Journal of Innovative and Creativity
Journal Homepage: https://joecy.org/index.php/joecy
e-mail: galungwaknook@gmail.co

Faculty of Education
University of Pahlawan Tuanku Tambusai
ISSN 2775-771X (Online)
ISSN 2962-570X (Printed)
6 (2) 2026, Page: 41656-41666

Perancangan Chatbot *Open-Domain* Berbasis *Retrieval Augmented Generation* dan *Large Language* Model untuk Generasi Digital

Afif Putra Romadhoni[1], Lusia Rakhmawati[2], I Gusti Putu Asto Buditjahjanto[3], Nurhayati[4]

[1,2,3,4] Universitas Negeri Surabaya

Corresponding Author: Afif Putra Romadhoni, Lusia Rakhmawati, I Gusti Putu Asto Buditjahjanto, Nurhayati
e-mail: afif.22097@mhs.unesa.ac.id[1], lusiarakhmawati@unesa.ac.id[2], asto@unesa.ac.id[3], nurhayati@unesa.ac.id[4]

Published: Juli 2026

ABSTRACT

The rapid advancement of artificial intelligence, particularly Large Language Models (LLMs), has driven the development of chatbots capable of generating more natural and interactive conversations. However, general-purpose chatbots still have limitations in providing personalized interactions that meet users' specific needs. This study aims to design and implement an open-domain chatbot based on a Large Language Model (LLM) and Retrieval-Augmented Generation (RAG) with mood personalization and response tone features. The system was developed as a web-based application using an LLM as the response generator and a keyword-based RAG mechanism to provide additional context from a knowledge base. The novelty of this research lies in the integration of the RAG mechanism, mood personalization, response tone selection, and a safety layer within a single chatbot system designed to deliver a more personalized and secure conversational experience. System evaluation was conducted through response time testing and user acceptance testing using the Technology Acceptance Model (TAM). The results indicate that all designed features were successfully implemented and that the system was capable of generating responses across various conversational scenarios. Furthermore, the evaluation demonstrated a high level of user acceptance in terms of perceived usefulness, perceived ease of use, attitude toward using, and behavioral intention to use. Therefore, the developed chatbot was successfully implemented and can serve as a digital interaction medium that leverages LLM and RAG technologies to provide a more personalized conversational experience for users.

Keywords: Chatbot; Large Language Model; Response Personalization; Retrieval-Augmented Generation; Technology Acceptance Model.

© The Author(s). 2021 Open Access This article is licensed under a Creative Commons Atibution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0

PENDAHULUAN

Perkembangan teknologi kecerdasan buatan (*Artificial Intelligence*/AI) dalam beberapa tahun terakhir telah memberikan dampak signifikan pada berbagai bidang kehidupan, termasuk pendidikan, kesehatan, dan hiburan. Salah satu implementasi AI yang paling banyak dikembangkan adalah chatbot, yaitu sistem percakapan berbasis komputer yang dirancang untuk meniru interaksi manusia melalui teks maupun suara (Nurul dkk., 2022). Chatbot awal seperti ELIZA dan ALICE hanya mampu merespons pola kalimat sederhana, sedangkan chatbot modern dengan dukungan *Large Language Model* (LLM) mampu menghasilkan percakapan yang lebih alami dan kontekstual (Zhou dkk., 2020). Di Indonesia, kebutuhan akan chatbot semakin meningkat seiring tingginya penggunaan internet dan media sosial, terutama di kalangan generasi muda berusia 17–25 tahun yang cenderung aktif mencari wadah komunikasi digital yang cepat, responsif, dan fleksibel untuk memenuhi kebutuhan informasi maupun hiburan (Putra & Santoso, 2021).

Chatbot merupakan agen perangkat lunak yang menyediakan akses layanan dan informasi melalui interaksi bahasa sehari-hari yang mudah dimengerti pengguna, baik melalui teks maupun suara

(Skjuve dkk., 2021). Perkembangan chatbot modern tidak lagi hanya bertumpu pada aturan sederhana, melainkan didukung oleh *Large Language Model* (LLM), yaitu model bahasa berskala besar yang dibangun menggunakan arsitektur *Transformer* dengan mekanisme *self-attention* untuk memahami hubungan antarkata dalam konteks percakapan (Vaswani dkk., 2017). LLM memiliki kemampuan *few-shot learning*, yaitu menyelesaikan berbagai tugas bahasa hanya dengan sedikit contoh atau instruksi, sehingga banyak digunakan pada chatbot, sistem tanya jawab, dan asisten virtual (Brown dkk., 2020). Meskipun demikian, LLM tetap memiliki keterbatasan berupa kemungkinan menghasilkan jawaban yang tidak akurat atau tidak sesuai konteks, sehingga diperlukan pendekatan pendukung untuk meningkatkan relevansi jawaban.

Untuk mengatasi keterbatasan tersebut, penelitian ini menerapkan *Retrieval-Augmented Generation* (RAG) yang menggabungkan proses pencarian informasi (*retrieval*) dengan kemampuan generatif model bahasa (*generation*) agar jawaban lebih relevan dan berbasis konteks (Lewis dkk., 2020). RAG terdiri atas tiga tahap utama, yaitu *retrieval*, *augmentation*, dan *generation*, dan penambahan konteks eksternal terbukti mampu meningkatkan relevansi serta konsistensi jawaban pada sistem berbasis LLM (Gao dkk., 2023). Pada sistem ini, RAG diterapkan melalui pendekatan *keyword-based retrieval* dengan mencocokkan kata kunci pada input pengguna terhadap *knowledge base* berbasis JSON karena lebih ringan dan sederhana. Selain itu, sistem dilengkapi *safety layer* berbasis aturan untuk menyaring masukan berisiko, mengingat LLM berpotensi menghasilkan bias, misinformasi, maupun konten tidak aman (Weidinger dkk., 2021). Tingkat penerimaan pengguna dievaluasi menggunakan *Technology Acceptance Model* (TAM) yang menilai persepsi kegunaan dan kemudahan penggunaan sebagai penentu penerimaan teknologi (Davis, 1989).

Sebagian besar chatbot yang tersedia saat ini masih bersifat *task-oriented* atau terbatas pada layanan tertentu, seperti pemesanan tiket, layanan pelanggan, atau asisten belanja daring. Keterbatasan tersebut menyebabkan minimnya chatbot yang mampu menjadi teman diskusi terbuka (*open-domain*) dengan percakapan yang adaptif. Selain itu, permasalahan lain yang muncul adalah kurangnya pengelolaan informasi yang akurat dan aman dalam sistem chatbot. Banyak chatbot generatif berpotensi memberikan jawaban yang tidak relevan, bias, atau bahkan menimbulkan dampak negatif bagi pengguna (Fan dkk., 2025).

Oleh karena itu, dibutuhkan sebuah pendekatan yang tidak hanya memanfaatkan kekuatan LLM, tetapi juga mengintegrasikan *Retrieval-Augmented Generation* (RAG) untuk mendukung penyediaan informasi berbasis sumber data yang terverifikasi. Dengan adanya RAG, chatbot dapat merujuk pada basis pengetahuan tertentu sehingga kualitas jawaban menjadi lebih faktual dan terarah (Fan dkk., 2025). RAG pertama kali diperkenalkan sebagai metode yang menggabungkan proses pencarian informasi eksternal dengan kemampuan generatif model bahasa agar jawaban yang dihasilkan lebih relevan dan berbasis konteks (Lewis dkk., 2020).

Meskipun integrasi LLM dan RAG mampu meningkatkan kualitas dan relevansi jawaban, pendekatan ini juga menambahkan beberapa tahapan pemrosesan sistem, seperti pencarian dokumen relevan dan penggabungan konteks ke dalam prompt. Tahapan tambahan tersebut berpotensi meningkatkan waktu respons atau latensi sistem. Latensi yang tinggi dapat menurunkan kualitas pengalaman pengguna dan menyebabkan berkurangnya kenyamanan dalam berinteraksi dengan sistem digital (Nielsen, 1993). Penelitian lain juga menunjukkan bahwa respons sistem yang lambat dapat berdampak negatif terhadap kepuasan dan keterlibatan pengguna, terutama pada aplikasi berbasis percakapan *real-time* (Seow, 2008). Oleh karena itu, pengukuran latensi menjadi aspek penting untuk memastikan chatbot open-domain berbasis LLM dan RAG tetap responsif dan layak digunakan.

Selain aspek teknis, keberhasilan implementasi chatbot juga sangat dipengaruhi oleh tingkat penerimaan pengguna (*user acceptance*). Menurut *Technology Acceptance Model* (TAM), penerimaan teknologi ditentukan oleh persepsi kegunaan (*perceived usefulness*) dan kemudahan penggunaan (*perceived ease of use*) (Davis, 1989). Dalam konteks chatbot, meskipun sistem mampu menghasilkan jawaban yang akurat, pengguna belum tentu menerima dan menggunakan sistem tersebut apabila merasa interaksinya rumit, lambat, atau tidak nyaman. Generasi muda sebagai target pengguna memiliki ekspektasi tinggi terhadap kecepatan respons, kemudahan penggunaan, dan pengalaman interaksi yang natural (Venkatesh & Davis, 2000).

Berdasarkan kondisi tersebut, penelitian ini bertujuan mengembangkan sebuah chatbot open-domain yang memanfaatkan kombinasi LLM dan RAG agar dapat berperan sebagai teman virtual bagi generasi muda. Kehadiran *safety layer* juga ditambahkan untuk meminimalisir keluaran jawaban yang tidak sesuai konteks atau berpotensi menimbulkan risiko bagi pengguna. Dengan demikian, penelitian ini diharapkan tidak hanya menghasilkan chatbot yang informatif dan aman, tetapi juga memiliki latensi yang dapat diterima serta tingkat penerimaan pengguna yang baik.

METODE

Penelitian ini menggunakan pendekatan kuantitatif deskriptif karena berfokus pada pengukuran data numerik yang diperoleh dari hasil pengujian sistem dan penilaian pengguna, meliputi nilai *response time* chatbot dan skor *user acceptance* melalui kuesioner berbasis skala *Likert* 1–5. Subjek penelitian pada pengujian penerimaan pengguna adalah 30 responden yang ditentukan menggunakan teknik *purposive sampling*, yakni pengguna yang telah mencoba chatbot secara langsung sebelum mengisi kuesioner. Instrumen yang digunakan terdiri atas dua alat ukur, yaitu pencatatan otomatis *response time* melalui fungsi *performance.now()* yang direkam ke *Google Spreadsheet*, serta kuesioner *User Acceptance Test* (UAT) yang disusun berdasarkan empat konstruk *Technology Acceptance Model* (TAM), yaitu *Perceived Usefulness* (PU), *Perceived Ease of Use* (PEOU), *Attitude Toward Using* (ATU), dan *Behavioral Intention* (BI) (Davis, 1989).

Alur penelitian dilakukan secara kronologis, dimulai dari identifikasi masalah, studi literatur, analisis kebutuhan sistem, perancangan arsitektur chatbot, implementasi sistem menggunakan *Next.js* dan *TypeScript* dengan mengintegrasikan GPT-4o mini melalui *OpenAI API*, hingga pengujian dan analisis hasil. Chatbot dibangun dengan mengintegrasikan modul deteksi emosi, *safety layer*, *keyword-based retrieval* terhadap *knowledge base* JSON, *prompt builder*, dan LLM sebagai pembangkit respons. Pengujian *response time* dilakukan sebanyak sepuluh kali percobaan pada tiga variasi *tone*, yaitu *friendly*, *professional*, dan *angry*, kemudian dianalisis menggunakan nilai rata-rata (*mean*). Data *user acceptanœ* dianalisis menggunakan perangkat lunak *Jamovi*, dengan uji validitas melalui nilai *Item-Rest Correlation* dan uji reliabilitas melalui koefisien *Cronbach's Alpha*, selanjutnya nilai setiap konstruk dan rata-rata TAM dihitung untuk menentukan tingkat penerimaan pengguna terhadap sistem.

HASIL DAN PEMBAHASAN

Implementasi Sistem

Implementasi sistem dilakukan berdasarkan rancangan arsitektur yang telah disusun, dengan menerjemahkan rancangan menjadi sebuah chatbot yang dapat menerima masukan pengguna, memahami konteks percakapan, menghasilkan respons yang sesuai, serta menyimpan data hasil pengujian. Sistem dikembangkan menggunakan arsitektur *client-server*, di mana antarmuka pengguna (*frontend*) berfungsi sebagai media interaksi, sedangkan *backend* menjalankan seluruh mekanisme pemrosesan pesan. Pada *backend*, sistem mengintegrasikan modul deteksi emosi (*Emotion Detection*), *Safety Layer*, *Keyword-Based Retrieval*, *Prompt Builder*, *Large Language Model* (GPT-4o mini), serta modul pencatatan *response time*. Proses dimulai ketika pengguna mengirimkan pesan, kemudian pesan divalidasi, dianalisis emosinya, diperiksa keamanannya, dicarikan konteks dari *knowledge base*, disusun menjadi *prompt*, dan dikirim ke LLM untuk menghasilkan respons yang selanjutnya dikembalikan kepada pengguna beserta pencatatan *response time*.

Tabel 1. Komponen Implementasi Sistem

Komponen	Fungsi
Frontend	Menyediakan antarmuka bagi pengguna untuk mengirim pesan dan menerima respons chatbot.
Backend	Mengelola seluruh proses pemrosesan pesan pengguna.
Emotion Detection	Mengidentifikasi kondisi emosi berdasarkan isi pesan pengguna.
Safety Layer	Melakukan penyaringan masukan yang berpotensi tidak aman sebelum diteruskan ke proses berikutnya.

Komponen	Fungsi
Keyword-Based Retrieval	Mencari informasi yang relevan dari *knowledge base* berdasarkan kata kunci pada pesan pengguna.
Prompt Builder	Menggabungkan pesan pengguna, konteks hasil *retrieval*, emosi, bahasa, dan gaya respons menjadi *prompt*.
Large Language Model	Menghasilkan respons berdasarkan *prompt* yang telah disusun.
Logger	Mencatat *response time* sebagai data pengujian sistem.

Berdasarkan Tabel 1, seluruh komponen sistem bekerja secara terintegrasi dalam satu alur pemrosesan yang utuh. Sistem menggunakan pendekatan modular sehingga setiap komponen memiliki fungsi spesifik namun tetap saling terhubung, mulai dari penerimaan masukan pengguna hingga pengiriman respons. Pendekatan modular tersebut mempermudah proses implementasi, pengembangan, serta pemeliharaan karena setiap modul dapat dikembangkan secara terpisah tanpa memengaruhi keseluruhan sistem. Deteksi emosi diterapkan menggunakan pendekatan *rule-based* melalui pencocokan kata kunci, di mana kata seperti "capek", "sedih", "lelah", dan "down" dikategorikan sebagai *sad*; "marah", "kesal", dan "emosi" sebagai *angry*; "bingung", "stress", dan "overwhelmed" sebagai *overwhelmed*; sedangkan masukan lain diberi nilai *neutral*.

Tabel 2. Kategori Emosi

Kata Kunci	Emosi
capek, sedih, lelah, down	*Sad*
marah, kesal, emosi	*Angry*
bingung, stress, overwhelmed	*Overwhelmed*
Selain kata di atas	*Neutral*

Berdasarkan Tabel 2, sistem mengklasifikasikan kondisi emosional pengguna ke dalam empat kategori berdasarkan kecocokan kata kunci pada teks masukan yang sebelumnya telah dinormalisasi menjadi huruf kecil (*lowercase*). Hasil deteksi emosi tidak ditampilkan secara langsung kepada pengguna, tetapi digunakan sebagai salah satu parameter dalam penyusunan *prompt* bersama konteks hasil *retrieval*, gaya respons (*tone*), bahasa, riwayat percakapan, dan profil pengguna. Dengan demikian, respons yang dihasilkan oleh *Large Language Model* dapat disesuaikan dengan kondisi emosional pengguna sehingga interaksi terasa lebih empatik dan personal.

Setelah deteksi emosi, sistem melakukan pemeriksaan *Safety Layer* menggunakan pendekatan *rule-based moderation* untuk mendeteksi kata atau frasa berisiko sebelum pesan diproses lebih lanjut. Selanjutnya, sistem menjalankan *Keyword-Based Retrieval* dengan mencocokkan kata kunci pada pesan pengguna terhadap dokumen pada *knowledge base* dan memilih dokumen dengan skor kecocokan tertinggi sebagai konteks utama. Contoh proses tersebut disajikan pada Tabel 3.

Tabel 3. Contoh Proses *Keyword-Based Retrieval*

Input Pengguna	*Keyword* Cocok	Dokumen Terpilih	Konteks
"Aku sulit tidur setiap malam."	Tidur	*Sleep*	Informasi mengenai kebiasaan tidur sehat.
"Bagaimana cara mengurangi stres?"	Stress	*Stress Management*	Informasi mengenai pengelolaan stres.
"Aku kurang minum air."	Air	Dehidrasi	Informasi mengenai pentingnya konsumsi air putih.

Berdasarkan Tabel 3, mekanisme *Keyword-Based Retrieval* mampu menemukan konteks yang relevan dari *knowledge base* berdasarkan kata kunci yang terdapat pada pesan pengguna. Apabila ditemukan kecocokan, sistem mengambil dokumen paling relevan sebagai konteks tambahan; namun apabila tidak ditemukan dokumen yang cocok, sistem mengembalikan konteks kosong sehingga proses pembangkitan respons tetap dapat dilakukan menggunakan kemampuan generatif LLM.

Konteks hasil *retrieval* tersebut kemudian diteruskan ke tahap *Prompt Builder* untuk digabungkan dengan informasi lain sebelum dikirimkan ke *Large Language Model*.

Tabel 4. Informasi yang Disusun pada *Prompt Builder*

Komponen	Fungsi
Pesan pengguna	Menjadi pertanyaan utama yang akan dijawab oleh LLM.
Context	Memberikan informasi tambahan hasil *Keyword-Based Retrieval*.
Emotion	Menyesuaikan respons dengan kondisi emosional pengguna.
Style	Menentukan gaya bahasa chatbot (*Friendly*, *Professional*, *Angry*).
Language	Menentukan bahasa yang digunakan chatbot.
History	Menjaga kesinambungan percakapan sebelumnya.
Profile	Menambahkan informasi pengguna dan identitas chatbot.

Berdasarkan Tabel 4, *Prompt Builder* menggabungkan tujuh komponen utama menjadi satu instruksi yang utuh sebelum dikirim ke LLM. Penyertaan konteks hasil *retrieval*, hasil deteksi emosi, gaya respons, bahasa, riwayat percakapan, dan profil pengguna memungkinkan model menghasilkan respons yang lebih personal, konsisten, dan relevan. Riwayat percakapan disertakan agar chatbot tidak hanya merespons berdasarkan pesan terakhir, tetapi juga mempertimbangkan percakapan sebelumnya sehingga jawaban yang dihasilkan lebih berkesinambungan.

Tabel 5. Data yang Dicatat oleh *Logger*

Data	Keterangan
Timestamp	Waktu pencatatan interaksi.
Name	Nama pengguna.
Message	Pesan pengguna.
Response	Respons chatbot.
Risk Level	Tingkat risiko hasil *Safety Layer*.
Tone	Gaya respons chatbot.
Mood	Suasana hati pengguna.
Language	Bahasa yang digunakan.
Response Time	Waktu respons sistem (ms).

Berdasarkan Tabel 5, modul *Logger* mencatat sembilan atribut data secara otomatis pada setiap interaksi chatbot ke dalam *Google Spreadsheet* melalui *Google Apps Script*, sehingga seluruh data terdokumentasi tanpa memerlukan pencatatan manual. Data *response time* yang tercatat kemudian menjadi dasar analisis performa sistem pada tahap pengujian, sedangkan atribut lain seperti *risk level* dan *mood* mendukung pemantauan perilaku sistem selama interaksi berlangsung. Setelah respons dihasilkan, sistem mengembalikan hasil pemrosesan ke antarmuka pengguna dalam bentuk objek JSON melalui *HTTP Response* yang memuat jawaban chatbot beserta informasi pendukungnya.

Tabel 6. Atribut Pengiriman *Response* kepada Pengguna

Atribut	Fungsi
answer	Menampilkan jawaban chatbot kepada pengguna.
emotion	Menyimpan hasil deteksi emosi pengguna.
riskLevel	Menunjukkan tingkat risiko hasil *Safety Layer*.
safetyCategory	Menyimpan kategori risiko yang terdeteksi.
responseTime	Menyimpan waktu yang dibutuhkan sistem dalam menghasilkan respons.

Berdasarkan Tabel 6, sistem mengirimkan lima atribut utama sebagai hasil akhir satu siklus interaksi chatbot. Selain jawaban chatbot (*answer*), sistem juga menyertakan hasil deteksi emosi, tingkat

risiko, kategori hasil *Safety Layer*, serta nilai *response time* yang telah melalui seluruh tahapan pemrosesan. Dengan demikian, seluruh komponen sistem bekerja secara terintegrasi untuk menghasilkan respons yang relevan, empatik, dan aman bagi pengguna.

Antarmuka pengguna dibangun menggunakan *framework Next.js* dengan komponen React sehingga menghasilkan tampilan yang responsif dan interaktif. Antarmuka terdiri atas *Landing Page* sebagai halaman pengenalan sistem, halaman *onboarding* untuk mengumpulkan informasi awal pengguna, serta halaman percakapan sebagai media interaksi utama. Komponen masukan pada halaman *onboarding* disajikan pada Tabel 7.

Tabel 7. Komponen Halaman *Onboarding*

Komponen	Fungsi
Nama Pengguna	Digunakan sebagai identitas pengguna selama percakapan berlangsung.
Umur	Menjadi informasi profil pengguna yang dapat digunakan sebagai konteks tambahan.
Jenis Kelamin	Digunakan untuk menentukan nama chatbot sesuai aturan sistem.
Bahasa	Menentukan bahasa yang digunakan chatbot dalam memberikan respons.
Kondisi Emosi (*Mood*)	Menentukan suasana hati awal pengguna sebagai konteks penyusunan *prompt*.
Gaya Respons (*Tone*)	Menentukan gaya penyampaian respons chatbot.

Berdasarkan Tabel 7, halaman *onboarding* menyediakan enam komponen masukan yang harus diisi pengguna, yaitu nama, umur, jenis kelamin, bahasa, kondisi emosional awal (*mood*), dan gaya respons (*tone*). Data tersebut disimpan sementara dan dikirimkan ke *backend* ketika pengguna memulai percakapan, kemudian digunakan bersama pesan pengguna, hasil deteksi emosi, konteks hasil *Keyword-Based Retrieval*, dan riwayat percakapan dalam proses penyusunan *prompt*. Dengan demikian, chatbot mampu menghasilkan respons yang lebih personal, konsisten, dan sesuai dengan preferensi pengguna sejak awal interaksi.

Pengujian Sistem

Pengujian sistem dilakukan untuk mengevaluasi kinerja chatbot yang telah diimplementasikan melalui dua jenis pengujian, yaitu pengujian *response time* dan *User Acceptance Test* (UAT) menggunakan metode *Technology Acceptance Model* (TAM). Pengujian *response time* dilakukan menggunakan tiga variasi *tone*, yaitu *Friendly* (Santai), *Angry* (Marah), dan *Professional* (Profesional), yang masing-masing diuji sebanyak sepuluh kali sehingga diperoleh total tiga puluh data pengujian pada perangkat dan koneksi internet yang sama.

Hasil Pengujian *Response Time*

Tabel 8. Hasil Pengujian *Response Time Tone* Santai

No	Input Pengguna	*Response Time* (ms)
1	Aku sedih karena tidak bisa menabung beberapa minggu terakhir ini	3.552
2	Gatau gimana caranya mengembalikan tabungan lagi	2.346
3	Aku benar-benar menekan pengeluaran sampai tidak jajan sama sekali	2.285
4	Saya capek nih	1.539
5	Cemas karena sedang mengambil data penelitian dan target sidang semakin dekat	7.904
6	Kira-kira sampai bulan Juli waktunya cukup tidak ya?	2.718

No	Input Pengguna	*Response Time* (ms)
7	Bisa kasih contoh rencana harian atau mingguan?	6.468
8	I have so many tasks to do but I feel lazy	4.048
9	Kenapa mas-mas Batam tidak mau membalas chat aku lagi?	2.335
10	Aku selalu merasa semua orang menyebalkan	2.055
Total		35.286

Berdasarkan Tabel 8, pengujian *response time* pada *tone* Santai menghasilkan total 35.286 ms dari sepuluh percobaan sehingga diperoleh rata-rata sebesar 3.528,6 ms (dihitung dari total dibagi jumlah percobaan). Hasil tersebut menunjukkan bahwa chatbot mampu menghasilkan respons dengan waktu yang relatif stabil meskipun terdapat beberapa fluktuasi pada percobaan tertentu, seperti pada percobaan kelima (7.904 ms) yang memuat pertanyaan lebih kompleks. Fluktuasi tersebut dipengaruhi oleh panjang masukan dan waktu inferensi model, namun secara keseluruhan waktu respons masih berada pada rentang yang dapat diterima untuk interaksi sehari-hari.

Tabel 9. Hasil Pengujian *Response Time Tone* Marah

No	Input Pengguna	*Response Time* (ms)
1	Solusinya apa?	1.589
2	Gabisa	2.417
3	Kenapa kamu ambigu?	2.117
4	Obat untuk cemas	2.255
5	Sejauh apa kamu bisa membantu?	2.125
6	Gua lulusan PTN top 10, kerja di perusahaan biasa tidak apa-apa kan?	3.516
7	Masalahnya bapak gua suka meremehkan gua	3.848
8	Mending gua diem aja	2.981
9	Kenapa gua dilahirkan di keluarga seperti ini?	4.312
10	Doain gua dapat kerjaan yang tidak toxic	2.816
Total		27.976

Berdasarkan Tabel 9, pengujian *response time* pada *tone* Marah menghasilkan total 27.976 ms dari sepuluh percobaan sehingga diperoleh rata-rata sebesar 2.797,6 ms. Nilai rata-rata tersebut merupakan yang terendah di antara ketiga *tone* yang diuji, hal ini disebabkan respons pada *tone* Marah cenderung lebih singkat dan langsung sehingga waktu inferensi model menjadi lebih cepat. Hasil ini menunjukkan bahwa chatbot tetap mampu memberikan respons secara konsisten meskipun menggunakan gaya komunikasi yang lebih tegas dan langsung, tanpa mengorbankan kecepatan pemrosesan sistem.

Tabel 10. Hasil Pengujian *Response Time Tone* Profesional

No	Input Pengguna	*Response Time* (ms)
1	Apa perbedaan dari nada jawabanmu?	7.094
2	Apa kamu bisa menjawab pertanyaan mata kuliah?	2.688
3	Bagaimana cara menghapus halaman di Word?	11.048
4	Kenapa cemas itu ada?	4.364
5	Obat cemas terampuh	4.270

No	Input Pengguna	*Response Time* (ms)
6	Psikoterapi untuk cemas	3.998
7	Saya habis perjalanan jauh dan malas mengerjakan skripsi	8.879
8	Saya mengantuk tetapi deadline sore ini	6.894
9	Males ngobrol sama dia pasti gamau ngalah	3.354
10	Gimana menghilangkan rasa hampa dalam hidup	3.856
Total		56.445

Berdasarkan Tabel 10, pengujian *response time* pada *tone* Profesional menghasilkan total 56.445 ms dari sepuluh percobaan sehingga diperoleh rata-rata sebesar 5.644,5 ms. Nilai tersebut merupakan rata-rata tertinggi dibandingkan *tone* Santai dan Marah karena chatbot menghasilkan respons yang lebih formal, panjang, dan terstruktur. Beberapa percobaan menunjukkan waktu respons yang tinggi, seperti percobaan ketiga (11.048 ms) dan percobaan pertama (7.094 ms), yang dipengaruhi oleh kompleksitas jawaban serta kondisi layanan API pada saat pengujian. Meskipun demikian, sistem tetap mampu menghasilkan respons secara lengkap pada seluruh percobaan.

Perbandingan rata-rata *response time* pada ketiga variasi *tone* disajikan pada Tabel 11 untuk memperlihatkan perbedaan performa masing-masing gaya respons.

Tabel 11. Perbandingan Rata-Rata *Response Time* tiap *Tone*

Tone	Rata-Rata *Response Time* (ms)
Santai (*Friendly*)	3.528,6
Marah (*Angry*)	2.797,6
Profesional (*Professional*)	5.644,5

Berdasarkan Tabel 11, *tone* Profesional memiliki rata-rata *response time* tertinggi sebesar 5.644,5 ms, sedangkan *tone* Marah memiliki rata-rata terendah sebesar 2.797,6 ms, dan *tone* Santai berada di posisi tengah dengan 3.528,6 ms. Perbedaan tersebut dipengaruhi oleh panjang dan kompleksitas respons yang dihasilkan pada masing-masing *tone*, di mana respons profesional yang lebih terstruktur membutuhkan waktu inferensi lebih lama. Perubahan *tone* sendiri hanya memengaruhi instruksi pada *Prompt Builder*, sedangkan mekanisme utama chatbot tetap sama, sehingga variasi waktu respons lebih dipengaruhi oleh waktu inferensi LLM dan kondisi layanan API pada saat pengujian.

Hasil Pengujian *User Acceptance*

Pengujian *User Acceptance* dilakukan terhadap 30 responden yang telah menggunakan chatbot dan mengisi kuesioner berdasarkan empat konstruk TAM. Sebelum data dianalisis, dilakukan uji validitas menggunakan nilai *Item-Rest Correlation* dengan batas 0,30, yang hasilnya disajikan pada Tabel 12.

Tabel 12. Hasil Uji Validitas

Variabel	Jumlah Item	Hasil
Perceived Usefulness (PU)	4	Valid
Perceived Ease of Use (PEOU)	5	Valid
Attitude Toward Using (ATU)	4	Valid
Behavioral Intention to Use (BI)	4	Valid

Berdasarkan Tabel 12, seluruh item pernyataan pada variabel *Perceived Usefulness* (PU), *Perceived Ease of Use* (PEOU), *Attitude Toward Using* (ATU), dan *Behavioral Intention to Use* (BI) memenuhi kriteria validitas sehingga seluruh item dinyatakan valid. Dengan demikian, seluruh butir pernyataan layak digunakan sebagai instrumen penelitian dan dapat dilanjutkan pada tahap uji reliabilitas untuk mengetahui konsistensi instrumen.

Tabel 13. Hasil Uji Reliabilitas

Variabel	*Cronbach's Alpha*	Keterangan
Perceived Usefulness (PU)	0,794	Reliabel
Perceived Ease of Use (PEOU)	0,600	Cukup Reliabel
Attitude Toward Using (ATU)	0,673	Cukup Reliabel
Behavioral Intention to Use (BI)	0,664	Cukup Reliabel

Berdasarkan Tabel 13, variabel *Perceived Usefulness* (PU) memperoleh nilai *Cronbach's Alpha* sebesar 0,794 yang menunjukkan tingkat reliabilitas yang baik, sedangkan variabel PEOU, ATU, dan BI masing-masing memperoleh nilai 0,600, 0,673, dan 0,664 yang menunjukkan tingkat konsistensi yang cukup. Secara keseluruhan, hasil uji reliabilitas menunjukkan bahwa instrumen kuesioner memiliki tingkat konsistensi yang memadai sehingga layak digunakan untuk mengukur tingkat penerimaan pengguna terhadap chatbot menggunakan metode TAM. Setelah instrumen dinyatakan valid dan reliabel, dilakukan perhitungan nilai rata-rata setiap konstruk, yaitu PU sebesar 4,18, PEOU sebesar 4,47, ATU sebesar 4,17, dan BI sebesar 4,19, sehingga diperoleh rata-rata TAM keseluruhan sebesar 4,25.

Tabel 14. Hasil Rekapitulasi Pengujian TAM

Variabel	Nilai Rata-rata	Interpretasi
Perceived Usefulness (PU)	4,18	Diterima
Perceived Ease of Use (PEOU)	4,47	Sangat Diterima
Attitude Toward Using (ATU)	4,17	Diterima
Behavioral Intention (BI)	4,19	Diterima
Rata-rata TAM	4,25	Sangat Diterima

Berdasarkan Tabel 14, variabel *Perceived Ease of Use* (PEOU) memperoleh nilai tertinggi sebesar 4,47 dengan kategori Sangat Diterima, yang menunjukkan bahwa pengguna menilai chatbot mudah dipahami dan mudah digunakan. Sementara itu, variabel *Attitude Toward Using* (ATU) memperoleh nilai terendah sebesar 4,17, namun masih berada pada kategori Diterima. Nilai rata-rata TAM keseluruhan sebesar 4,25 termasuk dalam kategori Sangat Diterima. Hasil tersebut menunjukkan bahwa implementasi *Keyword-Based Retrieval*, *Prompt Builder*, *Safety Layer*, dan *Large Language Model* telah berhasil menghasilkan sistem chatbot yang dapat diterima oleh pengguna, sehingga tujuan penelitian untuk mengembangkan chatbot dengan tingkat penerimaan pengguna yang baik telah tercapai.

Pembahasan Implementasi Sistem

Berdasarkan hasil implementasi, chatbot berhasil dikembangkan menggunakan pendekatan *Keyword-Based Retrieval*, *Prompt Builder*, *Safety Layer*, dan *Large Language Model* (LLM) yang saling terintegrasi sehingga membentuk alur pemrosesan mulai dari penerimaan masukan pengguna, deteksi emosi, pemeriksaan keamanan, pencarian konteks, penyusunan *prompt*, hingga pembangkitan respons menggunakan GPT-4o mini. Implementasi *Keyword-Based Retrieval* memungkinkan chatbot memperoleh konteks yang relevan dari *knowledge base* berdasarkan kata kunci pada pesan pengguna, yang kemudian digabungkan dengan hasil deteksi emosi, gaya respons, bahasa, dan riwayat percakapan melalui *Prompt Builder* sebelum dikirimkan ke LLM.

Penerapan *Safety Layer* berperan sebagai mekanisme penyaringan awal terhadap masukan pengguna yang mengandung kata atau frasa berisiko, sehingga chatbot mampu memberikan respons yang lebih aman sebelum proses pembangkitan respons dilakukan. Hal ini sejalan dengan kebutuhan pengendalian keluaran pada sistem berbasis LLM yang berpotensi menghasilkan bias maupun konten tidak aman (Weidinger dkk., 2021). Secara keseluruhan, seluruh komponen sistem telah berjalan sesuai rancangan sehingga tujuan penelitian dalam mengembangkan chatbot berbasis *Retrieval-Augmented Generation* (RAG) dan LLM berhasil direalisasikan.

Pembahasan Hasil Pengujian

Hasil pengujian *response time* menunjukkan bahwa chatbot mampu menghasilkan respons dengan waktu yang relatif cepat pada seluruh variasi *tone*, yaitu *Friendly*, *Professional*, dan *Angry*.

Perbedaan waktu respons yang diperoleh tidak menunjukkan selisih yang signifikan karena perubahan *tone* hanya memengaruhi instruksi pada *Prompt Builder*, sedangkan mekanisme utama chatbot tetap menggunakan proses yang sama. Variasi nilai *response time* lebih dipengaruhi oleh waktu inferensi LLM dan kondisi layanan API pada saat pengujian dilakukan. Hasil ini konsisten dengan pentingnya menjaga latensi tetap rendah agar pengalaman pengguna tidak terganggu (Nielsen, 1993).

Berdasarkan hasil *User Acceptance Test* menggunakan metode TAM, pengguna memberikan penilaian yang baik terhadap chatbot pada aspek *Perceived Usefulness*, *Perceived Ease of Use*, *Attitude Toward Using*, dan *Behavioral Intention to Use*, dengan rata-rata TAM sebesar 4,25 pada kategori Sangat Diterima. Hasil tersebut menunjukkan bahwa chatbot tidak hanya berfungsi dengan baik secara teknis, tetapi juga dapat diterima oleh pengguna dari aspek kegunaan, kemudahan penggunaan, pengalaman penggunaan, serta keinginan untuk menggunakan kembali. Temuan ini memperkuat asumsi TAM bahwa persepsi kegunaan dan kemudahan penggunaan menjadi faktor utama penerimaan teknologi (Davis, 1989). Secara keseluruhan, hasil implementasi dan pengujian menunjukkan bahwa chatbot yang dikembangkan telah memenuhi tujuan penelitian, yaitu menghasilkan chatbot *open-domain* yang mampu memberikan respons secara relevan, empatik, dan aman.

KESIMPULAN

Penelitian ini berhasil merancang dan mengimplementasikan chatbot *open-domain* berbasis *Large Language Model* (LLM) dan *Retrieval-Augmented Generation* (RAG) yang dilengkapi fitur personalisasi suasana hati (*mood*), pilihan nada respons (*tone response*), riwayat percakapan, serta *safety layer* untuk memfilter masukan pengguna yang berpotensi berbahaya. Berdasarkan pengujian performa, chatbot memberikan respons dengan rata-rata *response time* sebesar 3.528,6 ms pada *tone* santai, 2.797,6 ms pada *tone* marah, dan 5.644,5 ms pada *tone* profesional, yang seluruhnya masih berada pada rentang waktu yang dapat diterima untuk penggunaan sehari-hari. Sementara itu, pengujian *user acceptance* menggunakan metode *Technology Acceptance Model* (TAM) memperoleh nilai rata-rata sebesar 4,25 yang termasuk dalam kategori Sangat Diterima, sehingga dapat disimpulkan bahwa chatbot yang dikembangkan dinilai bermanfaat, mudah digunakan, memberikan pengalaman positif, serta memiliki potensi untuk digunakan kembali oleh pengguna pada masa mendatang. Disarankan agar penelitian selanjutnya memperluas *knowledge base*, mengembangkan *safety layer* berbasis pemahaman konteks yang lebih kompleks, serta membandingkan beberapa model LLM untuk memperoleh kualitas dan efisiensi respons yang lebih optimal.

REFERENCES

Adiwardana, D., Luong, M., & So, D. (2020). Towards a human-like open-domain chatbot. *arXiv preprint arXiv:2001.09977*.

Brown, T. B., Mann, B., & Ryder, N. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901.

Chen, X., Wang, S., & Li, Y. (2023). Enhancing educational chatbots with retrieval-augmented generation. *Educational Technology & Society*, 26(4), 112–123.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340.

Fan, Z., Liu, X., & Huang, Y. (2025). Mitigating hallucinations in large language models: A survey. *AI Review*, 34(2), 155–178.

Gao, L., Chen, A., & Lin, J. (2023). Rethinking retrieval-augmented generation in the era of large language models. *arXiv preprint arXiv:2309.00000*.

Hussain, S., Raza, S., & Ali, H. (2021). A survey on conversational agents/chatbots. *Information Systems Frontiers*, 23, 1343–1371.

Lewis, P., Perez, E., & Piktus, A. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*, 33, 9459–9474.

Nielsen, J. (1993). *Usability engineering*. Academic Press.

Nurul, A., Putra, B., & Santoso, D. (2022). Analisis perkembangan chatbot berbasis kecerdasan buatan. *Journal of Information Technology*, 10(2), 45–53.

Putra, R., & Santoso, H. (2021). Preferensi komunikasi digital generasi muda di Indonesia. *Jurnal Teknologi Informasi*, 5(1), 22–31.

Sari, D. F., Rahman, A., & Yuniarto, B. (2022). Pengembangan chatbot edukasi untuk generasi muda berbasis NLP. *Jurnal Teknologi dan Pembelajaran*, 7(2), 88–97.

Seow, S. C. (2008). *Designing and engineering time: The psychology of time perception in software*. Addison-Wesley.

Skjuve, M., Følstad, A., & Brandtzaeg, P. B. (2021). Chatbots as social actors: Exploring user experiences and relationships. *Computers in Human Behavior*, 121, 106769.

Vaswani, A., Shazeer, N., & Parmar, N. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998–6008.

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, 46(2), 186–204.

Weidinger, L., Mellor, J., & Rifkin, H. (2021). Ethical and social risks of harm from language models. *arXiv preprint arXiv:2112.04359*.

Zhou, L., Gao, J., Li, D., & Shum, H. (2020). The design and implementation of XiaoIce, an empathetic social chatbot. *Computational Linguistics*, 46(1), 53–93.