Transformasi Layanan Informasi Berbasis *Chatbot Retrieval-Augmented Generation* dan *Large Language Model*

http://dx.doi.org/10.28932/jutisi.v12i2.13099

Riwayat Artikel
Received: 19 Agustus 2026 | *Final Revision*: 25 Maret 2026 | *Accepted*: 6 April 2026

Muhamad Komarudin ✉ #1, Chelly Sabrina*2, Yessy Mulyani#3, Puput Budi Wintoro#4

Program Studi Teknik Informatika, Universitas Lampung
Jl. Prof. Dr. Sumantri Brojonegoro No. 1, Bandar Lampung, Lampung 35145, Indonesia
[1]m.komarudin@eng.unila.ac.id
[3]yessi.mulyani@eng.unila.ac.id
[4]budi.wintoro@eng.unila.ac.id

* *PT Smart Multi Finance*
Foresta Business Loft 6 BSD City, Pagedangan, Tangerang, Banten, Indonesia
[2]chellysabrina1@gmail.com

✉Corresponding author: m.komarudin@eng.unila.ac.id

Abstrak — **Keterbukaan akses informasi publik diatur dalam Undang-Undang Nomor 14 Tahun 2008 tentang Keterbukaan Informasi Publik sebagai elemen utama untuk mewujudkan transparansi dan tata kelola pemerintahan yang baik. Di Universitas Lampung, Pejabat Pengelola Informasi dan Dokumentasi (PPID) menyediakan berbagai data publik melalui *website* resminya. Namun, pengguna sering mengalami kesulitan dalam menemukan informasi spesifik akibat tersebarnya dokumen serta keterbatasan fitur pencarian. Untuk mengatasi permasalahan tersebut, dikembangkan *chatbot* Telegram berbasis kecerdasan buatan menggunakan *Large Language Model* (LLM) Qwen2.5 VL 72B Instruct yang terintegrasi dengan arsitektur *Retrieval Augmented Generation* (RAG). Data diambil dari *website* PPID Universitas Lampung yang selanjutnya disusun menjadi dataset terstruktur sebanyak 375 pasangan pertanyaan–jawaban yang dikurasi dari 47 entri informasi publik. Data tersebut diproses melalui tahap tokenisasi, *embedding*, serta *indexing* menggunakan *vector database* FAISS untuk mendukung pencarian semantik. Evaluasi dilakukan dengan metrik RAGAS (*faithfulness, answer relevance,* dan *context recall*) dengan ambang batas minimal 0,8, serta perbandingan performa dengan LLM murni. Hasil pengujian menunjukkan bahwa sistem berhasil melampaui ambang batas RAGAS dengan rata-rata latensi respons yang sangat rendah yaitu 0,69 detik, serta memiliki akurasi lebih tinggi dibandingkan LLM murni dengan pengurangan signifikan terhadap halusinasi jawaban. Uji kegunaan menggunakan *Chatbot Usability Questionnaire* (CUQ) menghasilkan skor rata-rata 90,62 yang mengindikasikan pengalaman pengguna yang sangat baik. Penelitian ini menyimpulkan bahwa integrasi LLM dan RAG pada *chatbot* efektif meningkatkan akurasi, relevansi, dan kemudahan akses informasi publik secara *real-time*.**

Kata kunci— ***Chatbot; FAISS; Large Language Model; Retrieval Augmented Generation; Telegram.***

Transformation of Information Services Using Retrieval-Augmented Generation and Large Language Model

Abstract — ***The openness of access to public information is regulated in Law No. 14 of 2008 on Public Information Disclosure as a key element in realizing transparency and good governance. At the University of Lampung, Pejabat Pengelola Informasi dan Dokumentasi (PPID) provides various public data through its official website. However, users often face difficulties in finding specific information due to the dispersed nature of documents and the limitations of search features. To address this issue, a Telegram chatbot powered by Artificial Intelligence was developed using the Large Language Model (LLM) Qwen2.5 VL 72B Instruct integrated with the Retrieval-Augmented Generation (RAG) architecture. Data were collected from the University of Lampung's PPID website and organized into a structured dataset of 375 curated question–answer pairs derived from 47 public information entries. These were processed through tokenization, embedding, and indexing using the FAISS vector database to support semantic search. The evaluation was carried out using RAGAS metrics (faithfulness, answer relevance, and context recall) with a minimum threshold of 0.8, along with a performance comparison against a pure LLM. The results show that the developed system successfully surpassed the RAGAS minimum threshold with a remarkably low average response latency of 0.69 seconds and achieved higher accuracy compared to the pure LLM, with a significant reduction in answer hallucinations. Furthermore, usability testing using the Chatbot Usability Questionnaire (CUQ) produced an average score of 90.62, indicating an excellent user experience. This study concludes that integrating LLM and RAG in a chatbot effectively improves accuracy, relevance, and the ease of real-time access to public information.***

Keywords— ***Chatbot; FAISS; Large Language Model; Retrieval Augmented Generation; Telegram.***

I. PENDAHULUAN

Keterbukaan informasi publik merupakan elemen utama dalam mewujudkan tata kelola pemerintahan yang baik, transparan, dan akuntabel. Hal ini telah diatur dalam Undang-Undang Nomor 14 Tahun 2008 tentang Keterbukaan Informasi Publik (UU KIP) yang mewajibkan setiap badan publik untuk menyediakan informasi secara terbuka dan dapat diakses oleh masyarakat luas [1]. Prinsip ini berlaku tidak hanya pada instansi pemerintahan, tetapi juga mencakup organisasi non-pemerintah, institusi pendidikan, dan sektor swasta. Keterbukaan informasi ini memberikan ruang bagi partisipasi aktif publik dalam pengambilan keputusan serta pengawasan terhadap kebijakan dan pelaksanaan program, sehingga mendukung terciptanya tatanan yang lebih demokratis dan berorientasi pada kepentingan umum.

Sebagai institusi pendidikan tinggi, Universitas Lampung (Unila) memiliki tanggung jawab untuk melaksanakan prinsip keterbukaan informasi sesuai amanat UU KIP. Dalam implementasinya, Unila telah membentuk Pejabat Pengelola Informasi dan Dokumentasi (PPID) yang berperan sebagai pusat pengelolaan dan layanan informasi publik. PPID Unila bertugas menyediakan informasi institusional yang akurat, relevan, dan mudah diakses guna mendukung transparansi yang ada di universitas sekaligus meningkatkan kualitas layanan serta mendorong partisipasi publik [2].

Dalam mendukung keterbukaan informasi, PPID Unila menggunakan *website* sebagai sarana utama untuk menyampaikan informasi kepada publik. Melalui situs ini, khalayak dapat mengakses berbagai dokumen seperti laporan kinerja, kebijakan institusi, serta prosedur layanan. Namun, berdasarkan hasil observasi awal ditemukan bahwa tidak semua informasi tersedia dalam format yang langsung ditampilkan atau mudah diakses oleh pengguna. Informasi yang ada pada *website* PPID Unila tersebar dalam banyak dokumen dan struktur navigasi situs memerlukan pemahaman khusus agar pengguna dapat menemukan data yang relevan dengan kebutuhannya.

Temuan ini menunjukkan adanya kebutuhan akan inovasi dalam penyajian informasi yang lebih adaptif, interaktif, dan mudah dipahami oleh pengguna. Salah satu pendekatan potensial yang dapat diterapkan adalah penggunaan *chatbot* berbasis kecerdasan buatan atau *Artificial Intelligence* (AI). Saat ini, teknologi kecerdasan buatan tengah mengalami perkembangan pesat dan menjadi salah satu tren utama di berbagai bidang [3]. Hadirnya AI telah membawa banyak manfaat, terutama dalam mempercepat dan mempermudah akses terhadap informasi [4]. Dalam penelitian ini, *chatbot* dipilih sebagai solusi karena kemampuannya memberikan layanan pencarian informasi secara cepat, mudah, dan interaktif. Berbeda dengan metode pencarian manual di *website* yang mengharuskan pengguna membaca banyak dokumen atau halaman, *chatbot* memungkinkan pengguna memperoleh jawaban spesifik hanya dengan mengetikkan pertanyaan [5].

Penggunaan *chatbot* terbukti dapat meningkatkan efisiensi, hal ini disebutkan dalam penelit Xu et al. (2023) menunjukkan bahwa *chatbot* seperti ChatGPT mampu mengurangi waktu pencarian informasi hingga 65,20% dibandingkan *Google Search* dengan tingkat akurasi jawaban yang sebanding [6]. Arz Von Straussenburg (2023) juga menyatakan bahwa *chatbot* berbasis AI secara signifikan mempercepat akses terhadap informasi tanpa harus membaca banyak dokumen [7].

Untuk mencapai tingkat keakuratan yang tinggi dan menghindari kesalahan informasi (*hallucination*) yang umum terjadi pada *Large Language Model*, penelitian ini mengadopsi pendekatan *Retrieval Augmented Generation* (RAG). Arsitektur RAG menggabungkan kemampuan generatif dari LLM dengan sistem pencarian informasi eksternal yang relevan, sehingga respons yang dihasilkan lebih kontekstual dan didukung oleh sumber yang valid. Meskipun penggunaan chatbot berbasis RAG telah banyak dilaporkan dalam berbagai literatur terkini, terdapat research gap yang signifikan dalam optimalisasi akses informasi publik pada instansi pendidikan tinggi yang memiliki karakteristik dokumen kebijakan yang sangat spesifik,

statis, dan tersebar. Kebaruan (novelty) penelitian ini terletak pada pengembangan dataset *Question Answer* (QA) terstruktur sebanyak 375 entri yang dikurasi secara manual dari 47 entri informasi publik resmi Universitas Lampung, yang diintegrasikan dengan model Qwen2.5 VL 72B Instruct untuk menjamin akurasi tinggi tanpa halusinasi. Kontribusi utama penelitian ini adalah penyediaan kerangka kerja (framework) implementasi chatbot cerdas yang mampu mentransformasi dokumen publik menjadi basis pengetahuan interaktif dengan latensi rendah (0,69 detik) dan tingkat faithfulness mencapai 0,9629. Dengan demikian, penelitian ini tidak hanya menawarkan solusi teknis, tetapi juga memberikan model kebijakan layanan publik berbasis AI yang dapat direplikasi untuk meningkatkan transparansi informasi sesuai amanat UU Nomor 14 Tahun 2008 [8].

II. METODE PENELITIAN

A. Data Understanding

Tahap data understanding melibatkan identifikasi sumber data, pengumpulan data, dan eksplorasi awal terhadap konten yang akan digunakan dalam pengembangan chatbot. Sumber data berasal dari website PPID Universitas Lampung yang memuat berbagai dokumen dan informasi publik. Pengumpulan dilakukan secara manual (manual scraping) dengan mengakses halaman dan dokumen satu per satu, mencakup format HTML, PDF, dan gambar berisi teks yang kemudian dikonversi ke dalam format teks.

Hasil dari tahapan pengumpulan data mencakup 47 entri informasi yang dikelompokkan ke dalam delapan kategori utama, yaitu Tentang Kami, Layanan Informasi, Informasi Publik, PPID Pelaksana, My Unila, FAQ, Helpdesk, dan Kontak. Selanjutnya dilakukan eksplorasi untuk memahami karakteristik data, meliputi jenis informasi yang tersedia, struktur data, serta kelengkapan dan kejelasan konten. Tahap ini juga mencakup identifikasi potensi masalah seperti duplikasi, ketidakkonsistenan format, dan informasi yang memerlukan pembersihan. Hasil akhir berupa data terstruktur yang menjadi landasan dalam penyusunan dataset.

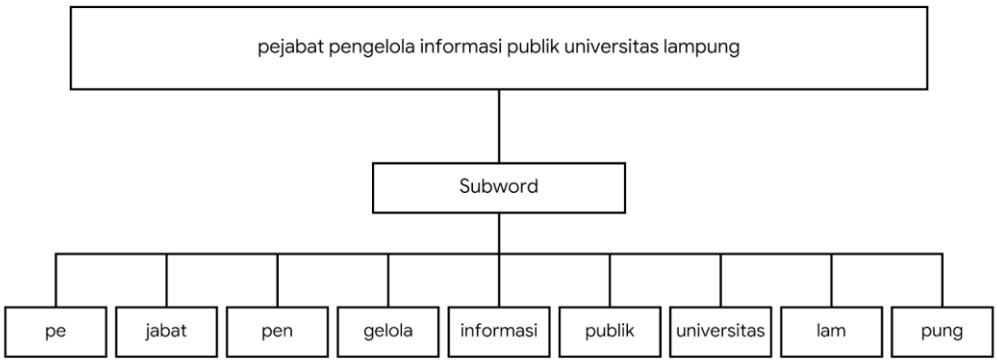
B. Pembuatan QA

Pembuatan *question–answer* (QA) merupakan tindak lanjut dari hasil eksplorasi pada tahap data understanding, di mana informasi publik yang telah diidentifikasi dan dikelompokkan kemudian diubah menjadi pasangan pertanyaan dan jawaban. Pertanyaan disusun berdasarkan konten dokumen resmi PPID Universitas Lampung dengan mempertimbangkan relevansi terhadap kebutuhan pengguna, meliputi topik seperti prosedur layanan informasi, regulasi, profil organisasi, dan data publik lainnya. Proses penyusunan dilakukan secara manual untuk memastikan keterkaitan langsung antara pertanyaan dan jawaban serta konsistensi format.

Dataset QA disimpan dalam format JSON, di mana setiap entri memiliki dua elemen utama, yaitu "*question*" dan "*answer*". Format JSON dipilih karena strukturnya sederhana, mudah dipahami, serta dapat diolah langsung oleh sistem pemrograman, sehingga sesuai untuk implementasi *chatbot*. Hasil akhir dari tahap ini adalah 375 pasangan QA yang disusun berdasarkan 47 entri informasi publik dari *website* PPID Unila. Setiap pasangan QA dirancang untuk merepresentasikan informasi yang paling relevan dengan memperhatikan konteks dan jenis konten dari masing-masing entri, sehingga siap digunakan pada proses *embedding* dan indexing dalam arsitektur RAG.

C. Tokenisasi

Setelah *dataset* dalam format QA selesai disusun, tahap berikutnya adalah melakukan proses *tokenisasi* untuk memecah teks menjadi satuan terkecil (*token*) yang dapat diproses oleh model sebagaimana ditunjukkan pada Gambar 1 [8]. *Tokenisasi* dilakukan menggunakan model *Sentence Transformer all-MiniLM-L6-v2* yang menerapkan teknik *tokenisasi* berbasis *WordPiece* [9]. Pemilihan model ini didasarkan pada kemampuannya menghasilkan representasi vektor yang mempertahankan makna semantik teks, sehingga efisien untuk pengolahan data berskala besar. Proses *tokenisasi* dilakukan menggunakan FAISS (*Facebook AI Similarity Search*) yang merupakan *database* vektor *open source* yang dirancang khusus untuk menyimpan, mencari, dan mengelola data vektor dari model AI [10].



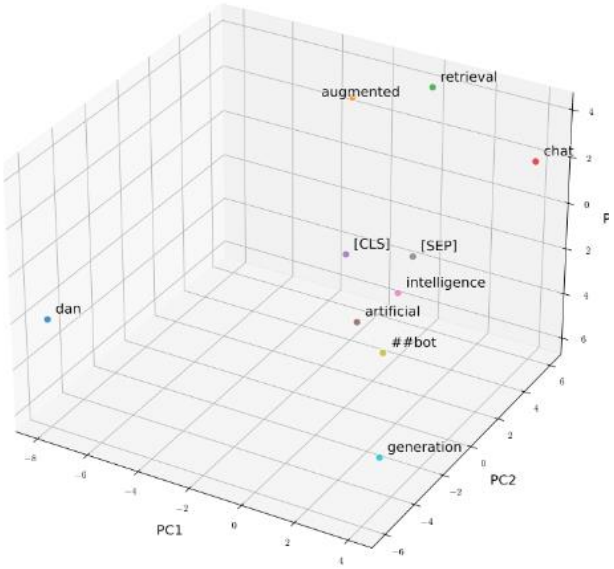
Gambar 1. Ilustras *Tokenisasi* Berbasis *WordPiece* [11]

D. Embedding

Embedding adalah proses yang digunakan untuk mengubah teks menjadi representasi numerik dalam bentuk angka. Proses ini penting karena algoritma *Machine Learning (ML)* tidak dapat memproses data dalam bentuk *string* atau teks biasa, sehingga diperlukan angka sebagai *input* [12]. Dalam konteks *Natural Language Processing* (NLP), *embedding* merepresentasikan kata-kata dalam bentuk vektor berdimensi tinggi yang dirancang untuk membantu model AI memahami, memproses, dan menganalisis teks. Setiap kata direpresentasikan sebagai vektor dalam ruang berdimensi tinggi, di mana kata-kata dengan makna atau konteks serupa memiliki posisi yang berdekatan [13]. Proses *tokenization* dan *embedding* dilakukan menggunakan model *all-MiniLM-L6-v2* untuk merepresentasikan teks ke dalam vektor multidimensi. Representasi ini tidak hanya mengubah kata menjadi angka, tetapi juga mencerminkan hubungan semantik antar kata sehingga mampu menghitung kemiripan kata berdasarkan nilai-nilai vektornya. *Embedding* dilakukan menggunakan model yang sama seperti pada tahap *tokenisasi*, yang mana setiap token dikonversi menjadi vektor berdimensi 384 untuk disimpan dan dikelola secara terstruktur, sehingga mendukung pencarian konteks yang relevan berdasarkan perhitungan jarak vektor. Fokus pada tahap ini adalah memastikan konsistensi dimensi vektor untuk mendukung pencarian semantik pada FAISS.

E. Indexing

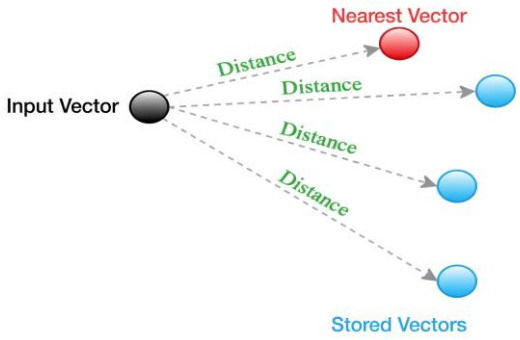
Indexing adalah proses menyimpan dan mengatur vektor hasil pemrosesan teks ke dalam struktur data tertentu agar dapat dicari kembali dengan cepat dan efisien. Dalam sistem ini, seluruh vektor disusun sedemikian rupa sehingga ketika pengguna mengajukan pertanyaan, sistem dapat menemukan vektor yang paling mirip secara semantik dalam waktu singkat. Proses ini memungkinkan pencarian informasi dilakukan berdasarkan kemiripan makna, bukan sekadar kecocokan kata secara literal sebagaimana divisualisasikan pada Gambar 2 [14].



Gambar 2. Visualisasi 3D *Indexing*

F. Data Retrieval

Data retrieval adalah proses pengambilan informasi yang relevan berdasarkan pertanyaan pengguna [15]. Dalam sistem *chatbot* berbasis RAG, setiap pertanyaan diubah menjadi vektor menggunakan model *Sentence Transformer all-MiniLM-L6-v2*. Vektor pertanyaan kemudian dibandingkan dengan vektor dokumen yang tersimpan di FAISS untuk menemukan potongan konteks paling relevan. Proses pencarian ini dilakukan menggunakan *IndexFlatL2* dengan ukuran kemiripan berbasis jarak *Euclidean* (L2), dan hasilnya berupa *top-k* konteks yang akan digunakan dalam pembuatan *prompt* bagi model LLM. Gambar 3 mengilustrasikan pencarian vektor dengan menggunakan *IndexFlatL2.* Pada penelitian ini, nilai *k* ditetapkan sebanyak 2 untuk mengambil dua konteks terdekat dari indeks. Pendekatan ini dipilih untuk memastikan model memperoleh informasi yang kaya dan relevan, sehingga jawaban yang dihasilkan lebih akurat dan informatif. Dengan menggabungkan dua konteks terdekat, *chatbot* dapat menyusun respons yang tidak hanya menjawab pertanyaan secara langsung, tetapi juga memberikan detail tambahan yang mendukung. Tahap ini berperan penting dalam meminimalkan risiko halusinasi jawaban dan meningkatkan relevansi informasi yang disampaikan kepada pengguna.



Gambar 3. Ilustrasi Pencarian Vektor

G. LLM Generation

Model Qwen2.5 VL 72B Instruct digunakan sebagai komponen utama untuk menghasilkan respons. Model ini dirilis oleh *Alibaba Cloud* pada 1 Februari 2025 yang memiliki 72 miliar parameter dan dapat digunakan secara gratis melalui layanan *OpenRouter* [16]. Pemilihan model ini didasarkan pada beberapa pertimbangan strategis. Pertama, model ini memiliki performa *state-of-the-art* dalam pengolahan bahasa alami multibahasa, termasuk pemahaman konteks formal dalam Bahasa Indonesia yang sangat krusial untuk dokumen publik. Kedua, dibandingkan dengan model tertutup seperti GPT-4, Qwen2.5 melalui API OpenRouter menawarkan keseimbangan optimal antara biaya operasional dan *throughput* (30,68 token/detik). Ketiga, arsitektur 72B memberikan kemampuan penalaran (*reasoning*) yang lebih stabil dalam membedakan informasi publik yang serupa, sehingga efektif dalam meminimalisir halusinasi saat diintegrasikan dengan RAG. Integrasi model dilakukan melalui API dengan pengelolaan kredensial secara aman untuk memastikan koneksi antara sistem dan layanan LLM tetap terjaga. Pada tahap generasi jawaban, sistem menyusun *prompt* yang merupakan gabungan antara konteks hasil pencarian dari FAISS dan pertanyaan pengguna. *Prompt* tersebut kemudian dikirim ke model *Qwen 2.5* untuk diproses sehingga menghasilkan jawaban yang relevan dan sesuai dengan informasi yang tersedia.

H. Deployment

Tahap *deployment* dilakukan dengan mengimplementasikan model yang telah dibangun ke dalam platform telegram, sehingga dapat diakses dan digunakan langsung oleh pengguna [17]. Proses ini mencakup pembuatan bot, integrasi dengan sistem RAG, serta penyediaan antarmuka komunikasi yang memudahkan interaksi antara pengguna dan *chatbot* PPID Universitas Lampung. Telegram dipilih sebagai platform karena ringan, populer, dan mendukung integrasi *bot* melalui API. Pembuatan bot dilakukan melalui layanan *BotFather* dengan perintah */newbot*, diikuti penentuan nama dan *username* bot. Dalam penelitian ini bot diberi nama *Chatbot PPID Universitas Lampung* dengan *user Retrieval Augmented Generation name* PPIDUnilaBot**.** Setelah pembuatan bot selesai, *BotFather* menghasilkan *HTTP API token* yang digunakan untuk menghubungkan sistem yang telah dibangun dengan *server* telegram.

III. HASIL DAN PEMBAHASAN

A. Dataset

Dalam penelitian ini dataset terdiri dari 375 pasangan QA yang disusun dalam format JSON. Format ini dipilih karena memudahkan proses embedding dan pencarian, serta memastikan bahwa informasi publik dapat direpresentasikan secara efektif dan relevan kepada pengguna. Tabel 1 menunjukkan struktur dataset pasangan pertanyaan dan jawaban (QA) yang disusun secara manual berdasarkan 47 entri informasi publik resmi dari situs web PPID Universitas Lampung. Dataset ini disimpan dalam format JSON untuk memudahkan integrasi pada proses embedding serta pencarian semantik dalam arsitektur RAG. Cakupan data di dalam tabel meliputi informasi kelembagaan fundamental seperti profil rektor dan status akreditasi "UNGGUL", hingga data teknis mengenai kode perguruan tinggi dan tautan dokumen laporan pengaduan. Setiap pasangan QA dirancang untuk merepresentasikan informasi paling relevan dengan memperhatikan konteks dokumen sumber guna memastikan respons chatbot tetap faktual dan dapat dipertanggungjawabkan. Melalui pengorganisasian data yang terstruktur ini, sistem mampu mentransformasi informasi publik yang sebelumnya tersebar menjadi dialog pengetahuan interaktif yang mendukung prinsip keterbukaan informasi.

TABEL 1
DATASET QA

No	Questions	Answer
1	Unila adalah?	Unila adalah singkatan dari Universitas Lampung perguruan tinggi negeri yang ada di Indonesia
2	Siapa rektor Universitas Lampung?	Prof. Dr. Ir. Lusmeilia Afriani. D. E. A., IPM., ASEAN Eng
3	Apa status akreditasi Universitas Lampung?	Akreditasi Universitas Lampung adalah UNGGUL
4	Apa status aktivitas Universitas Lampung?	Status aktivitas Universitas Lampung adalah AKTIF
5	Berapa kode PT Universitas Lampung?	kode PT Universitas Lampung adalah 001026
…	…	…
375	Laporan pengaduan internal dan eksternal Universitas Lampung?	Laporan pengaduan internal dan eksternal Universitas Lampung dapat diakses di https://ppid.unila.ac.id/wp- content/uploads/2024/09/Pengaduan-Masyarakat-2023- 2024.pdf

B. Monitoring Performa Model

Pemantauan terhadap performa model dilakukan sebagai langkah krusial untuk memastikan bahwa sistem chatbot beroperasi secara optimal, efisien, dan stabil saat menangani permintaan pengguna secara langsung melalui antarmuka Telegram. Parameter evaluasi yang dikumpulkan dalam proses ini mencakup waktu eksekusi (timestamp), volume data berupa jumlah token input dan output, kecepatan pemrosesan (tokens per second/tps), latensi (latency) atau waktu jeda respons, serta estimasi biaya penggunaan layanan.

Hasil monitoring performa model Qwen2.5 VL 72B Instruct dirangkum dalam Tabel 2, yang menyajikan data dari 20 sampel interaksi yang diambil pada 30 April 2025. Berdasarkan data tersebut, sistem menghadapi beban kerja yang bervariasi dengan jumlah token input berkisar antara 25 hingga 397 token dan token output antara 72 hingga 764 token. Meskipun beban kerja beragam, performa teknis sistem menunjukkan efisiensi yang tinggi, di mana rata-rata latensi tercatat hanya sebesar 0,69 detik dengan kecepatan pemrosesan mencapai 30,68 token per detik.

Secara sistematis, analisis ini menunjukkan bahwa integrasi model Qwen2.5 melalui API OpenRouter berhasil memberikan keseimbangan optimal antara kecepatan respons dan pemrosesan throughput yang diperlukan untuk layanan publik secara real-time. Selain itu, seluruh interaksi mencatatkan biaya operasional sebesar nol, yang menegaskan stabilitas dan keberlanjutan ekonomi sistem. Hasil pemantauan ini menyimpulkan bahwa model mampu menangani pertanyaan dengan berbagai tingkat kompleksitas secara cepat, stabil, dan proporsional sesuai dengan kebutuhan pengguna.

TABEL 2
HASIL MONITORING PERFORMA MODEL

No	*Tokens*	*Latency*	*Speed(tps)*	*Cost ($)*
1	83 → 321	0.48	35.4	0
2	65 → 377	0.47	29.4	0
3	111 → 382	0.4	29.5	0
4	130 → 249	1.61	22.2	0
5	252 → 536	1.52	35.5	0
6	397 → 421	0.86	37.6	0
7	89 → 255	0.41	26.6	0
8	137 → 350	1.1	35.9	0
9	103 → 84	0.37	36.3	0
10	141 → 293	0.35	30.1	0
11	386 → 445	1.27	22.7	0
12	129 → 553	0.41	28.9	0
13	121 → 376	0.47	29.4	0
14	133 → 764	0.57	29.7	0
15	50 → 188	0.68	14.1	0
16	27 → 147	0.41	31.3	0
17	25 → 187	0.49	19.6	0
18	46 → 72	0.49	24.1	0
19	115 → 200	0.95	35.6	0
20	83 → 435	0.69	35.7	0

C. Evaluasi Menggunakan Metrik RAGAS

Pengujian kualitas sistem dilakukan secara sistematis dengan mengajukan 30 pertanyaan acak kepada chatbot PPID Unila, yang kemudian dievaluasi menggunakan metrik Retrieval Augmented Generation Assessment Score (RAGAS). Instrumen evaluasi ini mengukur kualitas respon berdasarkan tiga aspek fundamental: faithfulness (akurasi jawaban terhadap konteks), answer relevance (kesesuaian jawaban dengan pertanyaan), dan context recall (cakupan informasi yang berhasil diserap dari konteks) [18]. Dalam prosedur pengujian ini, ditetapkan ambang batas minimal (threshold) sebesar 0,8 untuk setiap parameter metrik [19]. Apabila hasil evaluasi menunjukkan skor di bawah standar tersebut, dilakukan perbaikan secara langsung pada dataset QA melalui proses evaluasi iteratif hingga seluruh metrik melampaui skor minimum yang telah ditentukan. Pendekatan ini bertujuan untuk memastikan bahwa setiap jawaban yang dihasilkan tetap relevan dan selaras dengan informasi resmi yang tersedia pada situs web PPID Unila.

Hasil evaluasi metrik RAGAS yang dilakukan secara iteratif untuk menjamin kualitas respons sistem disajikan secara rinci pada Tabel 3. Data menunjukkan adanya peningkatan performa chatbot yang signifikan setelah melalui 20 kali iterasi perbaikan terhadap dataset QA. Pada akhir proses iterasi, sistem berhasil mencapai skor akhir yang impresif, yaitu faithfulness sebesar 0,9629, answer relevance sebesar 0,8760, dan context recall sebesar 0,8681.

Tingginya pencapaian skor pada aspek faithfulness dan relevansi menunjukkan bahwa integrasi arsitektur RAG bukan sekadar elemen otomasi teknis, melainkan instrumen kunci dalam mewujudkan tata kelola digital yang transparan dan akuntabel. Dengan kemampuan meminimalkan halusinasi yang sering ditemukan pada model LLM murni, sistem ini menjamin validitas informasi publik yang disampaikan agar tetap sesuai dengan dokumen hukum serta prosedur resmi universitas. Implementasi ini secara efektif mentransformasi layanan PPID dari sekadar pencarian dokumen statis menjadi dialog pengetahuan interaktif, yang pada gilirannya mendemokratisasi akses informasi bagi seluruh civitas akademika dan masyarakat luas melalui platform Telegram.

TABEL 3
HASIL EVALUASI METRIK RAGAS

i	*Faithfulness*	*Answer Relevance*	*Context Recall*
1	0.8517	0.7933	0.6887
2	0.8629	0.7740	0.6891
3	0.8634	0.7738	0.6891
4	0.8720	0.7636	0.6911
5	0.9209	0.7450	0.7213
…	…	…	…
20	0.9629	0.8760	0.8681

D. Evaluasi Efektivitas RAG pada LLM

Pengembangan chatbot berbasis RAG bertujuan untuk meningkatkan relevansi dan akurasi jawaban LLM, khususnya dalam menjawab pertanyaan seputar informasi PPID Unila, sekaligus meminimalisir halusinasi yang sering terjadi pada LLM murni. Untuk mengukur kontribusi mekanisme retrieval dalam meningkatkan kualitas respons, dilakukan pengujian perbandingan antara LLM murni Qwen 2.5 dan sistem berbasis RAG yang telah dibuat. Pemilihan model yang sama bertujuan untuk memastikan bahwa proses evaluasi dilakukan secara adil dan sebanding. Sebanyak 20 pertanyaan identik diajukan ke kedua sistem dan hasilnya dibandingkan untuk mengevaluasi efektivitas integrasi RAG terhadap ketepatan dan konteks jawaban.

Hasil pengujian menunjukkan bahwa LLM murni Qwen 2.5 sering menghasilkan jawaban yang tidak akurat atau bersifat halusinasi, terutama pada pertanyaan yang berkaitan dengan kebijakan, prosedur, dan dokumen spesifik PPID Unila. Dari seluruh pertanyaan yang diajukan, hanya empat pertanyaan umum yang dapat dijawab dengan tepat, sedangkan sisanya cenderung keliru atau asumtif. Sebaliknya, chatbot PPID Unila yang dikembangkan dengan pendekatan RAG mampu menjawab semua pertanyaan yang diajukan dengan akurat dan dapat dipertanggungjawabkan. Temuan ini menegaskan bahwa tanpa dukungan mekanisme RAG, model sulit memberikan jawaban faktual, sehingga integrasi RAG terbukti penting untuk meningkatkan relevansi dan ketepatan respons chatbot.

E. Chatbot Usability Questionnaire (CUQ)

Tahap pengujian usabilitas dilakukan untuk mengukur tingkat kepuasan dan pengalaman pengguna terhadap sistem yang telah dikembangkan dengan menerapkan metode Chatbot Usability Questionnaire (CUQ). Instrumen CUQ ini terdiri dari 16 butir pernyataan yang dinilai menggunakan skala Likert 1–5, mencakup dimensi kemudahan penggunaan, kejelasan jawaban, serta relevansi informasi. Untuk menjamin objektivitas hasil dan memberikan gambaran pengalaman pengguna secara komprehensif, kuesioner disusun dengan mengkombinasikan pernyataan bernada positif dan negatif. Penggunaan instrumen yang berimbang ini bertujuan untuk menyeimbangkan aspek usabilitas, keterlibatan, dan efektivitas interaksi [20]. Rincian instrumen pernyataan yang digunakan dalam pengujian ini disajikan pada Tabel 4, yang secara khusus mengevaluasi kemampuan chatbot dalam menyajikan interaksi yang menyerupai manusia, efektivitas penanganan kesalahan pengetikan, serta nilai praktis sistem dalam mempermudah akses informasi publik.

TABEL 4
DAFTAR PERTANYAAN CUQ

No	Pertanyaan	Jenis
1	*Chatbot* memberikan interaksi layaknya manusia	Positif
2	*Chatbot* terasa kaku	Negatif
3	Saat pertama kali menggunakan *Chatbot* ini, saya merasa nyaman	Positif
4	*Chatbot* sering memberikan respons yang kurang baik	Negatif
5	*Chatbot* menjalankan fungsinya dengan baik	Positif
6	Saya tidak memahami tujuan *chatbot* setelah menggunakannya	Negatif
7	*Chatbot* mudah dipahami dan digunakan	Positif
8	Saya bingung saat berinteraksi dengan *chatbot*	Negatif
9	*Chatbot* dapat memahami pertanyaan saya dengan baik	Positif
10	*Chatbot* sering salah memahami atau tidak bisa menjawab pertanyaan saya	Negatif
11	Jawaban *chatbot* informatif	Positif
12	Jawaban *chatbot* tidak relevan	Negatif
13	Ketika saya salah mengetik atau bertanya, *chatbot* tetap bisa memberikan respons yang membantu	Positif
14	Jika saya melakukan kesalahan dalam mengetik, *chatbot* tidak bisa menangani dengan baik	Negatif
15	Saya merasa *chatbot* sangat praktis dan memberi kemudahan	Positif
16	Saya merasa *chatbot* terlalu rumit dan tidak memberi kemudahan	Negatif

Proses pengujian melibatkan 10 responden dengan latar belakang beragam, meliputi mahasiswa, dosen, tenaga kependidikan, serta anggota tim PPID Universitas Lampung. Sebelum melakukan penilaian, setiap responden diwajibkan melakukan uji coba langsung dengan mengajukan berbagai pertanyaan terkait informasi publik melalui antarmuka chatbot guna mendapatkan pengalaman penggunaan yang nyata. Pendekatan ini memastikan bahwa skor akhir yang diperoleh mencerminkan kualitas interaksi teknis dan kepuasan subjektif responden secara akurat dan seimbang.

Data hasil kuesioner selanjutnya diolah untuk memperoleh nilai CUQ yang merepresentasikan tingkat usabilitas sistem secara keseluruhan. Skor dihitung dengan menjumlahkan seluruh nilai dari responden, kemudian dibagi dengan total skor maksimum yang mungkin dicapai. Guna menjaga konsistensi penilaian, skor untuk pernyataan positif dihitung sesuai dengan nilai Likert yang diberikan (1–5), sedangkan untuk pernyataan negatif diterapkan transformasi menggunakan rumus Persamaan (1). Melalui teknik standarisasi ini, nilai yang lebih rendah pada pernyataan negatif tetap mencerminkan tingkat kepuasan yang tinggi, sehingga seluruh skor akhir dapat merepresentasikan kualitas layanan secara seimbang, sistematis, dan akurat.

$$CUQ\ Score = \left(\frac{\sum(\text{Skor Positif}) + \sum(6 - \text{Skor Negatif})}{jumlah\ pertanyaan\ x\ 5} \right) x\ 100 \quad (1)$$

Hasil analisis usabilitas yang melibatkan 10 responden dirangkum dalam Tabel 5, yang menyajikan rincian skor CUQ (*Chatbot Usability Questionnaire*) untuk setiap partisipan. Berdasarkan data tersebut, diperoleh hasil penilaian yang berada dalam rentang 83,75 hingga 97,50, dengan pencapaian rata-rata skor akhir sebesar 90,62. Secara spesifik, skor tertinggi diberikan oleh responden 3 sebesar 97,50, sementara skor terendah diperoleh dari responden 2 sebesar 83,75.

Secara keseluruhan, capaian rata-rata skor yang melampaui angka 90 tersebut mengindikasikan tingkat kepuasan pengguna yang sangat baik terhadap performa chatbot yang dikembangkan. Tingginya akumulasi skor ini merefleksikan keunggulan sistem dalam menyajikan respons yang relevan, akurat, dan efisien, sehingga mampu mentransformasi prosedur pencarian informasi publik menjadi lebih praktis dan mudah dipahami oleh pengguna. Hasil ini mengonfirmasi bahwa integrasi teknologi RAG dan LLM Qwen2.5 tidak hanya unggul secara teknis, tetapi juga berhasil memberikan pengalaman pengguna yang memuaskan dalam implementasi nyata.

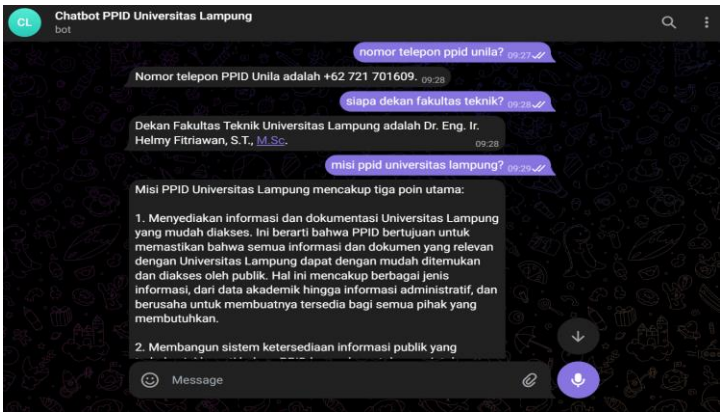
TABEL 5
HASIL PERHITUNGAN CUQ SCORE

Responden	**CUQ *Score***
Responden 1	87.50
Responden 2	83.75
Responden 3	97.50
Responden 4	91.25
Responden 5	85.00
Responden 6	85.00
Responden 7	87.50
Responden 8	96.25
Responden 9	96.25
Responden 10	96.25
Rata-rata	90.62

F. Visualisasi Chatbot PPID Universitas Lampung

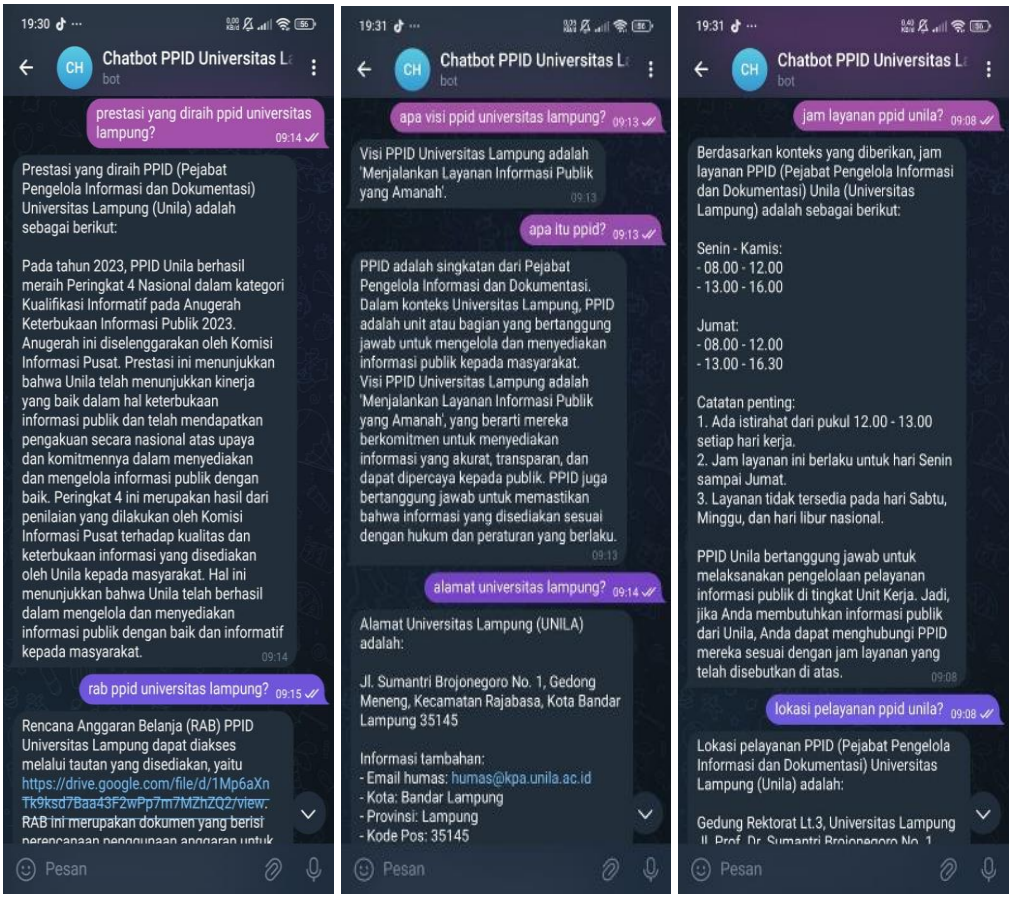
Setelah integrasi sistem *chatbot* dengan telegram berhasil dilakukan, pengguna dapat langsung berinteraksi dengan *chatbot* PPID Universitas Lampung melalui telegram. Proses tanya jawab berlangsung secara *real-time*, dimana pengguna mengirimkan pertanyaan dan *chatbot* merespons berdasarkan hasil pencarian kontekstual dari data yang tersedia. Respons yang diberikan bukanlah jawaban statis, melainkan hasil dari pemrosesan cerdas yang melibatkan pencocokan *embedding* dan inferensi menggunakan LLM. Gambar 4 menyajikan visualisasi antarmuka *chatbot* PPID Universitas Lampung saat diakses melalui platform Telegram versi *website* atau desktop. Melalui tampilan ini, terlihat bahwa interaksi antara pengguna dan sistem berlangsung secara *real-time*, dimana pengguna dapat mengajukan berbagai pertanyaan spesifik seperti nomor telepon kontak, nama pimpinan fakultas, hingga poin-poin misi organisasi. Respons yang dihasilkan oleh *chatbot* merupakan hasil dari pemrosesan cerdas yang melibatkan pencocokan *embedding* dan inferensi menggunakan model LLM, sehingga jawaban yang diberikan bersifat dinamis, akurat, dan relevan dengan basis pengetahuan yang tersedia.

p-ISSN : 2443-2210
e-ISSN : 2443-2229



Gambar 4. Visualisasi Chatbot di Telegram Website

Gambar 5 mendemonstrasikan fungsionalitas *chatbot* yang diakses melalui perangkat *smartphone* menggunakan aplikasi Telegram *mobile*. Visualisasi ini menunjukkan kemampuan sistem dalam menangani beragam kategori informasi secara mendalam, mulai dari capaian prestasi nasional tahun 2023, akses dokumen rencana anggaran (RAB), visi organisasi, hingga detail operasional seperti jam dan lokasi layanan. Penggunaan antarmuka yang familiar dan intuitif pada perangkat seluler memudahkan pengguna dalam memahami alur percakapan, sekaligus membuktikan bahwa sistem ini merupakan solusi layanan keterbukaan informasi publik yang praktis, modern, dan adaptif untuk digunakan di mana saja.



Gambar 5. Visualisasi *Chatbot* di Telegram Mobile

JuTISI
Jurnal Teknik Informatika dan Sistem Informasi

Hal ini menunjukkan bahwa *chatbot* dapat dijalankan lintas platform, baik melalui perangkat komputer maupun *smartphone*. Selain itu, tampilan antarmuka telegram yang familiar dan intuitif memberikan kemudahan bagi pengguna dalam mengakses dan memahami alur percakapan chatbot. Dengan demikian, pengguna dapat berinteraksi menggunakan *chatbot* secara dinamis dan adaptif sesuai dengan perangkat yang digunakan, tanpa mengurangi fungsionalitas maupun kualitas respons yang diberikan. Kemampuan ini memperkuat posisi *chatbot* sebagai solusi layanan keterbukaan informasi yang praktis, modern, dan mudah diakses. Meskipun menunjukkan performa unggul, sistem ini memiliki beberapa keterbatasan yang menjadi landasan penelitian selanjutnya. Pertama, akurasi jawaban sangat bergantung pada kualitas dan kelengkapan dataset QA yang disusun secara manual. Jika terdapat perubahan kebijakan atau dokumen baru pada situs PPID yang belum diperbarui dalam vector database, sistem berpotensi memberikan informasi yang kedaluwarsa. Kedua, domain informasi saat ini terbatas pada delapan kategori utama PPID Universitas Lampung. Pengembangan di masa depan perlu mempertimbangkan mekanisme automated scraping dan pembaruan indeks secara real-time untuk menjaga relevansi informasi tanpa intervensi manual yang intensif.

IV. SIMPULAN

Berdasarkan hasil penelitian, chatbot PPID Universitas Lampung berhasil dikembangkan dan diintegrasikan ke Telegram menggunakan pendekatan Retrieval Augmented Generation (RAG) dengan Large Language Model LLM Qwen 2.5 VL 72B Instruct. Evaluasi menunjukkan performa yang sangat baik, dengan skor RAGAS di atas 0,8 pada seluruh metrik (faithfulness sebesar 0.9629, answer relevance sebesar 0.8760, dan context recall sebesar 0.8681) serta skor CUQ dengan rata-rata 90,62 yang mencerminkan pengalaman pengguna yang positif. Pendekatan RAG terbukti efektif dalam mengurangi halusinasi jawaban, di mana chatbot mampu menjawab seluruh pertanyaan uji dengan tepat, sementara model LLM tanpa mekanisme retrieval hanya mampu menjawab benar sebagian kecil dari pertanyaan yang diajukan. Penelitian ini membuktikan bahwa integrasi Large Language Model (LLM) Qwen2.5 dengan arsitektur RAG mampu mentransformasi layanan informasi publik yang statis menjadi sistem dialog pengetahuan yang dinamis dan akurat. Pendekatan RAG terbukti menjadi solusi fundamental dalam mengatasi keterbatasan LLM generatif terhadap data lokal, khususnya dalam meminimalisir risiko halusinasi informasi pada dokumen kebijakan publik yang bersifat krusial. Keberhasilan sistem ini, yang divalidasi melalui metrik RAGAS dan tingkat kepuasan pengguna yang tinggi, menegaskan bahwa kecerdasan buatan dapat menjadi pilar utama dalam mendukung transparansi dan tata kelola pemerintahan yang baik di lingkungan akademis. Sebagai langkah pengembangan lebih lanjut, penelitian ini dapat diperluas melalui tiga arah strategis. Pertama, otomatisasi pembaruan data menggunakan mekanisme web scraping terjadwal untuk menjamin relevansi informasi tanpa ketergantungan pada penyusunan dataset QA secara manual. Kedua, perluasan domain informasi yang mencakup integrasi layanan data akademik dan administratif lainnya untuk menciptakan asisten virtual yang lebih komprehensif. Ketiga, model ini memiliki potensi replikasi pada institusi publik lain guna mempercepat transformasi digital dalam layanan keterbukaan informasi publik di tingkat nasional.

UCAPAN TERIMA KASIH

Penulis menyampaikan apresiasi kepada Universitas Lampung, khususnya Pejabat Pengelola Informasi dan Dokumentasi (PPID), atas dukungan administratif serta penyediaan akses data yang menjadi basis utama dalam penelitian ini. Ucapan terima kasih juga ditujukan kepada rekan sejawat atas diskusi produktif dan masukan konstruktif yang diberikan selama tahap pengembangan sistem hingga penyelesaian artikel ilmiah ini. Penelitian ini diharapkan dapat memberikan kontribusi nyata bagi pengembangan teknologi keterbukaan informasi publik di lingkungan akademis..

DAFTAR PUSTAKA

[1] Kementerian Komunikasi dan Informatika, 'Undang-Undang Nomor 14 Tahun 2008 tentang Keterbukaan Informasi Publik', 2010. [Online]. Available: https://jdih.komdigi.go.id/produk_hukum/unduh/id/172/t/undangundang+nomor+14+tahun++2008+tanggal+30+april+2008

[2] PPID Unila, 'Profil Pejabat Pengelola Informasi dan Dokumentasi Universitas Universitas Lampung', 2026. [Online]. Available: https://ppid.unila.ac.id/profil/

[3] I. Zaenuddin and A. B. Riyan, 'Perkembangan Kecerdasan Buatan (AI) Dan Dampaknya Pada Dunia Teknologi', *J. Inform. Utama*, vol. 2, no. 2, pp. 128–153, 2024.

[4] E. Yulianto, T. Murdianto, and Al-Amin, 'Peran Artificial Intelligence (AI) dalam Manajemen Arsip dan Dokumen', *J. Ilmu Inf. dan Perpust.*, vol. 1, no. 6, pp. 484–499, 2024.

[5] C. R. Hidayat, Y. Sumaryana, D. S. Anwar, A. L. N. Fadilah, and M. R. Saputra, 'Pengembangan Virtual Assistant (Chatbot) Bebrbasis NLP (Natural Language Processing) Untuk Portal Informasi Terpadu Pariwisata Tasikmalaya', *CSRID (Computer Sci. Res. Its Dev. Journal)*, vol. 17, no. 1, pp. 136–148, 2025.

[6] R. Xu, Y. Feng, and H. Chen, 'ChatGPT vs. Google: A Comparative Study of Search Performance and User Experience', *SSRN Electron. J.*, 2023.

[7] A. F. Arz von Straussenburg and A. Wolters, 'Towards Hybrid Architectures: Integrating Large Language Models in Informative Chatbots', in Wirtschaftsinformatik 2023. 2023. [Online]. Available: https://aisel.aisnet.org/wi2023/9

[8] L. R. Hidayat, I. G. P. S. Wijaya, and R. Dwiyansaputra, 'Optimalisasi Layanan Sistem Informasi Mahasiswa Dengan Integrasi Telegram:

Chatbot Retrieval-Augmented-Generation Berbasis Large Language Model', *J. Teknol. Informasi, Komputer, dan Apl.*, vol. 7, no. 1, pp. 121–131, 2025.

[9] N. Krawczyk, B. Probierz, and J. Kozak, 'Towards AI-Generated Essay Classification Using Numerical Text Representation', *Appl. Sci.*, vol. 14, no. 21, 2024.

[10] A. A. Nur Hakim, A. C. Murti, and R. Nindyasari, 'Implementasi Artificial Intelligence Dalam Sistem Pencarian Orang Hilang Dengan Face Recognition Studi Kasus Polres Kudus', *SKANIKA Sist. Komput. dan Tek. Inform.*, vol. 8, no. 1, pp. 168–180, 2025.

[11] C. Si and et Al., 'Sub-Character Tokenization for Chinese Pretrained Language Models', *Trans. Assoc. Comput. Linguist.*, vol. 11, pp. 469–487, 2023.

[12] F. A. Nugraha, N. H. Harani, and R. Habibi, *Analisis Sentimen Terhadap Pembatasan Sosial Menggunakan Deep Learning*. Bandung: Kreatif, 2020.

[13] S. Ghannay, B. Favre, Y. Estève, and N. Camelin, 'Word embeddings evaluation and combination', in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 2016, pp. 300–305.

[14] J. R. Smith, M. Naphade, and A. Natsev, 'Multimedia semantic indexing using model vectors', in *2003 International Conference on Multimedia and Expo. ICME '03. Proceedings (Cat. No.03TH8698)*, 2003, pp. 445–448.

[15] H. Tohir, N. Merlina, and M. Haris, 'Utilizing Retrieval-Augmented Generation in real-times To Enhance Indonesian Language Nlp', *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 10, no. 2, pp. 352–360, 2024.

[16] OpenRouter, 'PromptLayer Quickstart'. [Online]. Available: https://openrouter.ai/docs/quickstart

[17] I. P. G. A. Sudiatmika, 'E-Learning Berbasis Telegram Bot', *KERNEL J. Ris. Inov. Bid. Inform. dan Pendidik. Inform.*, vol. 1, no. 2, pp. 49–60, 2021.

[18] S. Es, J. James, L. Espinosa-Anke, and S. Schockaert, 'RAGAS: Automated Evaluation of Retrieval Augmented Generation', in *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, 2024, pp. 150–158.

[19] Shahul, 'Evaluating RAG pipelines with Ragas and Openlayer'. [Online]. Available: https://www.openlayer.com/blog/post/evaluating-rag-pipelines-with-ragas-and-openlayer

[20] A. Alabbas and K. Alomar, 'A Weighted Composite Metric for Evaluating User Experience in Educational Chatbots: Balancing Usability, Engagement, and Effectiveness', *Futur. Internet*, vol. 17, no. 2, pp. 1–35, 2025.