



Tersedia Online : <http://e-journals.unmul.ac.id/>

ADOPSI TEKNOLOGI DAN SISTEM INFORMASI (ATASI)

Alamat Jurnal : <http://e-journals2.unmul.ac.id/index.php/atasi/index>



## Pengembangan Chatbot WhatsApp Berbasis *Retrieval-Augmented Generation (RAG)* dan OpenAI untuk Penyediaan Jawaban Akurat dari Dokumen Aturan Organisasi

Mohamad Imaduddin <sup>1)\*</sup>, Rudi Susanto <sup>2)</sup>, Afu Ichsan Pradana <sup>3)</sup>

<sup>1)</sup> Teknik Informatika, Fakultas Ilmu Komputer, Universitas Duta Bangsa Surakarta

<sup>2,3)</sup> Fakultas Ilmu Komputer, Universitas Duta Bangsa Surakarta

E-Mail : [imaduddin.mohamad@gmail.com](mailto:imaduddin.mohamad@gmail.com) <sup>1)</sup>; [rudi\\_susanto@udb.ac.id](mailto:rudi_susanto@udb.ac.id) <sup>2)</sup>; [afu\\_ichsan@udb.ac.id](mailto:afu_ichsan@udb.ac.id) <sup>3)</sup>;

### ARTICLE INFO

#### Article history:

Received : 30-06-2025

Revised : 11-07-2025

Accepted : 11-10-2025

Available online : 29-04-2026

#### Keywords:

Chatbot

OpenAI

Organizational Rules

RAG

WhatsApp

### ABSTRACT

The scattered and unstructured nature of organizational regulatory documents makes it difficult for civil servants (ASN) to obtain information quickly and accurately. This study develops a WhatsApp chatbot based on Retrieval-Augmented Generation (RAG) and OpenAI language models to generate relevant automated answers from organizational documents. The system is built with two main modules: a WhatsApp gateway and the RAG system, each consisting of backend and frontend components. The WhatsApp backend handles session authentication as well as message transmission and reception, while the frontend monitors connection status and message activity in real time. The RAG backend processes user questions by performing semantic searches on documents that have been converted into vector representations, while the frontend allows admins to manage documents automatically. The two systems are integrated using a specific hashtag pattern and inter-module communication via API and Socket.IO. Evaluation using F1 Score, BLEU Score, and ROUGE-L metrics shows strong system performance, with an average F1 Score of 0.80, BLEU Score of 0.54, ROUGE-L Precision of 0.76, Recall of 0.88, and F1 Score of 0.81. These results demonstrate that integrating RAG and OpenAI technologies in a WhatsApp chatbot is effective for providing automated and efficient access to personnel information.

### ABSTRAK

Akses dokumen aturan yang tersebar dan tidak terstruktur menyulitkan ASN memperoleh informasi secara cepat dan akurat. Penelitian ini mengembangkan chatbot WhatsApp berbasis RAG dan model bahasa OpenAI untuk menghasilkan jawaban otomatis yang relevan dari dokumen organisasi. Sistem dikembangkan dalam dua modul utama, yaitu gateway WhatsApp dan sistem RAG, masing-masing terdiri atas backend dan frontend. Backend WhatsApp menangani autentikasi sesi serta pengiriman dan penerimaan pesan, sedangkan frontend berfungsi memantau status koneksi dan aktivitas pesan secara real-time. Backend RAG memproses pertanyaan pengguna dengan melakukan pencarian semantik terhadap dokumen yang telah dikonversi menjadi representasi vektor, sementara frontend memungkinkan admin mengelola dokumen secara otomatis. Integrasi kedua sistem dilakukan melalui tagar khusus dengan komunikasi antarmodul menggunakan API dan Socket.IO. Evaluasi dengan metrik F1 Score, BLEU Score, dan ROUGE-L menunjukkan performa sistem yang baik. Rata-rata F1 Score sebesar 0,80, BLEU Score 0,54, ROUGE-L Precision 0,76, Recall 0,88, dan F1 Score 0,81. Dengan demikian, integrasi teknologi RAG dan OpenAI dalam chatbot WhatsApp terbukti efektif sebagai solusi akses informasi kepegawaian secara otomatis dan efisien.

2026 Adopsi Teknologi dan Sistem Informasi (ATASI) with CC BY SA license.

## 1. PENDAHULUAN

Dalam sebuah organisasi, keberadaan aturan berfungsi sebagai pedoman operasional yang mengatur perilaku dan tindakan setiap anggota untuk mencapai tujuan bersama. Aturan tidak hanya menetapkan batasan dan standar perilaku, tetapi juga menciptakan kepastian dalam pengambilan keputusan. Secara umum, aturan organisasi terbagi menjadi dua: aturan internal dan aturan eksternal. Aturan internal ditujukan untuk mengatur kegiatan anggota organisasi, seperti kebijakan jam kerja, tata tertib, serta pengaturan hak dan kewajiban pegawai. Sementara itu, aturan eksternal ditujukan kepada pihak di luar organisasi yang berinteraksi dengan organisasi tersebut, seperti prosedur layanan, tata cara pengurusan dokumen, syarat dan ketentuan penjualan, atau aturan pengembalian barang. Aturan-aturan tersebut menjadi panduan dalam menjalankan organisasi secara tertib dan terarah untuk mencapai kinerja yang optimal. Lingkungan kerja dan keberadaan aturan organisasi juga terbukti berpengaruh signifikan terhadap peningkatan kinerja karyawan, sebagaimana ditunjukkan dalam model regresi  $Y = 11,771 + 0,353 X_1 + 0,469 X_2$  (Manihuruk, 2024, p. 74).

Permasalahan sering terjadi terkait dengan akses dan pemahaman terhadap aturan yang berlaku. Dokumen aturan organisasi banyak tersimpan dalam format tidak terstruktur, tersebar di berbagai sumber, atau sulit diakses dengan cepat. Pengguna yang membutuhkan informasi sering kali harus mencarinya secara manual atau menghubungi bagian terkait untuk klarifikasi. Proses ini tidak hanya memakan waktu, tetapi juga dapat menyebabkan ketidakkonsistenan dalam interpretasi aturan yang berlaku. Selain itu, sistem pencarian informasi dalam organisasi umumnya bersifat statis dan tidak mendukung interaksi dinamis yang memungkinkan pengguna memperoleh jawaban yang spesifik sesuai kebutuhan mereka.

Permasalahan serupa juga terjadi dalam pengelolaan informasi terkait regulasi Aparatur Sipil Negara (ASN). Dokumen seperti Undang-Undang, Peraturan Pemerintah, Peraturan Presiden, serta kebijakan teknis dari instansi seperti Kementerian PANRB dan Badan Kepegawaian Negara (BKN) tersebar di berbagai sumber. Kebutuhan akan sistem yang mampu menyajikan jawaban akurat dari dokumen-dokumen tersebut menjadi semakin mendesak, terutama bagi instansi pemerintah yang ingin meningkatkan efisiensi layanan dan akses informasi.

Teknologi kecerdasan buatan (*Artificial Intelligence*), khususnya chatbot berbasis metode *Retrieval-Augmented Generation* (RAG), menjadi salah satu solusi potensial. Metode ini menggabungkan proses pencarian informasi (*retrieval*) dengan kemampuan generatif untuk membentuk jawaban yang kontekstual dan akurat. Penggunaan pendekatan ini terbukti mampu meningkatkan kualitas jawaban pada sistem tanya-jawab berbasis dokumen, seperti pada pengembangan chatbot untuk diagnosis radiologi dengan tingkat akurasi yang tinggi (Fink et al., 2025).

Penggunaan chatbot dalam layanan informasi juga telah banyak diadopsi oleh institusi. WhatsApp chatbot untuk layanan perpustakaan UIN Walisongo berhasil dikembangkan dan terbukti meningkatkan efisiensi pustakawan dalam menjawab pertanyaan pengguna (Ahwan et al., 2024). Di sisi lain, integrasi NLP dan teknologi *text-to-speech* terbukti mampu meningkatkan pengalaman pengguna dalam berinteraksi dengan chatbot secara signifikan (Saputra & Harefa, 2025).

WhatsApp merupakan salah satu platform komunikasi yang paling banyak digunakan secara global. Mengintegrasikan chatbot dengan WhatsApp akan memberikan kenyamanan dan kemudahan akses kepada pengguna tanpa perlu menginstal aplikasi tambahan. Dengan antarmuka yang familiar, pengguna dapat memperoleh informasi secara cepat dan efisien. WhatsApp sebagai platform yang telah dikenal luas berpotensi meningkatkan aksesibilitas terhadap informasi aturan organisasi (Puthiyaveetil, 2024).

Diharapkan, sistem ini dapat mempercepat akses informasi, mengurangi beban administrasi, serta meningkatkan efisiensi komunikasi dalam organisasi. Untuk memastikan bahwa sistem yang dikembangkan mampu memberikan jawaban yang relevan dan berkualitas, penelitian ini menggunakan metode evaluasi yang objektif melalui pengujian fungsional dan metrik evaluasi otomatis, seperti F1 Score, BLEU Score, dan ROUGE-L Score. Metrik-metrik ini digunakan untuk mengukur kinerja chatbot berdasarkan kesesuaian jawaban dengan referensi dari dokumen aturan. Dengan latar belakang tersebut, penelitian ini bertujuan untuk membangun sistem chatbot berbasis WhatsApp yang mampu menjawab pertanyaan secara akurat dan kontekstual dari dokumen aturan organisasi, dengan menerapkan metode *Retrieval-Augmented Generation* (RAG) dan teknologi model bahasa dari OpenAI. Fokus utama diarahkan pada pemrosesan kebijakan yang berkaitan dengan Aparatur Sipil Negara (ASN).

## 2. TINJAUAN PUSTAKA

WhatsApp memiliki potensi tinggi sebagai media interaksi chatbot berbasis RAG karena mampu menyampaikan informasi secara cepat, interaktif, dan mudah dipahami melalui fitur seperti teks, gambar, suara, dan video. Selain itu, notifikasi langsung dan efisiensi biaya menjadikan WhatsApp sarana edukasi dan komunikasi yang efektif dalam layanan publik berbasis dokumen (Mujib et al., 2025).

Dalam konteks perpustakaan, WhatsApp chatbot telah dimanfaatkan untuk meningkatkan efisiensi layanan. WhatsApp Chatbot menjadi solusi bagi pustakawan dalam menjawab pertanyaan pemustaka secara otomatis, menghemat waktu, dan mempercepat respons terhadap permintaan informasi. Kehadiran chatbot ini juga memperluas jangkauan layanan tanpa menambah beban kerja staf secara langsung (Ahwan et al., 2024).

Selain penggunaan pada layanan informasi umum, pengembangan teknologi chatbot juga diarahkan untuk meningkatkan kualitas interaksi. Implementasi *Natural Language Processing* (NLP) yang dikombinasikan dengan fitur *Text-to-Speech* (TTS) terbukti dapat meningkatkan kualitas komunikasi, menjadikan interaksi lebih alami, serta memperbaiki pemahaman chatbot terhadap maksud pengguna. Chatbot berbasis AI juga mampu memberikan

\*) Corresponding Author

<https://doi.org/10.30872/atasi.v5i1.3472>

2026 Adopsi Teknologi dan Sistem Informasi (ATASI) with CC BY SA license.

respons yang konsisten, bersifat personal, dan tersedia selama 24 jam, sehingga berkontribusi pada peningkatan pengalaman serta kepuasan pengguna. Studi ini merekomendasikan pengembangan lebih lanjut, terutama dalam hal peningkatan akurasi pemrosesan bahasa alami dan integrasi AI yang lebih luas (Saputra & Harefa, 2025).

Penggunaan chatbot juga telah diterapkan dalam bidang pariwisata. Chatbot berbasis *Natural Language Processing* (NLP) yang dikembangkan untuk Kota Serang mampu menyajikan informasi wisata secara efektif dan mendukung pengalaman berwisata yang efisien serta mudah diakses (Hadinata & Stianingsih, 2025).

Lebih jauh lagi, chatbot berbasis AI menunjukkan potensi dalam mendukung pengembangan fungsi-fungsi kognitif pengguna. Chatbot berbasis AI berpotensi meningkatkan keterampilan eksekutif seperti memori kerja, perhatian, regulasi emosi, dan pengambilan keputusan. Interaksi yang dirancang secara terstruktur dan adaptif dapat membantu pengguna dalam mengelola fokus, merespons emosi secara terkendali, serta mengambil keputusan secara lebih reflektif. Meskipun demikian, diperlukan studi lanjutan dengan desain metodologis yang lebih kuat, ukuran sampel yang lebih besar, dan pengujian jangka panjang untuk memahami dampak sebenarnya terhadap pengembangan fungsi eksekutif (Pergantis et al., 2025).

Pada sisi teknis, pendekatan *Retrieval-Augmented Generation* (RAG) yang dikombinasikan dengan model bahasa besar seperti OpenAI telah terbukti meningkatkan relevansi konteks jawaban chatbot. Chatbot berbasis RAG mampu memberikan respons yang akurat melalui integrasi proses web scraping, pemrosesan teks, pencarian semantik, dan penyusunan jawaban dari LLM. Keunggulan pendekatan ini terletak pada kemampuannya mengakses informasi terbaru dari berbagai sumber secara real-time sambil mempertahankan koherensi jawaban (Pokhrel et al., 2024).

Efektivitas RAG juga terbukti dalam aplikasi domain khusus seperti bidang kesehatan. Penggunaan RAG dapat meningkatkan akurasi dan transparansi chatbot berbasis GPT-4 dalam diagnosis radiologi darurat. Studi ini mengembangkan TraumaCB, chatbot yang diperkuat dengan pengetahuan dari jurnal *RadioGraphics*, sehingga menghasilkan jawaban yang lebih akurat dan terpercaya dibandingkan GPT-4 standar. TraumaCB mencapai akurasi 100% dalam diagnosis, 96% dalam klasifikasi cedera, dan 87% dalam penilaian keparahan. Temuan ini menegaskan bahwa integrasi RAG tidak hanya memperkuat aspek faktual dalam respons chatbot, tetapi juga meningkatkan kepercayaan pengguna dalam konteks kritis seperti layanan medis berbasis AI (Fink et al., 2025).

Secara umum, RAG telah menjadi bagian penting dalam ekosistem pencarian modern. Pendekatan ini menggantikan metode tradisional dengan kemampuan menghasilkan ringkasan informatif dan teratribusi secara real-time (Pradeep et al., 2024). *Retrieval-Augmented Generation* (RAG) meningkatkan kinerja model bahasa besar (LLM) dengan mengambil informasi eksternal yang relevan, sehingga menghasilkan jawaban yang lebih akurat dan selalu diperbarui. Pendekatan ini cocok untuk kebutuhan pendidikan yang menuntut informasi terpercaya dan mutakhir (Li et al., 2025).

### 3. METODE PENELITIAN

Penelitian ini bertujuan mengembangkan sistem chatbot yang mampu menjawab pertanyaan berbasis dokumen aturan organisasi secara otomatis melalui platform WhatsApp. Sistem ini menggunakan pendekatan *Retrieval-Augmented Generation* (RAG), yang menggabungkan teknik pencarian informasi dan penyusunan jawaban berbasis model bahasa OpenAI. RAG dikembangkan sebagai metode untuk menjawab pertanyaan berbasis dokumen dengan mengintegrasikan *retrieval* semantik dan model *generatif*, sehingga meningkatkan relevansi dan akurasi jawaban secara signifikan (Lewis et al., 2020).

Pengembangan sistem dilakukan dengan pendekatan rekayasa perangkat lunak model iteratif dan inkremental, yang memungkinkan pembangunan bertahap, pengujian awal, serta fleksibilitas dalam penyesuaian fitur sesuai kebutuhan pengguna. Evaluasi akhir dilakukan untuk mengukur akurasi jawaban dan keandalan sistem dalam lingkungan pengguna yang sesungguhnya.

#### A. Studi Literatur dan Pengumpulan Data

Studi literatur dilakukan untuk mendalami konsep *Retrieval-Augmented Generation* (RAG), *Natural Language Processing* (NLP), arsitektur sistem chatbot, serta integrasi aplikasi dengan platform WhatsApp. Literatur yang dikaji meliputi artikel ilmiah, dokumentasi teknis, dan referensi resmi terkait pengolahan bahasa alami dan teknologi chatbot. Salah satu referensi utama menjelaskan arsitektur dasar RAG dan penerapannya untuk tugas-tugas NLP berbasis pengetahuan (Lewis et al., 2020).

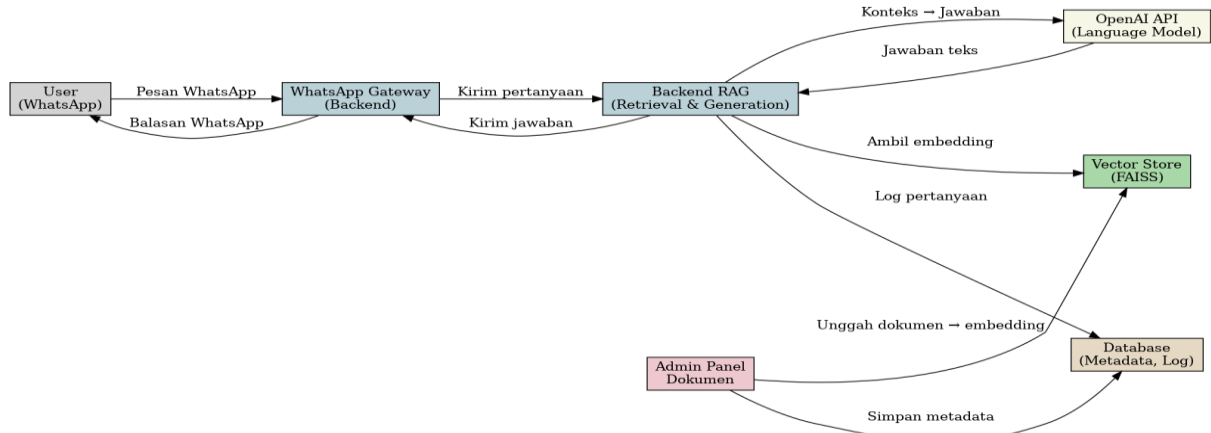
Selain itu, dokumen aturan organisasi seperti peraturan pemerintah, peraturan badan, dan surat edaran dikumpulkan sebagai sumber pengetahuan utama. Dokumen-dokumen tersebut diolah ke dalam format teks terstruktur, kemudian diproses melalui tahap *embedding* untuk memungkinkan pencarian semantik yang efisien oleh sistem.

#### B. Arsitektur dan Alur Sistem

Sistem yang dikembangkan terdiri atas lima komponen utama, yaitu: (1) Backend WhatsApp Gateway yang berfungsi untuk menerima dan mengirim pesan melalui WhatsApp, (2) Frontend WhatsApp Gateway sebagai dashboard untuk memantau status koneksi dan histori pesan secara real-time, (3) Backend *Retrieval-Augmented Generation* (RAG) yang menangani proses pencarian dokumen dan penyusunan jawaban, (4) Frontend Admin Dokumen yang digunakan untuk mengelola dokumen aturan yang menjadi sumber pengetahuan sistem, serta (5) penyimpanan vektor berbasis FAISS untuk mendukung pencarian semantik. FAISS (*Facebook AI Similarity*

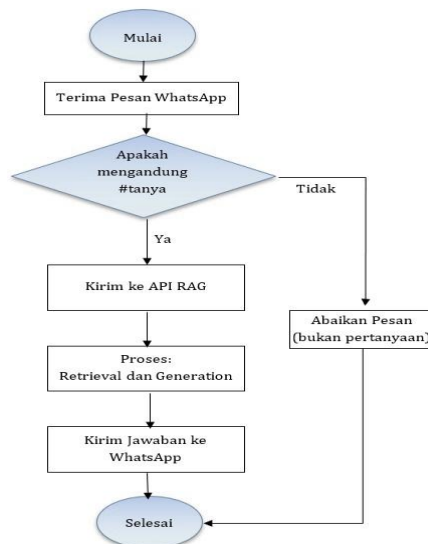
*Search*) merupakan pustaka pencarian berbasis vektor yang dirancang untuk skala besar dan efisien, serta telah banyak digunakan dalam implementasi sistem RAG modern (Johnson et al., 2019).

Kelima komponen ini saling terhubung membentuk alur kerja sistem chatbot secara terpadu. Struktur dan hubungan antar komponen ditampilkan pada Gambar 1, yang memperlihatkan aliran data dari pengguna hingga sistem menghasilkan dan mengirimkan kembali jawaban. Alur dimulai saat pengguna mengirim pesan melalui WhatsApp dengan tagar khusus seperti #tanya. Backend WhatsApp Gateway memfilter pesan dan meneruskannya ke API backend RAG jika sesuai pola. Backend RAG kemudian melakukan pencarian dokumen relevan dari vector store (FAISS) dan menghasilkan jawaban menggunakan model bahasa OpenAI. Jawaban yang dihasilkan dikirim kembali ke pengguna melalui jalur komunikasi yang sama.



Gambar 1. Diagram Komponen Sistem Chatbot WhatsApp RAG

Seluruh proses komunikasi antarkomponen dan logika pengambilan keputusan di dalam sistem divisualisasikan secara lebih rinci dalam Gambar 2 dalam bentuk flowchart. Diagram ini membantu memperjelas urutan kerja sistem dan peran masing-masing modul dalam memproses pertanyaan pengguna hingga menghasilkan respons yang sesuai.



Gambar 2. Flowchart alur proses chatbot WhatsApp dalam menangani pertanyaan pengguna

Alur interaksi antar komponen dalam sistem tanya-jawab ini berjalan secara berurutan, dimulai dari pengguna yang mengirimkan pesan melalui WhatsApp dengan awalan tagar #tanya. Pesan tersebut diteruskan dari WhatsApp Gateway ke backend sistem RAG untuk diproses dan dijawab. Urutan komunikasi tersebut divisualisasikan pada Gambar 3.



Gambar 3. Sequence diagram proses interaksi chatbot WhatsApp berbasis RAG

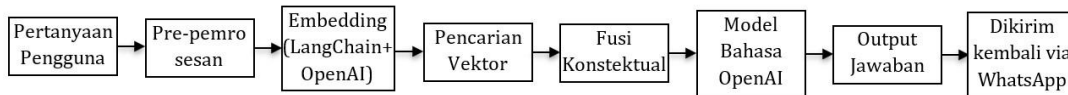
\*) Corresponding Author

<https://doi.org/10.30872/atasi.v5i1.3472>

2026 Adopsi Teknologi dan Sistem Informasi (ATASI) with CC BY SA license.

Selain arsitektur dan interaksi antarkomponen, proses internal sistem *Retrieval-Augmented Generation* (RAG) juga dijelaskan pada Gambar 4. Diagram ini menggambarkan tahapan utama dalam pipeline RAG. Pertama, sistem menerima input berupa teks pertanyaan dari pengguna. Pertanyaan ini kemudian diproses dan diubah menjadi representasi vektor menggunakan model embedding. Vektor tersebut digunakan untuk melakukan pencarian semantik terhadap dokumen yang telah disimpan dalam vector store berbasis FAISS.

Dokumen relevan yang ditemukan digabungkan secara kontekstual dengan pertanyaan pengguna (*contextual fusion*) untuk membentuk masukan yang utuh bagi model bahasa, sebagaimana diterapkan dalam RAG-Fusion (Rackauckas, 2024). Konteks ini kemudian diberikan sebagai masukan kepada model bahasa OpenAI untuk menghasilkan jawaban. Jawaban yang dihasilkan dikemas dan dikirimkan kembali ke pengguna melalui kanal WhatsApp. Proses ini memastikan bahwa jawaban yang diberikan tidak hanya relevan secara semantik, tetapi juga dikontekstualisasikan berdasarkan isi dokumen yang tersedia, sehingga meningkatkan akurasi dan keandalan sistem dalam menjawab pertanyaan berbasis aturan organisasi.



Gambar 4. Pipeline Proses Retrieval-Augmented Generation (RAG)

### C. Teknologi dan Implementasi

Sistem dikembangkan menggunakan bahasa pemrograman Python. Backend sistem tanya-jawab berbasis RAG dibangun menggunakan framework FastAPI, dengan LangChain sebagai pustaka utama untuk membangun pipeline *retrieval-augmented generation*. Dokumen yang digunakan dalam sistem diubah menjadi vektor embedding dan disimpan menggunakan FAISS untuk mendukung pencarian semantik.

Untuk komunikasi real-time antara sistem dan pengguna melalui WhatsApp, digunakan mekanisme berbasis Socket.IO. Data terkait pesan dan status koneksi disimpan dalam basis data relasional MySQL. Pengujian antarmuka pemrograman aplikasi (API) dilakukan menggunakan Postman, sedangkan antarmuka pengguna dirancang dengan tampilan sederhana dan responsif untuk memfasilitasi monitoring dan manajemen dokumen.

### D. Evaluasi

Evaluasi sistem dilakukan untuk menilai kinerja chatbot WhatsApp berbasis RAG dalam menjawab pertanyaan berdasarkan dokumen aturan organisasi. Evaluasi mencakup dua aspek utama, yaitu pengujian fungsional dan pengujian kualitas jawaban.

Pengujian fungsional bertujuan memastikan setiap komponen sistem, mulai dari WhatsApp Gateway, backend RAG, hingga antarmuka admin dokumen, beroperasi sesuai alur proses yang dirancang. Setiap interaksi diuji untuk menjamin integrasi antarkomponen berjalan lancar dan responsif.

Untuk evaluasi kualitas jawaban, digunakan tiga metrik evaluasi otomatis, yaitu F1 Score, BLEU Score, dan ROUGE-L Score. F1 Score mengukur keseimbangan antara precision dan recall, BLEU Score menilai kesamaan n-gram secara literal, sementara ROUGE-L Score mengevaluasi kesamaan berbasis Longest Common Subsequence (LCS) yang toleran terhadap parafrase.

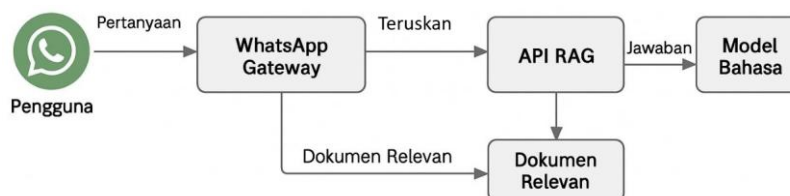
Dataset pengujian terdiri dari 30 pertanyaan yang disusun berdasarkan dokumen kepegawaian seperti peraturan pemerintah, peraturan BKN, PermenPANRB dan surat edaran terkait ASN. Jawaban yang dihasilkan sistem dibandingkan dengan jawaban referensi yang disusun secara manual untuk mengukur akurasi semantik dan kecocokan konteks.

## 4. HASIL DAN PEMBAHASAN

### A. Implementasi Sistem

Sistem chatbot dikembangkan berdasarkan pendekatan *Retrieval-Augmented Generation* (RAG) yang memanfaatkan layanan OpenAI untuk menjawab pertanyaan pengguna berdasarkan dokumen aturan organisasi (Lewis et al., 2020). Implementasi sistem mengacu pada komponen utama yang telah dijelaskan sebelumnya, yaitu WhatsApp Gateway (backend dan frontend), backend RAG, antarmuka admin dokumen, serta penyimpanan embedding menggunakan FAISS (Johnson et al., 2019).

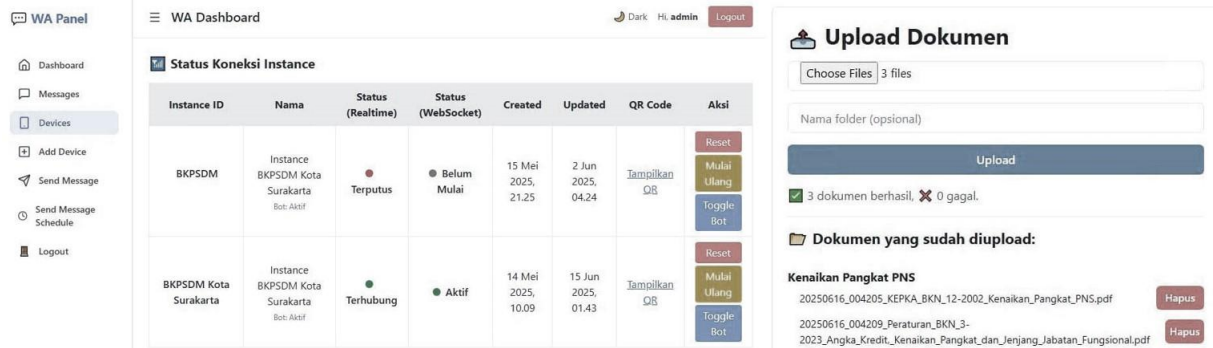
Seluruh komponen tersebut bekerja secara terintegrasi dalam alur proses tanya-jawab yang ditunjukkan pada Gambar 5. Diagram ini menggambarkan bagaimana pertanyaan yang dikirimkan oleh pengguna melalui WhatsApp diproses, mulai dari tahap penerimaan pesan, pencarian dokumen relevan, penyusunan jawaban, hingga pengiriman kembali jawaban tersebut ke pengguna.



Gambar 5. Diagram Alur Sistem Tanya-Jawab Chatbot WhatsApp

Backend WhatsApp Gateway dirancang untuk menangani koneksi WhatsApp secara real-time dalam berbagai instance sekaligus, serta mendukung komunikasi dua arah dengan sistem tanya-jawab. Backend RAG dikembangkan menggunakan Python dan FastAPI, dengan LangChain sebagai pipeline pemrosesan serta FAISS sebagai penyimpanan vektor. Antarmuka pengguna dirancang berbasis web yang responsif dan memanfaatkan Socket.IO untuk pembaruan status secara langsung. Data metadata disimpan dalam basis data MySQL.

Antarmuka untuk mengelola dokumen dan dashboard pemantauan WhatsApp Gateway ditampilkan pada Gambar 6. Antarmuka ini dirancang agar sederhana dan responsif, memungkinkan admin untuk memantau status koneksi setiap instance WhatsApp secara real-time, melihat histori pesan yang masuk dan keluar, serta mengunggah, menghapus, dan mengelola dokumen aturan yang menjadi sumber pengetahuan sistem. Desain antarmuka difokuskan pada kemudahan penggunaan agar proses pengelolaan dokumen dan pemantauan sistem dapat dilakukan dengan efisien.



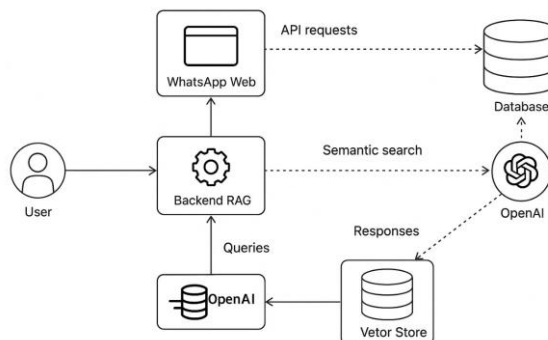
Gambar 6 Tampilan Antarmuka Monitoring WhatsApp Gateway dan Pengelolaan Dokumen

### B. Integrasi Sistem

Integrasi antara WhatsApp Gateway dan backend sistem tanya-jawab dilakukan melalui mekanisme REST API. Setiap pesan yang dikirimkan oleh pengguna melalui WhatsApp dan diawali dengan tagar khusus seperti #tanya akan dikenali dan disaring oleh gateway. Pesan tersebut kemudian diteruskan ke backend RAG untuk diproses lebih lanjut.

Di sisi backend, sistem melakukan dua tahap utama, yaitu pencarian potongan dokumen yang relevan menggunakan teknik retrieval berbasis vektor embedding, dan penyusunan jawaban secara otomatis menggunakan model bahasa dari OpenAI. Hasil jawaban yang dihasilkan kemudian dikembalikan ke pengguna melalui endpoint API dan disampaikan kembali melalui WhatsApp.

Seluruh proses integrasi ini melibatkan sejumlah komponen sistem yang saling berinteraksi, mulai dari antarmuka pengguna, gateway pesan, backend pemrosesan, hingga penyimpanan data. Hubungan antar komponen dan aliran data dalam sistem chatbot ini dapat dilihat pada Gambar 7.



Gambar 7. Arsitektur Sistem Chatbot Berbasis RAG

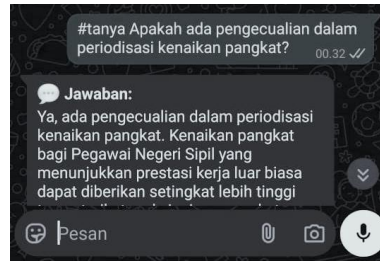
### C. Hasil Pengujian

Pengujian sistem dilakukan untuk mengevaluasi aspek fungsionalitas, akurasi jawaban terhadap dokumen referensi, serta efektivitas integrasi antar komponen sistem. Berdasarkan pengujian fungsional, sistem mampu menerima dan merespons pertanyaan pengguna dengan benar. Rata-rata waktu respons berkisar antara 2–5 detik. Gambar 8 memperlihatkan tangkapan layar hasil percakapan antara pengguna dan chatbot melalui WhatsApp.

\*) Corresponding Author

<https://doi.org/10.30872/atasi.v5i1.3472>

2026 Adopsi Teknologi dan Sistem Informasi (ATASI) with CC BY SA license.



Gambar 8. Tampilan Percakapan WhatsApp dengan Chatbot

Tabel 1 menyajikan beberapa contoh pertanyaan dari pengguna yang digunakan dalam proses pengujian sistem. Setiap pertanyaan dibandingkan dengan jawaban referensi dari dokumen dan hasil yang diberikan oleh chatbot. Dapat dilihat bahwa pada sebagian besar kasus, jawaban chatbot cukup sesuai dengan jawaban referensi, baik secara struktur maupun informasi inti yang disampaikan. Hal ini menunjukkan bahwa sistem mampu menangani pertanyaan eksplisit dengan cukup baik.

Tabel 1. Contoh Pertanyaan Pengujian dan Jawaban yang Dihasilkan Chatbot.

No	Pertanyaan	Jawaban Referensi	Jawaban Chatbot
1	Kapan saja periodisasi kenaikan pangkat ditetapkan dalam setahun?	Periodisasi kenaikan pangkat ditetapkan pada tanggal 1 Februari, 1 April, 1 Juni, 1 Agustus, 1 Oktober, dan 1 Desember setiap tahun.	Periodisasi kenaikan pangkat ditetapkan pada tanggal 1 Februari, 1 April, 1 Juni, 1 Agustus, 1 Oktober, dan 1 Desember setiap tahun.
2	Berapa lama minimal waktu dalam pangkat terakhir untuk dapat naik pangkat?	Minimal 1 tahun dalam pangkat terakhir.	Setidaknya 1 (satu) tahun dalam pangkat terakhir.
3	Siapa yang menetapkan keputusan kenaikan gaji berkala PPPK?	Keputusan kenaikan gaji berkala PPPK ditetapkan oleh PyB atau pejabat yang diberikan pendelegasian oleh PyB.	Keputusan kenaikan gaji berkala bagi PPPK ditetapkan oleh PyB atau pejabat yang diberikan pendelegasian oleh PyB.
4	Kapan keputusan kenaikan gaji berkala paling lambat ditetapkan?	Paling lambat 2 bulan sebelum kenaikan gaji mulai berlaku.	Kenaikan gaji berkala paling lambat ditetapkan 2 (dua) bulan sebelum kenaikan gaji berkala mulai berlaku.

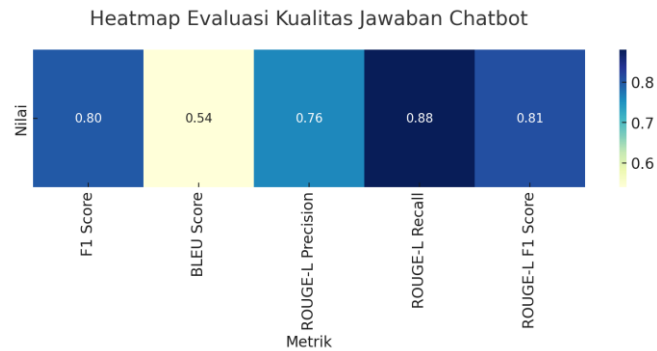
Evaluasi akurasi dilakukan dengan membandingkan jawaban chatbot terhadap jawaban referensi dari dokumen aturan. Berdasarkan 30 pertanyaan uji, diperoleh F1 Score sebesar 0,80, BLEU Score 0,54, dan ROUGE-L F1 Score 0,81. Hasil ini menunjukkan bahwa sistem mampu menghasilkan jawaban yang relevan dengan cukup baik. Rincian hasil evaluasi kuantitatif disajikan pada Tabel 2.

Tabel 2. Metrik Evaluasi dan Nilai yang Diperoleh.

Metrik Evaluasi	Nilai Rata-rata
<i>F1 Score</i>	0.80
<i>BLEU Score</i>	0.54
<i>ROUGE-L Precision</i>	0.76
<i>ROUGE-L Recall</i>	0.88
<i>ROUGE-L F1 Score</i>	0.81

Tabel 2 memperlihatkan nilai rata-rata dari metrik evaluasi yang digunakan untuk menilai kualitas jawaban chatbot. Nilai F1 Score sebesar 0,80 menunjukkan bahwa chatbot memiliki keseimbangan yang baik antara ketepatan dan kelengkapan informasi. BLEU Score sebesar 0,54 menunjukkan kesamaan literal kalimat yang cukup tinggi. Sementara ROUGE-L Recall yang mencapai 0,88 mengindikasikan bahwa sebagian besar informasi penting dari jawaban referensi berhasil ditangkap oleh chatbot.

Visualisasi hasil evaluasi dalam bentuk heatmap yang memperlihatkan perbandingan antar metrik secara intuitif ditampilkan pada Gambar 9. Heatmap ini menyajikan performa chatbot berdasarkan nilai F1 Score, BLEU Score, dan ROUGE-L dalam format visual berwarna untuk memudahkan analisis. Warna lebih gelap menunjukkan performa lebih tinggi.



Gambar 9. Heatmap Evaluasi Kualitas Jawaban Chatbot

Namun, kelemahan ditemukan pada pertanyaan dengan struktur negasi atau pengecualian, di mana sistem masih kurang tepat dalam memahami konteks. Temuan ini menjadi catatan penting untuk pengembangan sistem pada tahap berikutnya.

Hasil evaluasi sistem ini menunjukkan performa yang lebih tinggi dibandingkan studi Oro et al. (2023), yang menggunakan BLEU, ROUGE, dan F1 untuk menilai sistem RAG pada domain tanya-jawab publik. Perbedaan ini kemungkinan disebabkan oleh ruang lingkup pertanyaan yang lebih terfokus dan jumlah dokumen yang masih terbatas. Kondisi ini memungkinkan proses pencarian dan penyusunan jawaban berjalan lebih akurat dan tidak terlalu kompleks dibandingkan domain publik yang lebih beragam dan tidak terstruktur.

## 5. KESIMPULAN

Berdasarkan hasil evaluasi terhadap sistem chatbot berbasis *Retrieval-Augmented Generation* (RAG) yang diuji menggunakan 30 pertanyaan dari dokumen aturan kepegawaian, disimpulkan bahwa sistem mampu menghasilkan jawaban yang cukup relevan dan akurat. Evaluasi dilakukan menggunakan tiga metrik utama, yaitu F1 Score, BLEU Score, dan ROUGE-L Score, yang secara keseluruhan menunjukkan performa yang baik.

Rata-rata nilai F1 Score sebesar 0,80 mengindikasikan bahwa sistem memiliki keseimbangan yang baik antara *precision* dan *recall*. Hal ini menunjukkan bahwa chatbot mampu mengidentifikasi token penting dalam jawaban referensi sekaligus meminimalkan kesalahan isi. Metrik ini mencerminkan bahwa sistem memiliki pemahaman yang memadai terhadap konteks pertanyaan dan isi dokumen sumber.

Nilai BLEU Score sebesar 0,54 menunjukkan kesamaan literal antar n-gram antara jawaban chatbot dan jawaban referensi. Skor ini cukup baik untuk konteks sistem tanya jawab berbasis dokumen, meskipun BLEU tidak memperhitungkan sinonim atau parafrase, sehingga respons yang benar secara semantik namun berbeda secara redaksional dapat memperoleh nilai lebih rendah.

Rata-rata skor ROUGE-L F1 sebesar 0,81, dengan precision 0,76 dan recall 0,88, memperlihatkan bahwa sistem mampu menangkap informasi penting dari dokumen sumber. ROUGE-L yang berbasis *Longest Common Subsequence* (LCS) lebih toleran terhadap perbedaan struktur kalimat, sehingga mendukung analisis terhadap variasi jawaban yang valid. Perbedaan antara nilai precision dan recall menunjukkan bahwa sistem cenderung menghasilkan jawaban yang panjang atau mencakup informasi tambahan, yang dapat disesuaikan lebih lanjut berdasarkan konteks penggunaan.

Berdasarkan observasi selama proses pengujian, terdapat beberapa arah pengembangan yang disarankan. Pertama, sistem masih memiliki keterbatasan dalam memahami pertanyaan yang mengandung negasi atau pengecualian. Untuk mengatasi hal ini, dapat diterapkan teknik *prompt engineering* yang lebih spesifik atau menambahkan modul reranking berbasis logika (*logic-based reranker*) agar interpretasi terhadap konteks lebih akurat. Selain itu, mekanisme penyaringan atau re-ranking dokumen berdasarkan konteks pertanyaan dapat ditambahkan untuk mencegah dimasukkannya informasi yang tidak relevan dalam pembentukan jawaban.

Untuk pengembangan jangka panjang, pengujian lanjutan disarankan melibatkan dokumen tambahan dari berbagai jenis aturan, termasuk dokumen yang telah mengalami pembaruan, guna menguji kemampuan sistem dalam beradaptasi terhadap versi regulasi yang lebih mutakhir. Validasi jawaban juga perlu melibatkan pakar yang sesuai dengan bidang aturan yang diujikan, sehingga kualitas jawaban dapat dinilai secara lebih substantif dan kontekstual, tidak hanya berdasarkan metrik evaluasi otomatis. Selain itu, untuk menghadapi tantangan yang berkaitan dengan dokumen berdimensi panjang dan kompleks, pendekatan seperti LongRefiner dapat dipertimbangkan. Pendekatan ini mengadopsi penyaring hierarkis berbasis *multi-task learning* yang dirancang untuk meningkatkan akurasi proses ekstraksi informasi dalam sistem RAG.

## 6. DAFTAR PUSTAKA

Ahwan, M. A., Abiadhoh, N., Kusuma, A. B., & Alam, U. F. (2024). Pembangunan WhatsApp Chatbot sebagai layanan kecerdasan buatan di Perpustakaan UIN Walisongo Semarang. *Berkala Ilmu Perpustakaan dan Informasi*, 20(1), 119-132. <https://doi.org/10.22146/bip.v20i1.7043>

\*) Corresponding Author

<https://doi.org/10.30872/atasi.v5i1.3472>

2026 Adopsi Teknologi dan Sistem Informasi (ATASI) with CC BY SA license.

- Fink, A., Nattenmüller, J., Rau, S., et al. (2025). Retrieval-augmented generation improves precision and trust of a GPT-4 model for emergency radiology diagnosis and classification: a proof-of-concept study. *European Radiology*. <https://doi.org/10.1007/s00330-025-11445-z>
- Hadinata, W., & Stianingsih, L. (2025). Implementation of natural language processing on chatbot for tourist information services (Case study: Serang City). *INFOKUM*, 13(2), 227–236. <https://doi.org/10.58471/infokum.v13i02>
- Jin, J., Wang, W., Li, Z., & Sun, Y. (2024). LongRefiner: Efficient Long-Context RAG via Hierarchical Refinement and Multi-Task Learning. Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL). <https://github.com/ignorejijj/LongRefiner>
- Johnson, J., Douze, M., & Jégou, H. (2019). Billion-scale similarity search with GPUs. *IEEE Transactions on Big Data*. <https://arxiv.org/abs/1702.08734>
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Riedel, S. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. arXiv preprint arXiv:2005.11401. <https://arxiv.org/abs/2005.11401>
- Li, Z., Wang, Z., Wang, W., Hung, K., Xie, H., & Wang, F. L. (2025). Retrieval-Augmented Generation for Educational Application: A Systematic Survey. *Computers and Education: Artificial Intelligence*, 8, Article 100417. <https://doi.org/10.1016/j.caeai.2025.100417>
- Manihuruk, C. (2024). *Lingkungan kerja dan aturan organisasi terhadap kinerja karyawan*. *Jurnal Bisnis dan Ekonomi*, 2(1), 74–82. <https://doi.org/10.61597/jbe-ogzrp.v2i1.17>
- Mujib, H., Permatasari, A. P., Maryamah, A. M., & Walidah, A. N. (2025). *The education on the use of Mobile Maslahah for new customers through WhatsApp at Bank Jabar Banten Syariah Sub-Branch Office Banjar City*. <https://doi.org/10.59966/bnj0cz95>
- Oro, E., Granata, F. M., Lanza, A., Bachir, A., De Grandis, L., & Ruffolo, M. (2023). Evaluating Retrieval-Augmented Generation for Question Answering with Large Language Models. In *CEUR Workshop Proceedings (Vol. 3762)*. CEUR-WS.org. Retrieved from <https://iris.cnr.it/retrieve/5b673a72-40a9-4ea5-ad4f-4083fe1cecf1/495.pdf>
- Pergantis, P., Bamicha, V., Skianis, C., & Drigas, A. (2025). AI Chatbots and Cognitive Control: Enhancing Executive Functions Through Chatbot Interactions: A Systematic Review. *Brain Sci*, 15(1), 47. <https://doi.org/10.3390/brainsci15010047>
- Pokhrel, S., K C, B., & Shah, P. B. (2024). A practical application of retrieval-augmented generation for website-based chatbots: Combining web scraping, vectorization, and semantic search. *Journal of Trends in Computer Science and Smart Technology*, 6(4), 424–442. <https://doi.org/10.36548/jtcsst.2024.4.007>
- Pradeep, R., Thakur, N., Sharifymoghaddam, S., Zhang, E., Nguyen, R., Campos, D., Craswell, N., & Lin, J. (2024). Ragnarök: A reusable RAG framework and baselines for TREC 2024 retrieval-augmented generation track (arXiv:2406.16828). arXiv. <https://doi.org/10.48550/arXiv.2406.16828>
- Puthiyaveetil, A. S. (2024). How AI-Powered Chatbots Empowering the Telecom Industry to Deliver Exceptional Customer Experiences. *IEEE Xplore*. <https://doi.org/10.10157423>
- Rackauckas, Z. (2024). RAG-Fusion: A New Take on Retrieval-Augmented Generation. *International Journal on Natural Language Computing*, 13(1), 37–47. <https://arxiv.org/abs/2402.03367>
- Saputra, A. F., & Harefa, K. (2025). Penerapan Metode Natural Language Processing (NLP) dalam Implementasi Asisten Virtual Chatbot dengan Memanfaatkan API ChatGPT dan Gradio App. *JORAPI: Journal of Research and Publication Innovation*, 3(1), 1-15. <https://jurnal.portalpublikasi.id/index.php/JORAPI/article/view/1332>