JSAI : Journal Scientific and Applied Informatics
Vol. 09, No. 3, *October* 2026, *pages*. 400~408
E-ISSN: 2614-3054; P-ISSN: 2614-3062, *accredited by Kemenristekdikti*, Sinta 4
DOI: 10.36085

Integrasi Informasi Akademik Publik dan Data Personal Mahasiswa pada *Chatbot* Berbasis *Large Language Model* dengan Pendekatan *Retrieval-Augmented Generation*

[1]Erlangga Wijaya, [2]Ikrimach
[1,2]Universitas Teknologi Yogyakarta, Indonesia
[1]*erlanggawijayaofficial@gmail.com*; [2]*ikrimach@uty.ac.id*;

Article Info

Article history:

Received, 2026-10-12
Revised, 2026-10-24
Accepted, 2026-11-02

Kata Kunci:

Chatbot akademik
Large Language Model
Retrieval-Augmented Generation
Pencarian semantic
Sistem informasi akademik

ABSTRAK

Tujuan: Penelitian ini bertujuan mengembangkan chatbot akademik berbasis *Large Language Model* (LLM) dengan pendekatan *Retrieval-Augmented Generation* (RAG) untuk mengintegrasikan informasi akademik publik dan data akademik personal mahasiswa Universitas Teknologi Yogyakarta (UTY). Metode: Sistem dikembangkan dengan memanfaatkan dokumen akademik publik yang diproses melalui ekstraksi teks, chunking, dan embedding, kemudian disimpan dalam *ChromaDB*. Data akademik personal diperoleh melalui autentikasi Sistem Informasi Akademik (SIA) dan disimpan dalam PostgreSQL. Evaluasi dilakukan menggunakan *Black Box Testing* terhadap sembilan skenario dan pengujian RAG terhadap 50 pertanyaan dengan lima metrik evaluasi. Hasil: Pengujian fungsional menunjukkan tingkat keberhasilan sebesar 100% (9 dari 9 skenario). Evaluasi RAG menghasilkan *Retrieval Hit Rate* dan *Context Relevance* sebesar 96,0%, *Answer Relevance* 93,2%, *Faithfulness* 94,0%, serta skor *Accuracy ekuivalen* 92,8% dari skor maksimum. Kesimpulan: Integrasi RAG dan data akademik personal dapat diterapkan untuk menyediakan layanan informasi akademik melalui satu antarmuka percakapan dengan mekanisme akses terautentikasi.

Keywords:

Academic chatbot
Large Language Model
Retrieval-Augmented Generation
semantic search
academic information system

ABSTRACT

Objective: This study aims to develop an academic chatbot based on a Large Language Model (LLM) using a Retrieval-Augmented Generation (RAG) approach to integrate public academic information and students' personal academic data at Universitas Teknologi Yogyakarta (UTY). Methods: The system utilizes public academic documents processed through text extraction, chunking, and embedding, then stored in ChromaDB. Personal academic data are retrieved through authentication via the Academic Information System (SIA) and stored in PostgreSQL. Evaluation consists of Black Box Testing involving nine scenarios and RAG assessment using 50 questions across five evaluation metrics. Results: Functional testing achieved a success rate of 100% (9 out of 9 scenarios). RAG evaluation yielded 96.0% for Retrieval Hit Rate and Context Relevance, 93.2% for Answer Relevance, 94.0% for Faithfulness, and an equivalent Accuracy score of 92.8% of the maximum scale. Conclusion: Integrating RAG with personal academic data can provide academic information services through a single conversational interface with authenticated access mechanisms.

This is an open access article under the CC BY-SAlicense.

Penulis Korespondensi:

Erlangga Wijaya,
Universitas Teknologi Yogyakarta,
Email: *erlanggawijayaofficial@gmail.com*

1. PENDAHULUAN

Perkembangan *Large Language Model* (LLM) telah mendorong penggunaan *chatbot* sebagai antarmuka interaksi berbasis bahasa alami dalam berbagai bidang, termasuk pendidikan tinggi. Kemampuan LLM dalam memahami masukan pengguna dan menghasilkan respons secara kontekstual memungkinkan *chatbot* digunakan untuk mendukung interaksi dan layanan akademik. Kajian mengenai *Large Language Model*-based

chatbot di pendidikan tinggi menunjukkan bahwa teknologi tersebut telah digunakan pada berbagai kebutuhan akademik, sekaligus menyoroti kebutuhan terhadap keandalan informasi dan kesesuaian penggunaan teknologi dalam lingkungan pendidikan[1],[2]. Pengembangan *chatbot* untuk kebutuhan akademik juga dilakukan dengan memanfaatkan pemrosesan bahasa alami dan klasifikasi maksud pengguna untuk menyediakan informasi sesuai dengan kebutuhan pengguna [3],[4].

Penerapan *chatbot* berbasis LLM tidak hanya bergantung pada kemampuan model dalam menghasilkan teks, tetapi juga pada informasi yang diberikan sebagai konteks. Penambahan sumber data eksternal dapat digunakan untuk menyesuaikan respons model dengan informasi pada domain tertentu[5]. Penggunaan metode *in-context* juga menunjukkan bahwa konteks yang diberikan kepada model merupakan salah satu aspek yang dapat memengaruhi respons LLM pada berbagai bahasa[6]. Penerapan LLM yang dikombinasikan dengan komponen pemrosesan bahasa alami juga telah digunakan pada pengembangan *chatbot* untuk kebutuhan layanan tertentu[7],[8].

Dalam lingkungan perguruan tinggi, informasi akademik memiliki karakteristik yang beragam. Informasi publik dapat berupa pedoman akademik, ketentuan akademik, informasi administrasi, dan informasi layanan mahasiswa, sedangkan informasi personal dapat berupa nilai, presensi, pembayaran, dan jadwal mahasiswa. Perbedaan karakteristik tersebut menyebabkan mekanisme akses terhadap informasi juga berbeda. Informasi publik dapat digunakan sebagai sumber pengetahuan yang dapat diakses secara umum, sedangkan informasi personal memerlukan autentikasi dan pembatasan akses. Pengembangan *chatbot* untuk layanan akademik berbasis RAG telah dilakukan untuk menyediakan informasi akademik melalui sumber pengetahuan eksternal [9].Selain itu, pencarian informasi pada sistem informasi akademik juga dapat memanfaatkan pendekatan pencarian semantik untuk memperoleh informasi yang sesuai dengan kebutuhan pengguna[10].

Pada Universitas Teknologi Yogyakarta (UTY), informasi akademik tersedia melalui *website*, dokumen institusi, dan Sistem Informasi Akademik (SIA). Pengguna perlu mengakses sumber yang berbeda sesuai dengan jenis informasi yang dibutuhkan. Kondisi tersebut menyebabkan proses pencarian informasi belum terintegrasi dalam satu layanan percakapan. Selain itu, penggunaan LLM tanpa sumber pengetahuan eksternal menyebabkan jawaban tidak secara langsung mengacu pada informasi akademik yang digunakan oleh institusi. Kondisi tersebut menunjukkan perlunya pendekatan yang dapat menghubungkan kemampuan generatif LLM dengan sumber informasi akademik yang tersedia pada institusi.

Salah satu pendekatan yang dapat digunakan adalah *Retrieval-Augmented Generation* (RAG). RAG menggabungkan proses *retrieval* informasi dari sumber eksternal dengan kemampuan generatif LLM sehingga konteks yang diperoleh dapat digunakan dalam pembentukan jawaban. Kajian sistematis mengenai RAG pada aplikasi pendidikan menunjukkan bahwa pendekatan tersebut telah digunakan pada berbagai kebutuhan pendidikan, termasuk sistem tanya jawab dan *chatbot*[11]. Kajian mengenai *RAG chatbot* untuk pendidikan juga menunjukkan pemanfaatan sumber pengetahuan eksternal untuk mendukung penyediaan informasi melalui antarmuka percakapan[12].

Dalam implementasinya, RAG dapat menggunakan *vector database* untuk menyimpan representasi informasi dalam bentuk vektor. Pertanyaan pengguna kemudian direpresentasikan dalam bentuk *embedding* untuk memperoleh informasi yang memiliki kemiripan semantik. Implementasi RAG dengan *vector database* telah diterapkan pada pengembangan *chatbot* untuk mengambil informasi dari sumber eksternal sebelum proses generasi jawaban[13]. Pendekatan serupa juga digunakan pada pengembangan *chatbot* berbasis RAG dengan model bahasa untuk menjawab pertanyaan berdasarkan sumber informasi tertentu[14] . Pada proses *retrieval*, pendekatan *dense text retrieval* memungkinkan pencarian dilakukan berdasarkan representasi semantik dari teks, sehingga informasi yang relevan dapat diperoleh berdasarkan kedekatan representasi antara pertanyaan dan dokumen[15]. Penggunaan RAG juga dapat dikombinasikan dengan strategi *prompt engineering* dan penyesuaian model untuk kebutuhan domain tertentu. Penelitian mengenai RAG untuk sistem pengetahuan spesifik menunjukkan bahwa *fine-tuning* dan *prompt engineering* dapat digunakan untuk menyesuaikan proses pembentukan jawaban berdasarkan kebutuhan sistem[16]. Pada konteks bahasa Indonesia, RAG juga telah diterapkan untuk mendukung pemanfaatan LLM dalam pemrosesan bahasa alami sehingga informasi eksternal dapat digunakan dalam pembentukan respons berbahasa Indonesia[17].

Meskipun penelitian sebelumnya telah membahas *chatbot* berbasis LLM pada pendidikan tinggi, penerapan RAG pada bidang pendidikan, layanan akademik berbasis RAG, serta penggunaan *vector database* sebagai sumber informasi eksternal, masih terdapat kebutuhan untuk mengintegrasikan informasi akademik publik dan data akademik personal mahasiswa dalam satu layanan dengan mekanisme akses yang berbeda. Penelitian mengenai *chatbot* akademik berbasis RAG berfokus pada pemanfaatan sumber pengetahuan untuk menyediakan layanan akademik[9], sedangkan penelitian lain menunjukkan penggunaan RAG dan *vector database* untuk memperoleh informasi dari sumber eksternal[13]. Penelitian mengenai sistem informasi akademik juga telah mengembangkan pencarian informasi berdasarkan kebutuhan pengguna[10]. Berdasarkan

penelitian tersebut, penelitian ini memfokuskan integrasi dokumen akademik publik sebagai basis pengetahuan RAG dengan data akademik personal mahasiswa yang diperoleh melalui SIA dan mekanisme autentikasi.

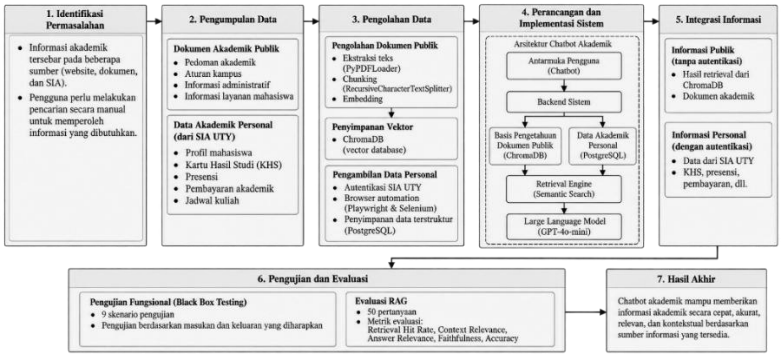
Kontribusi penelitian ini terletak pada penerapan integrasi dua sumber informasi dengan karakteristik akses yang berbeda dalam satu *chatbot* akademik. Informasi publik diproses melalui RAG untuk memperoleh konteks dari dokumen akademik, sedangkan informasi personal diperoleh dari data akademik mahasiswa yang telah terautentikasi. Dengan pendekatan tersebut, informasi umum dan informasi personal dapat diakses melalui satu antarmuka percakapan, sementara mekanisme akses terhadap data personal tetap dibedakan dari akses terhadap basis pengetahuan publik. Penelitian ini tidak diarahkan untuk mengembangkan algoritma RAG baru, tetapi untuk menerapkan dan mengevaluasi integrasi tersebut pada layanan informasi akademik.

Evaluasi diperlukan untuk mengetahui kemampuan sistem dalam memperoleh konteks yang relevan dan menghasilkan jawaban berdasarkan informasi yang tersedia. Pengujian pada sistem berbasis RAG perlu mempertimbangkan kualitas proses *retrieval* dan kualitas jawaban yang dihasilkan. Selain itu, pengujian fungsional diperlukan untuk memastikan fungsi sistem berjalan sesuai dengan kebutuhan yang telah ditentukan. Metode *Black Box Testing* dapat digunakan untuk menguji fungsi sistem berdasarkan masukan dan keluaran yang diamati tanpa melihat struktur internal program[18].

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk mengembangkan dan mengevaluasi *chatbot* berbasis LLM dengan pendekatan RAG yang mampu menyediakan informasi akademik publik dan data akademik personal mahasiswa dalam satu layanan percakapan. Evaluasi dilakukan terhadap 50 pertanyaan menggunakan *Retrieval Hit Rate*, *Context Relevance*, *Answer Relevance*, *Faithfulness*, dan *Accuracy*, serta pengujian fungsional menggunakan *Black Box Testing* terhadap sembilan skenario. Penelitian ini diharapkan memberikan kontribusi praktis dalam pengembangan layanan informasi akademik yang terintegrasi, berbasis sumber informasi, dan dapat diakses melalui antarmuka percakapan.

2. METODE PENELITIAN

Penelitian ini dilakukan untuk mengembangkan *chatbot* akademik berbasis *Large Language Model* (LLM) dengan pendekatan *Retrieval-Augmented Generation* (RAG) sebagai layanan informasi akademik di Universitas Teknologi Yogyakarta (UTY). Tahapan penelitian meliputi identifikasi permasalahan, pengumpulan dan pengolahan data, perancangan arsitektur sistem, implementasi RAG dan LLM, integrasi data akademik personal, serta pengujian dan evaluasi sistem. Tahapan tersebut diarahkan untuk mengembangkan *chatbot* yang memanfaatkan dokumen akademik sebagai sumber informasi publik dan data akademik mahasiswa sebagai sumber informasi personal.



Gambar 1. Kerangka penelitian

Tahap pertama adalah identifikasi permasalahan. Layanan informasi akademik UTY telah tersedia melalui *website* dan Sistem Informasi Akademik (SIA), tetapi informasi yang dibutuhkan pengguna berasal dari beberapa sumber. Kondisi tersebut menyebabkan pengguna perlu melakukan pencarian secara manual untuk memperoleh informasi yang dibutuhkan. Berdasarkan permasalahan tersebut, penelitian diarahkan pada pengembangan *chatbot* yang dapat mengintegrasikan sumber informasi akademik ke dalam layanan percakapan menggunakan bahasa alami.

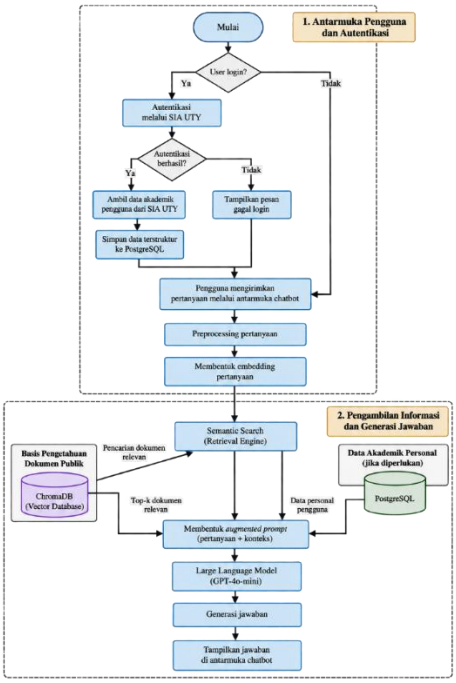
Tahap berikutnya adalah pengumpulan data. Data penelitian terdiri atas dokumen akademik dalam format PDF dan data akademik terstruktur. Dokumen PDF digunakan sebagai sumber pengetahuan untuk informasi akademik publik, seperti pedoman akademik, aturan kampus, informasi administratif, dan informasi layanan mahasiswa. Sementara itu, data akademik personal diperoleh dari SIA UTY dan mencakup data Kartu Hasil Studi (KHS), presensi, pembayaran akademik, serta informasi akademik lainnya. Data tersebut digunakan agar *chatbot* dapat memberikan informasi yang sesuai dengan kondisi pengguna yang telah melakukan autentikasi.

Pengambilan data akademik personal dilakukan melalui proses *automated data extraction* dengan *browser automation* menggunakan Playwright dan Selenium. Proses dimulai dengan autentikasi pengguna melalui SIA UTY. Setelah autentikasi berhasil, sistem mengambil data akademik yang diperlukan dan menyimpannya dalam basis data PostgreSQL dalam bentuk terstruktur. Mekanisme tersebut digunakan agar data personal yang diberikan kepada *chatbot* dapat dikaitkan dengan identitas pengguna yang sedang melakukan akses.

Dokumen akademik publik yang diperoleh kemudian diproses sebagai sumber pengetahuan pada sistem RAG. Tahapan pengolahan dokumen meliputi *text extraction*, *chunking*, dan pembentukan *embedding*. Dokumen yang telah dibagi menjadi beberapa bagian kemudian dikonversi menjadi representasi vektor dan disimpan dalam ChromaDB. Penggunaan representasi vektor memungkinkan sistem melakukan pencarian berdasarkan kemiripan semantik antara pertanyaan pengguna dan dokumen yang tersedia. Pendekatan *dense retrieval* menggunakan representasi vektor memungkinkan sistem memperoleh dokumen yang memiliki kedekatan semantik dengan pertanyaan pengguna.

Penggunaan RAG dan *vector database* juga telah diterapkan pada pengembangan *chatbot* untuk memperoleh informasi dari sumber eksternal sebelum menghasilkan jawaban. Penerapan RAG pada sistem berbasis LLM juga dapat dikombinasikan dengan *prompt engineering* untuk mengarahkan model dalam menghasilkan respons berdasarkan informasi yang diperoleh dari sumber eksternal. Pada implementasinya, dokumen PDF dibaca menggunakan PyPDFLoader. Dokumen yang telah diekstraksi kemudian diproses menggunakan RecursiveCharacterTextSplitter dengan ukuran *chunk* sebesar 600 karakter dan *chunk overlap* sebesar 150 karakter. Hasil pemotongan dokumen selanjutnya digunakan untuk membentuk *embedding* dan disimpan dalam ChromaDB. Tahapan tersebut dilakukan untuk mempersiapkan dokumen agar dapat digunakan dalam proses *retrieval.*

Proses *retrieval* dilakukan ketika pengguna mengirimkan pertanyaan. Pertanyaan pengguna diubah menjadi representasi *embedding*, kemudian digunakan untuk mencari dokumen yang memiliki tingkat kemiripan semantik tertinggi pada ChromaDB. Hasil pencarian berupa dokumen yang relevan digunakan sebagai konteks tambahan dalam proses pembentukan jawaban. Pendekatan tersebut sesuai dengan prinsip RAG yang menggabungkan proses *retrieval* dari sumber eksternal dengan kemampuan generatif LLM. Penerapan RAG pada bidang pendidikan juga menunjukkan pemanfaatan sumber eksternal sebagai konteks untuk mendukung jawaban *chatbot* pada lingkungan Pendidikan.



Gambar 2. Arsitektur sistem *chatbot* akademik

Arsitektur sistem terdiri atas antarmuka pengguna, *backend*, sumber data publik, SIA UTY, PostgreSQL, ChromaDB, *retrieval engine*, dan LLM. Pengguna berinteraksi dengan sistem melalui antarmuka *chatbot*. Pertanyaan yang diberikan kemudian diproses oleh *backend* untuk menentukan sumber informasi yang diperlukan. Informasi publik diperoleh melalui proses *retrieval* pada basis pengetahuan, sedangkan informasi

personal diperoleh dari data akademik pengguna yang telah terautentikasi. Hasil informasi yang diperoleh kemudian digunakan sebagai konteks untuk proses generasi jawaban.

Untuk data akademik personal, sistem menerapkan mekanisme autentikasi melalui SIA UTY sebelum informasi tersebut dapat digunakan. Setelah autentikasi berhasil, sistem melakukan sinkronisasi data akademik dan menyimpannya pada PostgreSQL. Data yang digunakan dalam layanan personal meliputi profil mahasiswa, KHS, presensi, dan pembayaran akademik. Pemisahan antara informasi publik dan informasi personal memungkinkan *chatbot* menyediakan layanan informasi umum tanpa autentikasi sekaligus menyediakan layanan informasi akademik personal bagi pengguna yang telah melakukan *login*.

Hasil *retrieval* dan data akademik personal yang diperlukan kemudian digabungkan dengan pertanyaan pengguna untuk membentuk *augmented prompt*. Konteks tersebut dikirimkan kepada LLM untuk menghasilkan jawaban dalam bahasa alami. Penggunaan sumber data tambahan pada sistem berbasis LLM memungkinkan model menghasilkan respons dengan mempertimbangkan informasi yang diberikan dari luar pengetahuan internal model. Pada penelitian ini digunakan GPT-4o-mini melalui API sebagai model LLM. Model menerima pertanyaan beserta konteks yang diperoleh dari proses *retrieval* dan data akademik sehingga jawaban yang dihasilkan tidak hanya berasal dari pengetahuan internal model, tetapi juga berdasarkan informasi yang tersedia pada sistem. Pendekatan penggunaan sumber eksternal untuk memberikan konteks pada LLM merupakan karakteristik utama RAG dan telah diterapkan pada berbagai *chatbot* berbasis pendidikan. Proses pemberian jawaban *chatbot* diringkas dalam Algoritma.

Prosedur pemberian jawaban *chatbot.*

Input: pertanyaan pengguna
Output: Jawaban *chatbot*

1. Mulai.
2. Pengguna mengirimkan pertanyaan melalui antarmuka *chatbot*.
3. Sistem memeriksa status autentikasi pengguna.
4. Jika pengguna telah terautentikasi, sistem mengambil data akademik personal dari PostgreSQL apabila diperlukan.
5. Sistem melakukan *preprocessing* terhadap pertanyaan.
6. Sistem membentuk *embedding* dari pertanyaan.
7. Sistem melakukan *semantic search* pada ChromaDB.
8. Sistem mengambil dokumen yang relevan berdasarkan hasil *retrieval*.
9. Sistem menentukan konteks informasi yang diperlukan.
10. Jika pertanyaan membutuhkan data personal, sistem menggabungkan hasil *retrieval* dengan data akademik personal. Jika tidak, sistem menggunakan konteks hasil *retrieval*.
11. Sistem membentuk *augmented prompt*.
12. Sistem mengirimkan *augmented prompt* kepada LLM.
13. LLM menghasilkan jawaban berdasarkan konteks yang diberikan.
14. Sistem menampilkan jawaban kepada pengguna.
15. Selesai.

Setelah implementasi selesai, dilakukan pengujian untuk mengetahui kesesuaian fungsi sistem dan kemampuan sistem dalam menghasilkan jawaban berdasarkan sumber informasi yang tersedia. Pengujian fungsional dilakukan menggunakan *Black Box Testing*. Metode tersebut menguji sistem berdasarkan masukan dan keluaran tanpa melihat struktur internal program. Pada penelitian ini, *Black Box Testing* dilakukan terhadap sembilan skenario, yaitu akses halaman utama, *login* SIA UTY, *Knowledge Base Chat*, *Private Chat*, sinkronisasi data akademik, informasi nilai KHS, informasi presensi, informasi pembayaran, dan informasi jadwal kuliah. Setiap skenario diberikan masukan tertentu kemudian diamati keluaran sistem untuk menentukan apakah fungsi berjalan sesuai dengan kebutuhan yang telah ditetapkan.

$$Success\ Rate\ = \frac{N_s}{N_t} \times\ 100\% \qquad (1)$$

dengan Ns merupakan jumlah skenario yang berhasil dan Nt merupakan jumlah seluruh skenario pengujian. Selain pengujian fungsional, dilakukan evaluasi terhadap kemampuan RAG dalam memperoleh konteks dan menghasilkan jawaban. Evaluasi dilakukan terhadap 50 pertanyaan pengujian dengan menggunakan lima metrik, yaitu *Retrieval Hit Rate*, *Context Relevance*, *Answer Relevance*, *Faithfulness*, dan *Accuracy*. Penggunaan beberapa metrik tersebut dilakukan untuk mengevaluasi sistem RAG dari aspek keberhasilan *retrieval*, relevansi konteks, kesesuaian jawaban, keterikatan jawaban terhadap konteks, dan ketepatan jawaban. Evaluasi sistem RAG perlu mempertimbangkan proses *retrieval* dan kualitas jawaban yang dihasilkan karena kedua komponen tersebut saling berkaitan dalam menentukan kinerja sistem.

Retrieval Hit Rate digunakan untuk mengukur proporsi pertanyaan yang memperoleh informasi relevan dari proses *retrieval*. Perhitungan dilakukan menggunakan Persamaan.

$$Retrieval\ Hit\ Rate\ = \frac{N_h}{N_q} \times 100\% \qquad (2)$$

dengan N_h merupakan jumlah pertanyaan yang menghasilkan *retrieval* relevan dan N_q merupakan jumlah seluruh pertanyaan pengujian.

Context Relevance digunakan untuk menilai tingkat relevansi konteks yang diperoleh terhadap pertanyaan pengguna. Nilai metrik diperoleh berdasarkan penilaian terhadap konteks yang dihasilkan oleh proses *retrieval*. *Answer Relevance* digunakan untuk menilai kesesuaian jawaban yang dihasilkan terhadap pertanyaan pengguna, sedangkan *Faithfulness* digunakan untuk menilai kesesuaian informasi pada jawaban dengan konteks yang diberikan. Evaluasi tersebut digunakan untuk mengetahui apakah sistem tidak hanya memperoleh konteks yang relevan, tetapi juga menghasilkan jawaban yang sesuai dengan konteks tersebut. Nilai rata-rata setiap metrik dihitung menggunakan Persamaan

$$Score_{avg} = \frac{\sum_{i=1}^{n} Score_i}{n} \qquad (3)$$

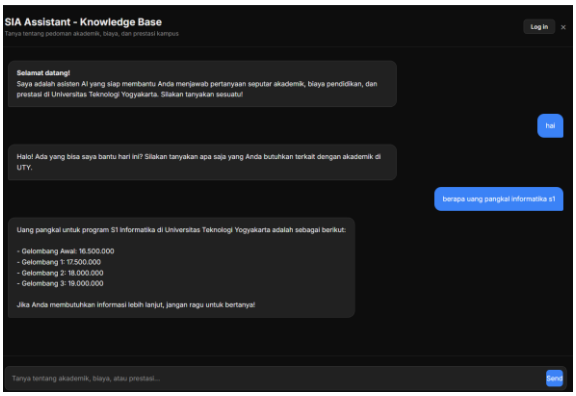
dengan Score_I merupakan nilai penilaian pada pertanyaan ke-i dan n merupakan jumlah pertanyaan pengujian. *Accuracy* digunakan untuk menilai tingkat kesesuaian jawaban sistem dengan jawaban atau informasi acuan yang digunakan dalam pengujian. Evaluasi terhadap *accuracy* digunakan sebagai pelengkap metrik RAG untuk memberikan gambaran mengenai ketepatan jawaban yang dihasilkan sistem.

Seluruh hasil pengujian kemudian digunakan untuk mengevaluasi sistem *chatbot* akademik yang telah dikembangkan. Hasil *Black Box Testing* digunakan untuk mengetahui keberhasilan fungsi utama sistem, sedangkan evaluasi RAG digunakan untuk menilai kemampuan sistem dalam memperoleh informasi yang relevan dan menghasilkan jawaban berdasarkan konteks. Dengan demikian, tahapan penelitian mencakup keseluruhan proses mulai dari identifikasi permasalahan, pengumpulan dan pengolahan data, pembentukan basis pengetahuan, perancangan arsitektur, implementasi RAG dan LLM, integrasi data akademik personal, hingga pengujian dan evaluasi sistem.

3. HASIL DAN ANALISIS

Hasil penelitian meliputi implementasi chatbot akademik berbasis *Large Language Model* (LLM) dengan pendekatan *Retrieval-Augmented Generation* (RAG), pengujian fungsional, dan evaluasi kualitas retrieval serta jawaban. Sistem mengintegrasikan dokumen akademik publik yang disimpan pada ChromaDB dengan data akademik personal mahasiswa yang disimpan pada PostgreSQL. Data personal digunakan setelah pengguna berhasil melakukan autentikasi melalui Sistem Informasi Akademik (SIA) Universitas Teknologi Yogyakarta (UTY). Evaluasi kuantitatif dilakukan melalui sembilan skenario pengujian fungsional dan 50 pertanyaan untuk menilai lima metrik RAG.

Implementasi chatbot menyediakan layanan informasi akademik publik melalui fitur Knowledge Base Chat dan layanan informasi akademik personal melalui fitur Private Chat. Pada layanan publik, pertanyaan diproses melalui preprocessing, pembentukan embedding, dan pencarian semantik pada ChromaDB. Dokumen yang diperoleh dari proses retrieval digunakan sebagai konteks tambahan bagi LLM dalam menghasilkan jawaban. Gambar 3 memperlihatkan implementasi antarmuka Knowledge Base Chat ketika pengguna mengajukan pertanyaan mengenai informasi akademik.



Gambar 3. Implementasi antarmuka *chatbot* akademik

Pada layanan Private Chat, sistem terlebih dahulu memeriksa status autentikasi pengguna. Setelah autentikasi berhasil, sistem dapat menggunakan data akademik personal yang tersimpan pada PostgreSQL, seperti Kartu Hasil Studi (KHS), presensi, pembayaran, dan jadwal kuliah, sesuai kebutuhan pertanyaan. Pengujian fungsional dilakukan menggunakan Black Box Testing terhadap sembilan skenario, meliputi akses halaman utama, login SIA UTY, Knowledge Base Chat, Private Chat, sinkronisasi data akademik, informasi nilai KHS, presensi, pembayaran, dan jadwal kuliah. Berdasarkan hasil pada Tabel 1, seluruh sembilan skenario memperoleh status berhasil.

Tabel 1. Hasil Pengujian *Black Box Testing*

No.	Fitur yang diuji	Skenario	Hasil pengujian	Status
1	Halaman utama	Akses halaman utama	Halaman utama ditampilkan	Berhasil
2	Login SIA UTY	Login menggunakan akun SIA UTY	Autentikasi pengguna berhasil	Berhasil
3	Knowledge Base Chat	Mengajukan pertanyaan akademik umum	Jawaban ditampilkan sesuai knowledge base	Berhasil
4	Private Chat	Mengajukan pertanyaan terkait data akademik pribadi	Informasi akademik pengguna ditampilkan	Berhasil
5	Sinkronisasi data akademik	Melakukan sinkronisasi data	Data akademik berhasil diperbarui	Berhasil
6	Informasi nilai KHS	Menanyakan nilai akademik	Data nilai ditampilkan sesuai basis data	Berhasil
7	Informasi presensi	Menanyakan data presensi	Informasi presensi mahasiswa ditampilkan	Berhasil
8	Informasi pembayaran	Menanyakan tagihan akademik	Informasi pembayaran mahasiswa ditampilkan	Berhasil
9	Informasi jadwal kuliah	Menanyakan jadwal kuliah	Jadwal ditampilkan sesuai KRS	Berhasil

Persentase keberhasilan pengujian fungsional dihitung berdasarkan perbandingan jumlah skenario yang berhasil dengan seluruh skenario yang diuji.

$$\textit{Keberhasilan fungsional} = (\textit{Jumlah skenario berhasil} / \textit{Jumlah seluruh skenario}) \times 100\%$$

$$= (9 / 9) \times 100\% = 100\%$$

Hasil tersebut menunjukkan bahwa seluruh fungsi yang tercakup dalam sembilan skenario telah menghasilkan keluaran sesuai dengan yang diharapkan pada saat pengujian. Dengan demikian, sistem memperoleh tingkat keberhasilan pengujian fungsional sebesar 100%. Nilai ini merepresentasikan keberhasilan pelaksanaan skenario *Black Box Testing*, bukan akurasi jawaban chatbot atau jaminan bahwa seluruh informasi yang dihasilkan selalu benar.

Evaluasi RAG dilakukan terhadap 50 pertanyaan pengujian menggunakan lima metrik, yaitu *Retrieval Hit Rate*, *Context Relevance*, *Answer Relevance*, *Faithfulness*, dan *Accuracy*. Nilai rata-rata yang dilaporkan pada pengujian kemudian dikonversi terhadap skor maksimum skala 5 untuk menyajikan persentase ekuivalen.

$$\textit{Persentase ekuivalen} = (\textit{Skor rata-rata} / 5) \times 100\%$$

Berdasarkan Tabel 2, *Retrieval Hit Rate* dan *Context Relevance* masing-masing memperoleh skor 4,80 dari 5 atau setara dengan 96,0% dari skor maksimum. Hasil tersebut menunjukkan bahwa penilaian terhadap proses pencarian dan relevansi konteks memperoleh skor rata-rata yang tinggi. Pada tahap generasi jawaban, Answer Relevance memperoleh skor 4,66 dari 5 atau 93,2%, sedangkan Faithfulness memperoleh skor 4,70 dari 5 atau 94,0%. Kedua hasil tersebut menunjukkan bahwa jawaban yang dihasilkan dinilai relevan terhadap pertanyaan dan memiliki kesesuaian yang tinggi dengan konteks yang diberikan.

Tabel 2. *Retrieval Hit Rate* dan *Context Relevance*

Metrik evaluasi	Skor rata-rata	Persentase *ekuivalen*
Retrieval Hit Rate	4,80/5	96,0%
Context Relevance	4,80/5	96,0%
Answer Relevance	4,66/5	93,2%
Faithfulness	4,70/5	94,0%
Accuracy	4,64/5	92,8%

Metrik *Accuracy* memperoleh skor rata-rata 4,64 dari 5 atau setara dengan 92,8% dari skor maksimum. Nilai ini menunjukkan hasil penilaian ketepatan jawaban yang relatif tinggi berdasarkan skala evaluasi yang digunakan. Namun, persentase 92,8% pada tabel merupakan konversi skor rata-rata terhadap skor maksimum, bukan persentase jumlah jawaban yang benar dari 50 pertanyaan. Untuk menyatakan akurasi aktual, diperlukan jumlah jawaban yang dinilai benar berdasarkan kunci jawaban atau ground truth yang telah ditetapkan. Oleh

karena itu, dalam penelitian ini istilah yang digunakan adalah persentase ekuivalen skor Accuracy, bukan akurasi aktual sebesar 92,8%.

Hal yang sama perlu diperhatikan pada pelaporan *Retrieval Hit Rate*. Dalam metode, metrik ini didefinisikan sebagai proporsi pertanyaan yang memperoleh hasil retrieval relevan. Akan tetapi, hasil yang disajikan saat ini berupa skor rata-rata 4,80/5. Oleh sebab itu, nilai 96,0% pada Tabel 2 hanya dapat disebut persentase ekuivalen skor apabila 4,80/5 memang berasal dari skala penilaian. Jika peneliti bermaksud melaporkan hit rate dalam pengertian proporsi, jumlah pertanyaan yang memperoleh *retrieval* relevan dari total 50 pertanyaan perlu dihitung kembali dari catatan pengujian.

Secara keseluruhan, skor evaluasi memperlihatkan bahwa lima aspek yang dinilai memperoleh persentase ekuivalen antara 92,8% hingga 96,0%. Hasil tersebut memberikan gambaran kuantitatif mengenai penilaian terhadap proses retrieval, relevansi konteks, relevansi jawaban, faithfulness, dan accuracy. Meskipun demikian, interpretasi hasil tetap bergantung pada prosedur pemberian skor, kriteria penilaian, dan konsistensi evaluator. Penjelasan mengenai siapa yang melakukan penilaian, rubrik skala 1–5, serta penggunaan jawaban acuan perlu dipastikan dalam metode agar hasil dapat direplikasi.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan *chatbot* akademik berbasis *Large Language Model* (LLM) dengan pendekatan *Retrieval-Augmented Generation* (RAG) untuk mengintegrasikan informasi akademik publik dan data akademik personal mahasiswa Universitas Teknologi Yogyakarta (UTY) melalui satu antarmuka percakapan. Sistem memanfaatkan ChromaDB sebagai basis pengetahuan dokumen akademik publik dan PostgreSQL untuk menyimpan data akademik personal yang diakses setelah pengguna melakukan autentikasi melalui Sistem Informasi Akademik (SIA). Hasil pengujian fungsional menggunakan *Black Box Testing* menunjukkan bahwa seluruh sembilan skenario berhasil dijalankan dengan tingkat keberhasilan sebesar 100%. Sementara itu, evaluasi terhadap 50 pertanyaan menghasilkan skor *Retrieval Hit Rate* dan *Context Relevance* masing-masing sebesar 4,80/5 (96,0%), *Answer Relevance* sebesar 4,66/5 (93,2%), *Faithfulness* sebesar 4,70/5 (94,0%), serta *Accuracy* sebesar 4,64/5 (92,8% dari skor maksimum). Hasil tersebut menunjukkan bahwa sistem mampu menjalankan fungsi yang diuji serta memperoleh penilaian yang tinggi pada aspek retrieval dan kualitas jawaban berdasarkan skala evaluasi yang digunakan. Penelitian selanjutnya disarankan memperluas variasi pertanyaan, memperjelas validasi jawaban berdasarkan ground truth, dan mengevaluasi sistem dengan lebih banyak pengguna.

REFERENSI

[1] C. McGrath, A. Farazouli, and T. Cerratto-Pargman, “Generative AI chatbots in higher education: a review of an emerging research area,” Jun. 01, 2025, *Springer Science and Business Media B.V.* doi: 10.1007/s10734-024-01288-w.

[2] D. Yigci, M. Eryilmaz, A. K. Yetisen, S. Tasoglu, and A. Ozcan, “Large Language Model-Based Chatbots in Higher Education,” Mar. 01, 2025, *John Wiley and Sons Inc*. doi: 10.1002/aisy.202400429.

[3] F. Arifin, H. Sibyan, and N. Hasanah, “RANCANG BANGUN CHATBOT PADA SISTEM EKAPTA BERBASIS NATURAL LANGUAGE PROCESSING DENGAN ALGORITMA ARTIFICIAL NEURAL NETWORK,” *Jurnal Ilmiah Informatika dan Komputer*, vol. 4, no. 1, pp. 1–8, 2025, [Online]. Available: https://ojs.unsiq.ac.id/index.php/biner

[4] N. R. P. Syallya, A. A. Pravitasari, and A. Helen, “NLP-Based Intent Classification Model for Academic Curriculum Chatbots in Universities Study Programs,” *Jurnal RESTI*, vol. 9, no. 1, pp. 111–117, Feb. 2025, doi: 10.29207/resti.v9i1.6276.

[5] H. Herlawati and R. T. Handayanto, “Fine-Tuning Large Language Model (LLM) for Chatbot with Additional Data Sources,” *PIKSEL : Penelitian Ilmu Komputer Sistem Embedded and Logic*, vol. 13, no. 1, pp. 125–132, Mar. 2025, doi: 10.33558/piksel.v13i1.10832.

[6] U. Ariq Prathama, A. Purwarianti, and S. Cahyawijaya, “Exploring The Effectiveness of In-Context Methods in Human-Aligned Large Language Models Across Languages,” *JUTI: Jurnal Ilmiah Teknologi Informasi*, vol. 23, no. 2, pp. 13–31, 2025, doi: 10.12962/j24068535.v23i1.a1323.

[7] M. A. Khuzaini, I. Ratna, I. Astutik, A. S. Fitrani, and Y. Findawati, “Chatbot WhatsApp Asisten Workout dengan Integrasi Large Language Model Gemini dan spaCy,” *semanTIK*, vol. x, No.x, no. 2, pp. 53–63, 2025.

[8] W. M. H. Sasmita, S. Sumpeno, and R. F. Rachmadi, “Improving Government Helpdesk Service With an AI-Powered Chatbot Built on the Rasa Framework,” *Jurnal RESTI*, vol. 9, no. 2, pp. 393–403, Apr. 2025, doi: 10.29207/resti.v9i2.6293.

[9] M. L. Husain, Y. Wibisono, and A. Anisyah, "Development of an Academic Services Chatbot Based on Retrieval-Augmented Generation (RAG)," *Brilliance: Research of Artificial Intelligence*, vol. 5, no. 2, pp. 727–735, Aug. 2025, doi: 10.47709/brilliance.v5i2.6719.

[10] A. Yuliana, N. L. Anggreini, R. Iskandar, and G. R. Prasanth, "A Hybrid LSTM–Smith Waterman Model for Personalized Semantic Search in Academic Information Systems," *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 4, pp. 2139–2152, Aug. 2025, doi: 10.52436/1.jutif.2025.6.4.4763.

[11] Z. Li, Z. Wang, W. Wang, K. Hung, H. Xie, and F. L. Wang, "Retrieval-augmented generation for educational application: A systematic survey," Jun. 01, 2025, *Elsevier B.V.* doi: 10.1016/j.caeai.2025.100417.

[12] J. Swacha and M. Gracel, "Retrieval-Augmented Generation (RAG) Chatbots for Education: A Survey of Applications," Apr. 01, 2025, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/app15084234.

[13] I. Pujiono, I. M. Agtyaputra, and Y. Ruldeviyani, "IMPLEMENTING RETRIEVAL-AUGMENTED GENERATION AND VECTOR DATABASES FOR CHATBOTS IN PUBLIC SERVICES AGENCIES CONTEXT," *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 10, no. 1, pp. 216–223, Aug. 2024, doi: 10.33480/jitk.v10i1.5572.

[14] M. R. Anam, A. S. Akbar, and H. Saputro, "QnA Chatbot with Mistral 7B and RAG method: Traffic Law Case Study," *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, vol. 15, no. 03, p. 207, Jan. 2025, doi: 10.24843/lkjiti.2024.v15.i03.p06.

[15] W. X. Zhao, J. Liu, R. Ren, and J. R. Wen, "Dense Text Retrieval Based on Pretrained Language Models: A Survey," *ACM Trans. Inf. Syst.*, vol. 42, no. 4, Feb. 2024, doi: 10.1145/3637870.

[16] Ahmad Fajri and Rila Mandala, "OPTIMIZING RETRIEVAL-AUGMENTED GENERATION FOR DOMAIN-SPECIFIC KNOWLEDGE SYSTEMS THROUGH FINE-TUNING AND PROMPT ENGINEERING," *JURTEKSI (jurnal Teknologi dan Sistem Informasi)*, vol. 12, no. 1, pp. 137–144, Dec. 2025, doi: 10.33330/jurteksi.v12i1.4338.

[17] H. Tohir, N. Merlina, and M. Haris, "UTILIZING RETRIEVAL-AUGMENTED GENERATION IN LARGE LANGUAGE MODELS TO ENHANCE INDONESIAN LANGUAGE NLP," *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 10, no. 2, pp. 352–360, Nov. 2024, doi: 10.33480/jitk.v10i2.5916.

[18] S. Ismiati, F. P. Aditiawan, and A. L. Nurlaili, "BLACK BOX TESTING WITH THE EQUIVALENCE PARTITIONING AND CAUSE EFFECT GRAPH METHOD IN ARCHIVE INFORMATION SYSTEM," *Jurnal Teknik Informatika (Jutif)*, vol. 5, no. 4, pp. 981–990, Jul. 2024, doi: 10.52436/1.jutif.2024.5.4.1944.