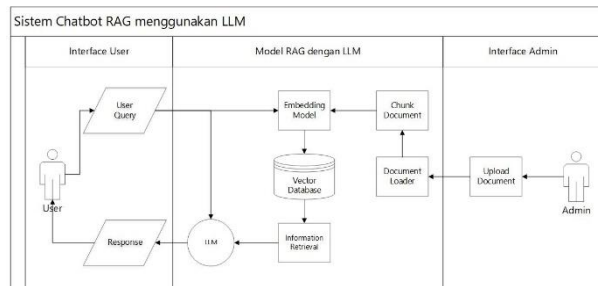


Gambar 1. Metode Waterfall

**Analisis Persyaratan (Requirements Analysis):** Tahap awal ini melibatkan kerjasama dengan Universitas Catur Insan Cendekia untuk memahami kebutuhan dan tujuan sistem *chatbot*. Analisis dilakukan untuk mengidentifikasi jenis informasi yang diperlukan oleh mahasiswa dan masyarakat umum lalu menetapkan fungsionalitas penting seperti menjawab pertanyaan tentang jadwal, proses registrasi mahasiswa baru, dan informasi biaya kuliah.

**Desain Sistem (Design):** Pada tahap ini, arsitektur sistem *chatbot* dirancang, mencakup alur kerja sistem, struktur basis data, pemilihan model bahasa, dan aspek desain lainnya. Rancangan ini bertujuan memastikan sistem

chatbot dapat beroperasi dengan lancar dan memenuhi kebutuhan pengguna.



Gambar 2. Alur Kerja Sistem

Berdasarkan gambar 2, yang merupakan alur kerja sistem chatbot dengan *retrieval augmented generation*, dengan keterangan sebagai berikut:

**User:** Pengguna yang berinteraksi dengan sistem, memasukkan *query* dan menerima *response*.

**User Query:** Permintaan yang diajukan oleh pengguna, berupa teks atau kalimat.

**Document Loader:** Modul yang memuat dokumen-dokumen yang akan diproses oleh sistem.

**Information Retrieval:** proses mengambil informasi yang relevan dari sumber data yang besar dan kompleks.

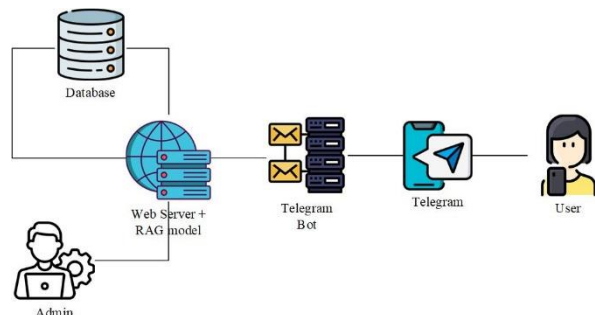
**Vector Database:** Penyimpanan *vector embedding* untuk setiap kata dalam skema semantik.

**Chunk Document:** Potongan-potongan dokumen yang digunakan proses pengambilan informasi.

**LLM (Large Language Model):** Model bahasa besar yang digunakan untuk menghasilkan *response* berdasarkan *query* dan *vector embedding*.

**Response:** Respon yang dihasilkan oleh sistem berdasarkan *query* dan *vector embedding*.

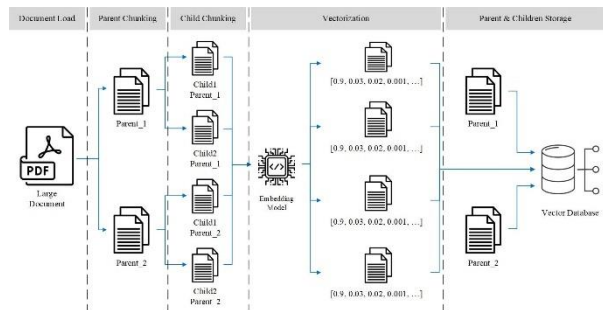
**Pengembangan (Development):** Implementasi sistem dilakukan sesuai desain yang telah dibuat. Tahap ini melibatkan pembangunan model bahasa dan pengembangan sistem chatbot secara keseluruhan. Dan integrasi ke aplikasi Telegram seperti pada Gambar 3.



Gambar 3. Arsitektur Sistem

Sistem *Retrieval Augmented Generation* (RAG) memiliki dua komponen utama: *Indexing* dan *Retrieval* serta *Generation*. Komponen *Indexing* mengumpulkan

dan mengindeks data dari sumber tertentu. Komponen *Retrieval* dan *Generation*, atau rantai RAG, bekerja saat pengguna mengajukan pertanyaan. Rantai ini mencari data relevan dari hasil *indexing* dan mengirimkannya ke model bahasa generatif untuk diolah menjadi teks yang sesuai dengan pertanyaan pengguna.



Gambar 4. Indexing

Pada Gambar 4, ditampilkan proses *indexing* dokumen menggunakan teknik *parent* dan *child chunking*. Proses ini bertujuan untuk mengatur dan mengorganisir konten dokumen dalam format hirarkis, sehingga memungkinkan pengambilan informasi dan analisis yang lebih efisien. Proses ini terdiri dari lima tahapan, yang dijelaskan sebagai berikut:



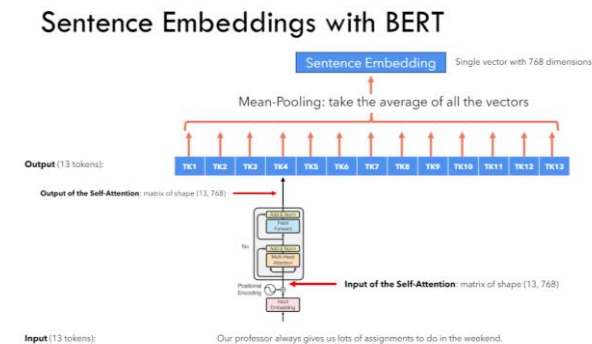
Gambar 5. Contoh Dokumen

Tahap pertama dimulai ketika admin mengunggah *file* dokumen, dilanjutkan dengan proses pemetaan dokumen. Tahap ini melibatkan ekstraksi konten teks dari dokumen, termasuk teks tersemat, gambar, dan elemen lainnya. Sebagai contoh, pada Gambar 5, sebuah dokumen PDF (*Portable Document Format*) yang dihasilkan dari proses *scanning* akan dimuat dan diekstrak menjadi teks dengan bantuan teknologi *Optical Character Recognition* (OCR).

Pada tahap *parent chunking*, sistem mengidentifikasi dan membagi dokumen menjadi bagian-bagian logis berdasarkan konten dan strukturnya. Bagian-bagian ini

mewakili *unit-level* informasi yang lebih tinggi. Setiap bagian disebut sebagai *parent chunk*.

Dalam setiap *parent chunk*, sistem lebih lanjut membagi konten menjadi unit-unit informasi yang lebih kecil. Unit-unit ini, yang dikenal sebagai *child chunks*, biasanya mewakili kalimat atau frasa. *Child chunks* bersarang di dalam *parent chunk* mereka yang sesuai, membentuk struktur hirarkis.



Gambar 6. Proses merubah kalimat menjadi vektor [10]

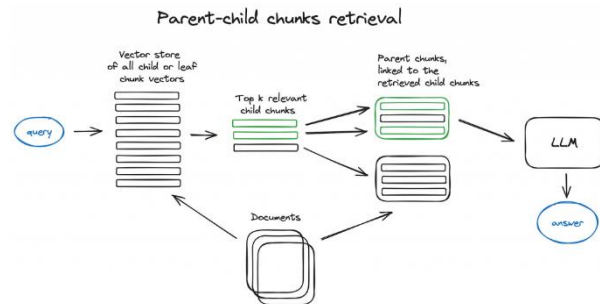
Proses pada gambar 6 terdiri dari beberapa langkah. Pertama, kalimat dimasukkan ke dalam model BERT dan diubah menjadi representasi token, di mana setiap token mewakili kata atau subkata. Selanjutnya, representasi token diproses melalui *encoder* BERT yang terdiri dari beberapa layer *transformer* untuk mempelajari hubungan antar token dalam kalimat. Terakhir, output dari *encoder* BERT adalah vektor representasi kalimat dengan dimensi 768 (untuk base model) yang mengandung informasi semantik tentang kalimat tersebut. Model *embedding* yang digunakan dalam proses ini adalah Sentence-BERT yang telah dilatih khusus untuk bahasa Indonesia dengan nama *firqaaa/indo-sentence-bert-base* dari platform Hugging Face, yang memungkinkan representasi kalimat yang lebih relevan dan akurat untuk tugas-tugas pemrosesan bahasa alami dalam bahasa Indonesia [10, 11].

*Parent chunks* dan *child chunks* yang sesuai, beserta vektor yang terkait, disimpan dalam basis data untuk vektor. Representasi terstruktur dari dokumen ini memudahkan pengambilan dan pemrosesan informasi spesifik.

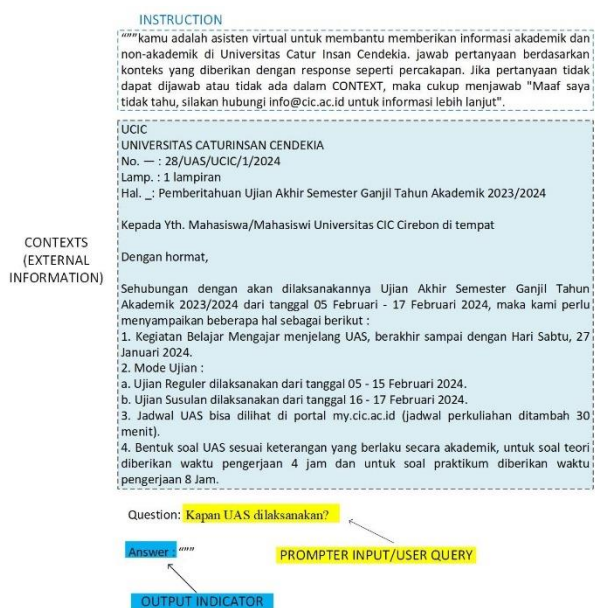
*Retrieval* mengacu pada proses mencari dan mengambil informasi yang relevan dari sumber yang tersedia. Dalam konteks pencarian informasi, *retrieve* dapat merujuk pada penggunaan mesin pencari untuk menemukan dokumen, artikel, atau data lainnya yang sesuai dengan kebutuhan pengguna.

*Parent Document Retriever* adalah komponen dalam sistem *Natural Language Processing* (NLP) yang bertugas untuk mengidentifikasi dokumen-dokumen yang berpotensi relevan terhadap permintaan atau konteks yang diberikan. Dimulai dengan mengambil segmen data yang lebih kecil (*child chunk*) yang paling

relevan untuk menjawab suatu pertanyaan, kemudian menggunakan bagian induk (*parent chunk*) yang terkait dengan segmen-segmen tersebut untuk mengakses dan mengembalikan bagian data induk (*parent chunk*) yang lebih besar yang akan digunakan sebagai konteks untuk LLM



Gambar 7. Parent Document Retriever [13]



Gambar 8. Proses Augmentasi

*Augmented* merupakan proses memperkaya atau memperluas informasi yang sudah ada dengan tambahan konteks atau detail dari proses *retrieval*. Terdapat 4 (empat) hal yang penting dalam proses augmentasi ini yaitu *instruction*, *contexts*, *user query*, dan *output indicator*. Berikut adalah penjelasannya:

*Instruction*, memberitahu model LLM apa yang harus dilakukan, bagaimana menggunakan informasi eksternal jika disediakan, apa yang harus dilakukan dengan permintaan, dan bagaimana membangun output.

*Contexts*, bertindak sebagai sumber pengetahuan tambahan bagi model. Ini dapat dimasukkan secara manual ke dalam prompt, diambil melalui database vektor (peningkatan pengambilan), atau ditarik melalui cara lain (API, perhitungan, dll.).

*User Query*, input kueri ke sistem oleh pengguna manusia (pemberi prompt).

*Output Indicator*, menandai awal teks yang akan dihasilkan.

*Generation* merupakan kemampuan sistem untuk menghasilkan teks atau jawaban berdasarkan informasi yang ditemukan selama proses pencarian, serta konteks yang diberikan oleh pengguna atau sistem itu sendiri. Ini melibatkan penggunaan model generasi teks yang canggih, seperti model bahasa besar (LLM), yang dapat memahami dan merespons pertanyaan atau permintaan pengguna dengan teks yang koheren dan informatif. Pada penelitian ini LLM yang digunakan merupakan LLM dari Google yaitu Gemini AI. Pada Gambar 9, contoh hasil dari generation menggunakan Gemini AI pada sistem RAG yang dikembangkan.

```
[ ] response = chain.invoke("Kapan UAS dilaksanakan?")
    display(Markdown(response))
```

➡ UAS dilaksanakan dari tanggal 05 Februari - 17 Februari 2024.

Gambar 9. *Output* hasil *generation*

Pengujian ini dilakukan dengan menggunakan RAGAS. RAGAS adalah sebuah *framework* yang digunakan untuk menguji dan mengevaluasi model *Retrieval-Augmented Generation* (RAG). Pengujian RAGAS dilakukan untuk memastikan sistem berfungsi dengan baik. Ini mencakup pengujian unit untuk setiap bagian dari sistem, pengujian integrasi untuk memastikan semua komponen bekerja bersama dengan baik, dan pengujian sistem dengan pengguna akhir untuk mengevaluasi kegunaan dan kinerja keseluruhan sistem.

Dalam RAGAS, terdapat beberapa metrik yang digunakan untuk menilai kinerja sistem. Metrik pertama adalah *context precision*, yang mengukur seberapa tepat konteks yang diambil oleh sistem dari sumber data. *Faithfulness* mengukur seberapa akurat informasi yang diberikan oleh sistem dibandingkan dengan sumber data aslinya. *Answer Relevancy* menilai sejauh mana jawaban yang diberikan oleh sistem relevan dengan pertanyaan yang diajukan. *Context Recall* mengukur seberapa lengkap konteks yang diambil oleh sistem dari sumber data. *Answer Correctness* menilai kebenaran jawaban yang diberikan oleh sistem, sedangkan *Answer Similarity* mengukur kemiripan jawaban yang diberikan oleh sistem dengan jawaban yang diharapkan atau standar.

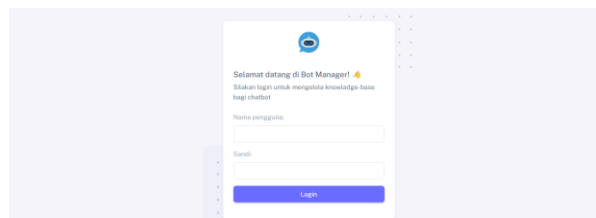
Setelah pengujian, sistem *chatbot* diluncurkan untuk digunakan oleh mahasiswa hingga masyarakat umum untuk mendapatkan informasi mengenai Universitas Catur Insan Cendekia. Staf BAAK akan memantau kinerja sistem agar tetap optimal dan sesuai dengan kebutuhan pengguna.

### 3. Hasil dan Pembahasan

Pada bagian ini akan menjelaskan hasil implementasi *chatbot*, termasuk tampilan antarmuka dan hasil pengujian menggunakan *framework* RAGAS.

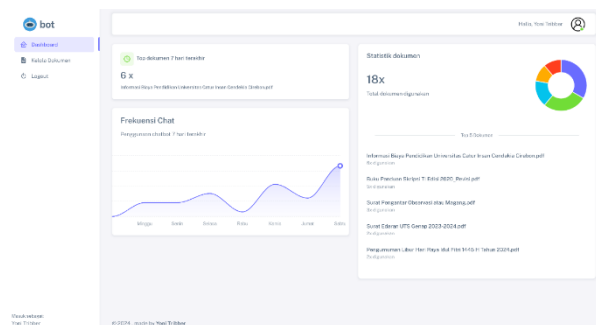
Implementasi ini mencakup penilaian terhadap kinerja *chatbot* dalam memberikan jawaban yang akurat dan relevan, serta analisis efektivitas metode *Retrieval-Augmented Generation* (RAG) yang diterapkan. Hasil pengujian RAGAS memberikan wawasan mendalam mengenai berbagai metrik yang menunjukkan kemampuan *chatbot* dalam memenuhi kebutuhan informasi pengguna secara efisien.

#### 3.1. Hasil Implementasi Tampilan



Gambar 10. Halaman *Login* Admin

Gambar 10 menunjukkan halaman *login* untuk aplikasi yang digunakan untuk mengelola *chatbot* seperti menambah, menghapus dan melihat dokumen yang digunakan sebagai pengetahuan *chatbot* (*knowledge-based chatbot*).



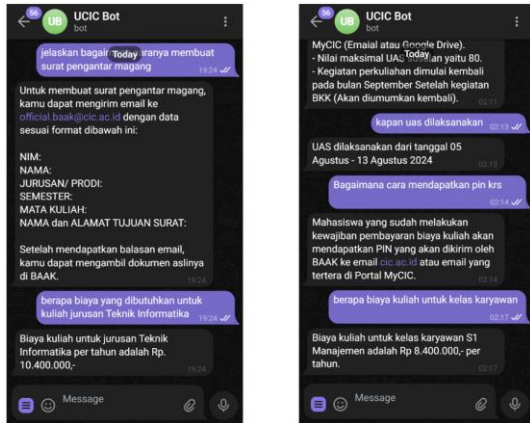
Gambar 11. Halaman *Dashboard*

Pada gambar 11, menunjukkan halaman *dashboard* admin. Halaman ini memungkinkan admin untuk melihat informasi tentang *chatbot*, seperti jumlah seberapa banyak sebuah dokumen digunakan pada proses pengambilan informasi (*information retrieval*), seberapa banyak penggunaan *chatbot* dalam rentang 7 (tujuh) hari terakhir, dan dokumen dengan penggunaan terbanyak (*information retrieval*) dalam 7 hari terakhir.

| ID | NAMA FILE                                                    | WAKTU UPLOAD | JUMLAH DOKUMEN | AKSI           |
|----|--------------------------------------------------------------|--------------|----------------|----------------|
| 1  | Skema Pembahasan UAS Semester Genap TA 2023-2024.pdf         | 10 Juli 2024 | 2 kali         | [Edit] [Hapus] |
| 2  | Buku Panduan Skripsi TI Edisi 2020, Revisi.pdf               | 10 Juli 2024 | 10 kali        | [Edit] [Hapus] |
| 3  | Skat Koneksi Rantai Kajian.pdf                               | 10 Juli 2024 | 0 kali         | [Edit] [Hapus] |
| 4  | Skat Edisi 1 UTS Genap 2023-2024.pdf                         | 10 Juli 2024 | 4 kali         | [Edit] [Hapus] |
| 5  | Pengumuman Ujian Harat Rupa (R&P) R&P 1440 14 Tahun 2024.pdf | 10 Juli 2024 | 0 kali         | [Edit] [Hapus] |

Gambar 12. Halaman *Kelola Dokumen*

Pada gambar 12, menunjukkan halaman Kelola Dokumen. Halaman ini memungkinkan admin untuk mengunggah, melihat, dan menghapus dokumen yang digunakan sebagai pengetahuan (*knowledge-base*) untuk *chatbot*.



Gambar 13. Tampilan Chatbot

Gambar 13 menunjukkan tampilan pengguna dari aplikasi Telegram. Pengguna dapat memasukkan pertanyaan pada kotak teks yang tersedia, kemudian sistem RAG akan dijalankan untuk menghasilkan jawaban.

### 3.2. Pengujian RAGAS

Pengujian ini bertujuan untuk mengevaluasi kualitas jawaban yang dihasilkan oleh sistem *Retrieval-Augmented Generation* (RAG) menggunakan alat/*library* RAGAS. RAGAS dipilih karena metrik ini secara khusus dirancang untuk menilai kinerja sistem RAG, dengan fokus pada relevansi dan kesesuaian jawaban dengan dokumen yang diambil selama proses pengambilan informasi. Terdapat sampel 10 pertanyaan digunakan untuk menilai kualitas jawaban yang dihasilkan oleh sistem RAG. Evaluasi dilakukan dengan membandingkan jawaban yang dihasilkan oleh sistem dengan jawaban referensi (*ground truth*) dan mengukur performa menggunakan metrik RAGAS.

Tabel 1. Sample Pertanyaan

| No. | Pertanyaan                                                                                         |
|-----|----------------------------------------------------------------------------------------------------|
| 1   | Berapa total biaya kuliah per tahun untuk Program Strata 1 (S1) Fakultas Teknologi Informasi UCIC? |
| 2   | Di mana mahasiswa dapat melihat jadwal UAS?                                                        |

- 3 Bagaimana mahasiswa dapat melakukan pembayaran biaya kuliah secara transfer?
- 4 Bagaimana cara mendaftar mahasiswa baru di UCIC?
- 5 Kapan UAS dilaksanakan?
- 6 Bagaimana cara mengajukan surat pengantar observasi di UCIC?
- 7 Di mana pendaftaran offline dapat dilakukan?
- 8 kapan libur idul fitri?
- 9 Berapa lama waktu pengerjaan soal untuk UAS?
- 10 jelaskan penilaian dalam skripsi

Tabel 2. Jawaban Referensi

| No. | Jawaban referensi ( <i>ground truth</i> )                                                                                                                                                                                     |
|-----|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1   | Rp. 10.400.000.                                                                                                                                                                                                               |
| 2   | Jadwal UAS bisa dilihat di portal my.cic.ac.id.                                                                                                                                                                               |
| 3   | Transfer ke Bank BRI atas nama Universitas Catur Insan Cendekia dengan nomor rekening 0406-01-000977-30-6.                                                                                                                    |
| 4   | Pendaftaran mahasiswa baru dapat dilakukan melalui tiga cara: Pendaftaran Online melalui Web Aplikasi, Pendaftaran Online melalui Aplikasi Whatsapp, dan Pendaftaran Offline.                                                 |
| 5   | UAS dilaksanakan dari tanggal 05 Agustus - 13 Agustus 2024.                                                                                                                                                                   |
| 6   | Mahasiswa harus mengirim data sesuai format yang telah ditentukan ke email official.baak@cic.ac.id, kemudian mengambil dokumen aslinya di BAAK setelah mendapatkan balasan email.                                             |
| 7   | Pendaftaran offline dapat dilakukan di Sekretariat Kampus CIC, atau di acara/kegiatan yang diselenggarakan atau diikuti oleh pihak kampus CIC seperti acara Kampus, Univday/Expo Pendidikan, Presentasi, Seminar, Bazar, dll. |
| 8   | Dari tanggal 6 April 2024 s/d 15 April 2024.                                                                                                                                                                                  |
| 9   | Waktu pengerjaan untuk soal teori adalah 4 jam dan untuk soal praktikum adalah 8 jam.                                                                                                                                         |
| 10  | penilaian terhadap hasil skripsi meliputi penilaian dari hasil bimbingan dengan dosen pembimbing, penilaian dari hasil pengujian dalam kegiatan seminar skripsi, dan ujian sidang skripsi.                                    |

Hasil pengujian dengan RAGAS menunjukkan bahwa *chatbot* berbasis RAG yang dikembangkan untuk Universitas Catur Insan Cendekia memiliki performa yang baik dalam memahami dan mengambil konteks dari sumber data, serta memberikan jawaban yang relevan dengan sumber data. Namun, terdapat beberapa area yang memerlukan perbaikan, terutama dalam hal kebenaran jawaban. Secara keseluruhan, *chatbot* ini telah menunjukkan potensinya sebagai alat yang efisien dan efektif untuk meningkatkan akses informasi bagi mahasiswa dan masyarakat umum.

Tabel 3. Hasil Evaluasi RAGAS

| Pertanyaan | Context precision | Faithfulness | Answer relevancy | Context recall | Answer correctness | Answer similarity |
|------------|-------------------|--------------|------------------|----------------|--------------------|-------------------|
| 1          | 100%              | 100%         | 100%             | 100%           | 70%                | 79%               |
| 2          | 100%              | 100%         | 99%              | 100%           | 99%                | 98%               |
| 3          | 100%              | 100%         | 97%              | 100%           | 66%                | 93%               |
| 4          | 33%               | 100%         | 93%              | 100%           | 23%                | 91%               |
| 5          | 100%              | 100%         | 99%              | 100%           | 100%               | 99%               |
| 6          | 100%              | 100%         | 92%              | 100%           | 74%                | 94%               |
| 7          | 100%              | 100%         | 97%              | 100%           | 85%                | 100%              |

| Pertanyaan | Context precision | Faithfulness | Answer relevancy | Context recall | Answer correctness | Answer similarity |
|------------|-------------------|--------------|------------------|----------------|--------------------|-------------------|
| 8          | 100%              | 100%         | 91%              | 100%           | 97%                | 88%               |
| 9          | 100%              | 100%         | 96%              | 100%           | 74%                | 96%               |
| 10         | 100%              | 100%         | 91%              | 100%           | 24%                | 95%               |
| Rata-rata  | 93%               | 100%         | 96%              | 100%           | 71%                | 93%               |

#### 4. Kesimpulan

Penelitian ini menunjukkan bahwa penggunaan *chatbot* berbasis AI efektif dalam mengatasi kesulitan mahasiswa dan masyarakat umum dalam mengakses informasi di Universitas Catur Insan Cendekia. Metode *Retrieval Augmented Generation* memungkinkan *chatbot* memberikan jawaban yang lebih spesifik, relevan, dan akurat dibandingkan dengan metode tradisional. Hasil pengujian menunjukkan skor *context precision* 93%, *faithfulness* 100%, *answer relevancy* 96%, *context recall* 100%, *answer correctness* 71%, dan *answer similarity* 93%.

Untuk meningkatkan efektivitas dan efisiensi layanan informasi, beberapa saran yang dapat dipertimbangkan adalah: melakukan evaluasi berkelanjutan terhadap kinerja *chatbot*; mengadopsi teknik RAG terbaru seperti *RAG Agent*; menggunakan teknik *parsing* yang mengenali struktur dokumen seperti data tabel dengan *tool* seperti *LlamaParse*; melakukan *pre-processing* pada dokumen untuk menghilangkan noise dan mengatur format dokumen sebelum *parsing*; serta menambahkan fitur yang memungkinkan pengguna menyesuaikan preferensi mereka, topik informasi yang sering dicari.

#### Daftar Rujukan

- [1] F. Febriyanto, F. Pramudya, P. Sokibi, K. Kusnadi, and R. Taufiq Subagio, "Sosialisasi Penerapan Sistem Informasi dan Pendataan Pengunjung pada Keraton Kasepuhan Kota Cirebon," *JPUCIC*, vol. 1, 2022.
- [2] M. Asfi, Kusnadi, and Kristianto, "Analisis Kepuasan Alumni Terhadap Pelayanan Akademik dengan Metode Importance Performance Analysis Berbasis Web (Studi Kasus: Universitas CIC)," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 4, pp. 1007–1015, 2020, doi: 10.30865/mib.v4i4.2343.
- [3] D. Bill and T. Eriksson, "Fine-tuning a LLM using Reinforcement Learning from Human Feedback for a Therapy Chatbot Application," 2023.
- [4] L. Indahsari, K. Kusnadi, and T. E. Putri, "Rancang Bangun LINE Chatbot Informasi dan Edukasi Kesehatan Mental Menggunakan Algoritma Jaro Winkler," *Jurnal Eksplorasi Informatika*, vol. 10, no. 2, pp. 68–79, Mar. 2021, doi: 10.30864/eksplorasi.v10i2.428.
- [5] L. Monigatti, "Retrieval-Augmented Generation (RAG): From Theory to LangChain Implementation," Towards Data Science. Accessed: Feb. 21, 2024. [Online]. Available: <https://towardsdatascience.com/retrieval-augmented-generation-rag-from-theory-to-langchain-implementation-4e9bd5f6a4f2>
- [6] Y. Gao et al., "Retrieval-Augmented Generation for Large Language Models: A Survey," *arXiv e-prints*, p. arXiv:2312.10997, Dec. 2023, doi: 10.48550/arXiv.2312.10997.
- [7] P. Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," *arXiv e-prints*, p. arXiv:2005.11401, May 2020, doi: 10.48550/arXiv.2005.11401.
- [8] Z. Feng, X. Feng, D. Zhao, M. Yang, and B. Qin, "Retrieval-Generation Synergy Augmented Large Language Models," *arXiv e-prints*, p. arXiv:2310.05149, Oct. 2023, doi: 10.48550/arXiv.2310.05149.
- [9] R. S. Pressman, *Software Engineering: A Practitioner's Approach*, 7th ed. New York: McGraw-Hill, 2010.
- [10] U. Jamil, "Retrieval Augmented Generation (RAG) Explained: Embedding, Sentence BERT, Vector Database (HNSW)," Youtube. Accessed: May 13, 2024. [Online]. Available: [https://www.youtube.com/watch?v=rhZgXNdhWDY&list=P\\_LilBdZxzK8rTnz7hpfxXNeSro15uGw\\_cO&index=3](https://www.youtube.com/watch?v=rhZgXNdhWDY&list=P_LilBdZxzK8rTnz7hpfxXNeSro15uGw_cO&index=3)
- [11] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," *CoRR*, vol. abs/1908.10084, 2019, [Online]. Available: <http://arxiv.org/abs/1908.10084>
- [12] F. Arasyi, "indo-sentence-bert: Sentence Transformer for Bahasa Indonesia with Multiple Negative Ranking Loss," huggingface. Accessed: Mar. 09, 2024. [Online]. Available: <https://huggingface.co/firqaaa/indo-sentence-bert-base>
- [13] P. Sai, "The Ultimate Guide on Retrieval Strategies – RAG (part-4)," Chatgen.ai. Accessed: May 06, 2024. [Online]. Available: <https://chatgen.ai/blog/the-ultimate-guide-on-retrieval-strategies-rag-part-4/>