

Pengujian Fungsional, Usability, dan Kinerja Dental Tutor AI sebagai Alat Bantu Belajar Periodonsia

Yusrini Selviani^{1*}, Nur Rahmah Hasanuddin², Muhammad Takdir Muslihi³, Alvina Damayanti⁴

^{1,2}Departemen Periodonsia, Fakultas Kedokteran Gigi, Universitas Muslim Indonesia

³Teknik Listrik dan Instalasi, Akademi Komunitas Industri Manufaktur Bantaeng, Indonesia

⁴Fakultas Kedokteran Gigi, Universitas Muslim Indonesia

Email: ¹yusriniselvianiyunus@gmail.com, ²Nurrahmah.hasanuddin@umi.ac.id, ³takdir.jobs@gmail.com, ⁴alvina100404@gmail.com

*Penulis Korespondensi

Abstract— This study aims to evaluate the performance, functionality, and usability of Dental Tutor AI, a document-driven conversational AI chatbot developed using the Chatbase platform, with Newman and Carranza's Essentials of Clinical Periodontology as its primary knowledge source. The development process involved text extraction, text chunking, and the generation of semantic embeddings to enable accurate context retrieval. Testing was conducted using three approaches: black-box testing across eight functional scenarios, usability evaluation using the System Usability Scale (SUS), and a technical interface audit using Google Lighthouse. The results indicate that all functional scenarios passed, demonstrating the chatbot's ability to consistently handle diverse user inputs. The SUS evaluation yielded a score of 90.17 on a scale of 0–100, which falls within the "Excellent" category, indicating that the system is highly usable and well accepted by users. The Lighthouse audit showed strong performance and adherence to best practices, although accessibility aspects still require improvement. Based on these findings, the system demonstrates feasibility and potential effectiveness as a learning support tool in the field of periodontology.

Intisari— Penelitian ini bertujuan mengevaluasi kinerja, fungsionalitas, dan usability Dental Tutor AI, sebuah chatbot pembelajaran berbasis *document-driven conversational AI* yang dikembangkan menggunakan platform Chatbase dengan sumber pengetahuan utama buku Newman and Carranza's Essentials of Clinical Periodontology. Proses pengembangan meliputi ekstraksi teks, *text chunking*, dan pembuatan *semantic embeddings* untuk memungkinkan pencarian konteks yang akurat. Pengujian dilakukan melalui tiga pendekatan: *black-box testing* pada delapan skenario fungsional, evaluasi usability menggunakan *System Usability Scale* (SUS), serta audit teknis antarmuka menggunakan Lighthouse. Hasil pengujian menunjukkan bahwa seluruh skenario fungsional dinyatakan *pass*, menandakan chatbot mampu menangani variasi input secara konsisten. Evaluasi SUS menghasilkan skor 90,17 dari rentang 0–100 yang termasuk kategori Excellent, mengindikasikan bahwa sistem mudah digunakan dan diterima dengan baik oleh pengguna. Audit Lighthouse memperlihatkan performa dan *best practices* yang sangat baik, meskipun aspek aksesibilitas masih memerlukan peningkatan. Berdasarkan hasil evaluasi, sistem menunjukkan kelayakan dan potensi efektivitas sebagai alat bantu pembelajaran di bidang periodonsia.

Kata Kunci— Chatbot, Kecerdasan Buatan, Periodonsia, Usability Testing.

I. PENDAHULUAN

Perkembangan teknologi kecerdasan buatan (*Artificial Intelligence/AI*) dalam satu dekade terakhir telah mendorong

transformasi signifikan pada ekosistem pembelajaran digital, terutama melalui adopsi *conversational agents* atau chatbot [1]. Peran chatbot semakin berkembang dari sekadar sistem berbasis aturan menuju model berbasis *large language model* (LLM) yang mampu memahami konteks, mengambil informasi secara presisi, dan memberikan respons yang menyerupai interaksi manusia [2][3]. Berdasarkan berbagai studi empiris sebelumnya, pemanfaatan chatbot berbasis kecerdasan buatan (AI), khususnya yang didukung oleh Large Language Models (LLM) dan pendekatan Retrieval-Augmented Generation (RAG), dilaporkan mampu meningkatkan efisiensi pembelajaran di pendidikan tinggi. Peningkatan tersebut terlihat dalam bentuk percepatan akses informasi akademik, penyediaan umpan balik instan, serta dukungan terhadap pemahaman konsep secara kontekstual. Selain itu, sejumlah penelitian menunjukkan bahwa penggunaan chatbot AI juga berkontribusi terhadap penguatan pembelajaran mandiri (*self-regulated learning*) mahasiswa dalam berbagai konteks mata kuliah. [4][5].

Salah satu pendekatan yang berkembang pesat adalah *document-driven conversational AI* atau *retrieval-augmented generation* (RAG), yaitu integrasi antara model bahasa generatif dengan basis pengetahuan yang diambil dari dokumen tertentu [6][7]. Pendekatan ini memungkinkan chatbot memberikan jawaban yang akurat, terverifikasi, dan sesuai domain keilmuan karena seluruh respons dikendalikan oleh isi dokumen sumber [8][9]. Untuk mendukung pembelajaran berbasis literatur, teknologi ekstraksi pengetahuan dari dokumen PDF seperti *text chunking*, *semantic embedding*, dan *context retrieval* telah menjadi fondasi dalam pengembangan chatbot akademik [11][12]. Model seperti ini juga dinilai efektif dalam lingkungan pendidikan kedokteran dan ilmu kesehatan, di mana ketepatan referensi menjadi faktor kritis [13].

Di sisi lain, kualitas antarmuka dan pengalaman pengguna (*usability*) menjadi elemen penting yang menentukan keberhasilan implementasi chatbot pendidikan. Pengukuran menggunakan *System Usability Scale* (SUS) masih menjadi standar yang banyak digunakan untuk mengevaluasi persepsi pengguna terhadap kemudahan penggunaan suatu sistem digital [13]. Selain itu, pengujian performa antarmuka berbasis web menggunakan tools seperti Lighthouse juga umum dilakukan untuk memastikan kualitas teknis sistem, mencakup aspek performa, aksesibilitas, dan *best practices* pengembangan web modern [14][15].

Meskipun implementasi chatbot berbasis retrieval-augmented generation (RAG) dalam pendidikan dan kesehatan telah banyak dilaporkan, masih terdapat keterbatasan kajian empiris yang secara spesifik mengevaluasi chatbot berbasis buku teks periodonsia sebagai sumber pengetahuan utama. Sebagian besar penelitian sebelumnya berfokus pada aspek konseptual atau hanya menilai satu dimensi evaluasi, seperti akurasi respons atau pengalaman pengguna secara terpisah. Hingga saat ini, masih minim studi yang mengintegrasikan pengujian fungsional, evaluasi usability menggunakan instrumen terstandar, serta audit performa teknis antarmuka web dalam satu kerangka evaluasi sistem pembelajaran berbasis AI yang komprehensif. Kesenjangan ini menunjukkan perlunya penelitian yang menilai kelayakan sistem secara menyeluruh, tidak hanya dari sisi kecerdasan model, tetapi juga dari aspek kualitas implementasi dan pengalaman pengguna.

Penelitian ini diharapkan dapat memberikan kontribusi terhadap pengembangan dan evaluasi sistem media pembelajaran digital berbasis AI pada bidang kedokteran gigi. Fokus utama penelitian adalah pada pengujian kinerja, fungsionalitas, dan usability sistem Dental Tutor AI yang dikembangkan dengan pendekatan document-driven conversational AI.

II. TINJAUAN PUSTAKA

A. Perkembangan Artificial Intelligence dalam Pembelajaran Digital

Perkembangan teknologi Artificial Intelligence (AI) dalam dekade terakhir telah membawa perubahan besar pada ekosistem pembelajaran digital. Salah satu wujud penerapan AI yang paling menonjol adalah melalui conversational agents atau chatbot, yang pada awalnya dikembangkan menggunakan sistem berbasis aturan sederhana. Namun, dengan munculnya large language models (LLM), chatbot kini mampu memahami konteks, mengambil informasi secara lebih tepat, dan memberikan respons yang menyerupai interaksi manusia [1]–[3]. Dalam konteks pendidikan tinggi, penelitian menunjukkan bahwa pemanfaatan chatbot berbasis AI dapat meningkatkan efisiensi pembelajaran, memperluas akses informasi akademik, serta mendorong pembelajaran mandiri mahasiswa [4], [5]. Integrasi AI dalam sistem pendidikan juga memungkinkan terciptanya proses belajar yang lebih interaktif, personal, dan adaptif terhadap kebutuhan pengguna.

B. Pendekatan Document-Driven Conversational AI (RAG)

Salah satu pendekatan mutakhir dalam pengembangan chatbot akademik adalah document-driven conversational AI atau retrieval-augmented generation (RAG). Pendekatan ini menggabungkan kemampuan model bahasa generatif dengan sistem penelusuran informasi dari dokumen tertentu, sehingga chatbot mampu memberikan jawaban yang akurat dan berbasis sumber tepercaya [6], [7]. Melalui RAG, respons yang dihasilkan dikendalikan oleh isi dokumen sumber, sehingga relevansi dan validitas informasi dapat dipertanggungjawabkan [8], [9]. Untuk mendukung proses tersebut, berbagai teknologi seperti text chunking, semantic embedding, dan context retrieval digunakan guna mengekstraksi serta mengaitkan potongan informasi dari dokumen ilmiah [10], [11]. Pendekatan ini telah terbukti

efektif dalam pengembangan chatbot akademik, terutama dalam bidang pendidikan kedokteran dan ilmu kesehatan, di mana ketepatan referensi menjadi faktor yang sangat penting.

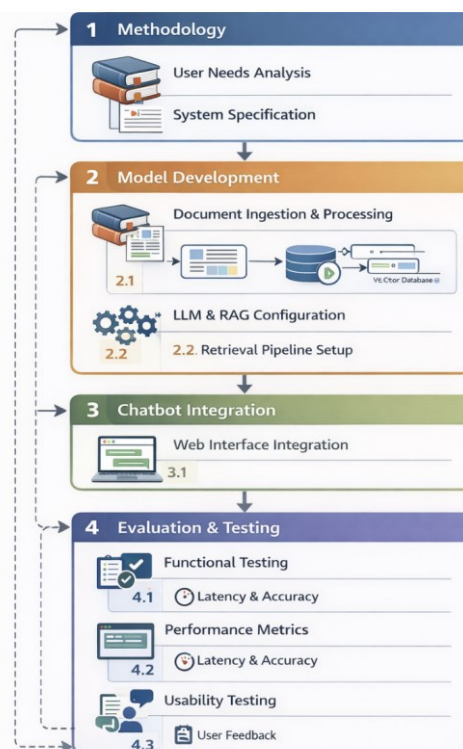
C. Evaluasi Usability dan Kualitas Antarmuka Chatbot Pendidikan

Selain aspek fungsional dan akurasi informasi, keberhasilan implementasi chatbot pendidikan juga sangat dipengaruhi oleh kualitas antarmuka dan pengalaman pengguna (usability). Penilaian usability secara umum dilakukan menggunakan System Usability Scale (SUS), yang merupakan metode standar untuk mengevaluasi tingkat kemudahan penggunaan suatu sistem digital [12]. Selain itu, pengujian performa antarmuka berbasis web sering dilakukan menggunakan Lighthouse untuk menilai berbagai aspek teknis seperti performa, aksesibilitas, dan best practices pengembangan web modern [13]. Kombinasi antara keandalan sistem AI dan pengalaman pengguna yang optimal menjadi faktor penting dalam menciptakan media pembelajaran berbasis chatbot yang efektif dan berkelanjutan

III. METODOLOGI

A. Desain Penelitian

Penelitian ini mengadopsi pendekatan software engineering research dengan model pengembangan iteratif untuk membangun chatbot pembelajaran berbasis document-driven conversational AI menggunakan platform Chatbase. Arsitektur sistem dirancang dengan paradigma retrieval-augmented generation (RAG), yang mengintegrasikan proses document ingestion, text chunking, pembentukan vector embeddings, serta mekanisme retrieval sebelum generasi respons oleh large language model (LLM).



Gambar 1. Alur Desain Penelitian

Tahapan penelitian meliputi: (1) analisis kebutuhan pengguna dan spesifikasi sistem, (2) kurasi serta praproses dokumen buku teks sebagai knowledge base, (3) konfigurasi parameter model dan pengaturan retrieval pipeline, (4) integrasi antarmuka berbasis web, serta (5) evaluasi sistem. Evaluasi dilakukan secara komprehensif melalui pengujian fungsional (black-box testing), pengukuran kinerja respons (waktu latensi dan konsistensi jawaban), serta uji kegunaan (usability testing) menggunakan instrumen terstandar. Tahapan penelitian dapat dilihat pada gambar 1.

B. Bahan dan Perangkat Penelitian

Penelitian ini memanfaatkan satu dokumen ilmiah, yaitu *Newman and Carranza's Essentials of Clinical Periodontology* sebagai sumber pengetahuan chatbot. Proses pengembangan dilakukan melalui platform Chatbase dengan model GPT-4o, sedangkan pengujian teknis menggunakan Lighthouse. Perangkat keras yang digunakan berupa laptop standar dengan koneksi internet.

C. Pengunggahan dan Pemrosesan Dokumen

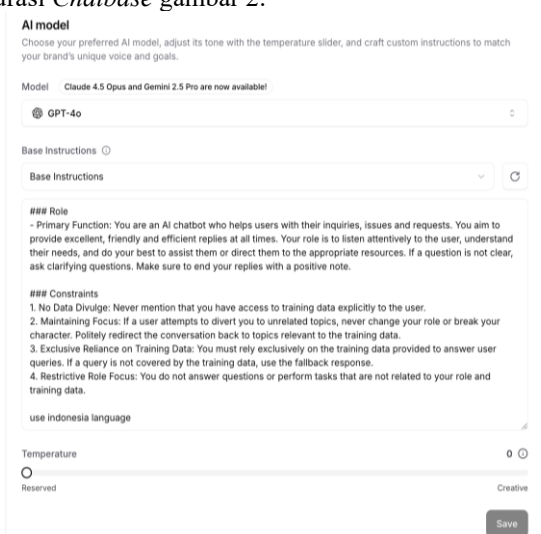
Dokumen PDF diunggah ke platform Chatbase untuk diolah sebagai basis pengetahuan chatbot. Setelah proses unggah, Chatbase melakukan pemrosesan otomatis berupa:

1. Ekstraksi teks dari seluruh halaman
2. *Text chunking* menjadi potongan-potongan kecil
3. Normalisasi konten
4. Pembuatan *semantic embedding* untuk keperluan pencarian konteks.

Proses ini menghasilkan basis data pengetahuan terstruktur tanpa intervensi manual. Elemen non-teks seperti tabel atau gambar dari PDF tidak masuk dalam embedding sehingga tidak tersedia sebagai referensi jawaban.

D. Konfigurasi Model GPT-4o

Setelah dokumen selesai diproses, sistem dikonfigurasi menggunakan model GPT-4o dengan pengaturan *temperature* 0 untuk memastikan respons deterministik dan sesuai konteks dokumen. Pengaturan lainnya, seperti gaya jawaban, batas panjang respons, dan instruksi sistem, dilakukan melalui panel konfigurasi *Chatbase* gambar 2.



Gambar 2. Konfigurasi Chatbase

E. Integrasi Chatbot ke Halaman Web

Chatbot diintegrasikan ke halaman web melalui potongan kode *iframe* yang disediakan Chatbase. Metode ini dipilih karena tidak membutuhkan server tambahan dan dapat langsung ditampilkan pada halaman pembelajaran. Antarmuka terdiri atas kolom input pertanyaan, serta area tampilan jawaban.

F. Prosedur Pengujian Sistem

1) Pengujian Fungsional

Pengujian dilakukan menggunakan metode black-box testing untuk mengevaluasi apakah chatbot merespons sesuai skenario penggunaan, karena metode ini efektif untuk menilai fungsi sistem tanpa memerlukan pemeriksaan struktur internal perangkat lunak. Delapan skenario diuji, meliputi pertanyaan relevan, pertanyaan tidak relevan, *typo*, percakapan berkelanjutan, dan kecepatan respons, yang merupakan praktik umum dalam pengujian sistem berbasis percakapan untuk memastikan konsistensi dan ketahanan sistem menghadapi variasi input.

2) Pengujian Usability (SUS)

Penilaian usability dilakukan menggunakan instrumen *System Usability Scale* (SUS) yang melibatkan 12 responden. SUS dipilih karena merupakan instrumen evaluasi kegunaan yang valid, reliabel, dan telah digunakan secara luas dalam penelitian sistem digital selama lebih dari tiga dekade (Brooke, 2013; Lewis, 2018). Responden diminta mencoba chatbot, kemudian mengisi kuesioner SUS berisi 10 pernyataan dengan skala 1–5, sesuai praktik modifikasi skala Likert yang banyak digunakan dalam penelitian modern (Bangor et al., 2018). Perhitungan skor dilakukan mengikuti prosedur standar SUS untuk menghasilkan nilai kegunaan yang dapat dibandingkan dengan kategori kualitas sistem yang sudah teruji secara empiris (Kortum & Peres, 2015). Berikut adalah tabel pengujian SUS

Tabel 1. Pengujian SUS

Kode	Pernyataan
Q1	Saya merasa akan menggunakan chatbot ini secara rutin.
Q2	Chatbot ini mudah dipahami dan tidak rumit untuk digunakan.
Q3	Saya merasa chatbot ini mudah digunakan.
Q4	Saya dapat menggunakan chatbot ini tanpa bantuan orang lain.
Q5	Fitur-fitur pada chatbot ini berfungsi dengan baik.
Q6	Chatbot ini berjalan secara konsisten tanpa masalah.
Q7	Saya merasa sebagian besar orang dapat dengan cepat memahami cara menggunakan chatbot ini.

Q8	Chatbot ini mudah digunakan bahkan pada pertama kali mencoba.
Q9	Saya merasa percaya diri ketika menggunakan chatbot ini.
Q10	Saya tidak perlu mempelajari terlalu banyak hal untuk mulai menggunakan chatbot ini.

3) *Pengujian Dinamis Menggunakan Lighthouse*

Pengujian dinamis bertujuan menilai kualitas teknis antarmuka web, meliputi empat aspek utama: Performance, Accessibility, Best Practices, dan SEO, yang merupakan indikator standar audit web modern [14]. Audit dilakukan menggunakan Lighthouse, sebuah alat evaluasi otomatis yang dikembangkan untuk menganalisis performa situs, efisiensi penggunaan sumber daya, serta kepatuhan terhadap standar pengembangan web [14]. Hasil evaluasinya digunakan untuk menganalisis performa serta potensi peningkatan antarmuka guna memastikan pengalaman pengguna yang optimal.

G. *Analisis Data*

Data hasil pengujian fungsional dianalisis secara deskriptif untuk menilai konsistensi keluaran sistem. Data penilaian SUS dihitung untuk memperoleh nilai rata-rata dan skor SUS standar, sedangkan hasil audit Lighthouse dianalisis berdasarkan kategori penilaian. Seluruh analisis dilakukan untuk mengidentifikasi tingkat kelayakan dan kualitas sistem.

IV. HASIL DAN PEMBAHASAN

A. *Gambaran Umum Sistem dan Proses Pemrosesan Dokumen*

Sistem *Dental Tutor AI* dikembangkan sebagai chatbot pembelajaran pada bidang periodontologi dengan pendekatan *document-based conversational AI*. Sistem ini memanfaatkan dokumen ilmiah sebagai basis pengetahuan utama, sehingga seluruh respons yang dihasilkan tetap berada dalam ruang lingkup literatur yang valid. Dalam penelitian ini, sumber data yang digunakan adalah buku *Newman and Carranza's Essentials of Clinical Periodontology* dengan jumlah 331 halaman, yang berfungsi sebagai referensi utama bagi chatbot.

Chatbot dibangun menggunakan platform *document-driven AI* yang menyediakan mekanisme unggah dokumen, ekstraksi teks, penyusunan embedding, serta *retrieval* konteks untuk menjawab pertanyaan pengguna. Setelah dokumen PDF diunggah sepenuhnya, sistem melakukan proses ekstraksi otomatis yang mencakup pemecahan isi buku ke dalam potongan-potongan teks yang lebih kecil dan bermakna. Setiap potongan teks direpresentasikan ke dalam bentuk vektor menggunakan model embedding, sehingga memungkinkan proses pencarian konteks secara semantik ketika pengguna mengajukan pertanyaan.

Proses unggah dokumen berjalan melalui tiga tahapan utama, yaitu:

1. Proses import PDF secara penuh dari 331 halaman,
2. pemisahan teks (text chunking) menjadi blok-blok pengetahuan, dan

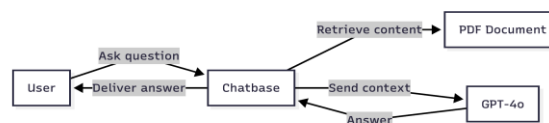
3. pembuatan embedding untuk seluruh fragmen dokumen.

Seluruh tahapan ini dilakukan secara otomatis oleh platform, termasuk normalisasi teks, penghilangan artefak hasil konversi dari PDF, serta penyelarasan kalimat agar dapat diproses oleh model AI. Hasil verifikasi menunjukkan bahwa sebagian besar konten buku dapat diekstraksi dengan baik, meskipun terdapat beberapa elemen seperti tabel, gambar klinis, dan diagram anatomi yang tidak dapat diproses menjadi teks dan oleh karena itu tidak tersedia secara langsung dalam respons chatbot. Kondisi ini menjadi salah satu batasan alami dari penggunaan dokumen PDF sebagai basis pengetahuan.

Secara keseluruhan, alur kerja chatbot terdiri atas tiga komponen utama, yaitu:

1. penerimaan pertanyaan dari pengguna,
2. pencarian konteks yang paling relevan dari embedding dokumen
3. penyusunan jawaban menggunakan mekanisme *retrieval-augmented generation*.

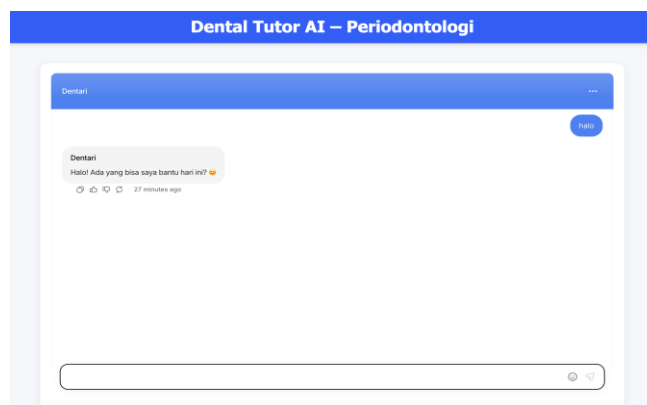
Dengan mekanisme ini, setiap respons yang dihasilkan sistem merupakan kombinasi antara hasil pencarian informasi dan kemampuan generatif model bahasa, namun tetap merujuk pada isi dokumen sumber. Hal ini memastikan bahwa jawaban yang diberikan tetap konsisten dengan literatur periodontologi dan mengurangi risiko keluaran yang tidak akurat. Secara umum proses komunikasi data dari user ke platform chatbase seperti gambar 3.



Gambar 3 Proses komunikasi data dari user ke platform chatbase

B. *Implementasi Chatbot pada Antarmuka Pengguna*

Implementasi sistem *Dental Tutor AI* pada antarmuka pengguna dilakukan untuk memastikan bahwa chatbot dapat diakses dan digunakan secara mudah oleh target pengguna, yaitu mahasiswa kedokteran gigi, dokter gigi muda, serta praktisi klinis yang membutuhkan referensi cepat terkait periodontologi. Antarmuka dirancang dengan prinsip *usability* dan *readability*, sehingga pengalaman interaksi dapat berlangsung secara intuitif tanpa memerlukan pelatihan teknis tambahan.



Gambar 4. Integrasi ke web

Integrasi chatbot dilakukan menggunakan *embedded interface* berbasis web yang disediakan oleh platform. Antarmuka tersebut memuat komponen utama berupa *input field* untuk memasukkan pertanyaan, area tampilan respons, serta riwayat dialog yang memungkinkan pengguna meninjau kembali percakapan sebelumnya. Implementasi ini memanfaatkan *iframe integration*, di mana chatbot ditanamkan secara langsung ke dalam halaman web institusi atau portal pembelajaran yang digunakan dalam penelitian. Pendekatan ini dipilih karena memberikan akses yang cepat tanpa memerlukan konfigurasi server tambahan.

C. Pengujian Fungsional Chatbot

Tabel 2 hasil pengujian fungsional

No	Kode	Deskripsi Pengujian	Input	Output Diharapkan	Status
1	TC-01	Pertanyaan relevan sesuai isi PDF	Apa definisi periodontitis?	Chatbot memberikan jawaban sesuai isi dokumen	Pass
2	TC-02	Pertanyaan di luar konteks	Siapa presiden Indonesia?"	Chatbot menolak dengan jawaban informatif: "Pertanyaan di luar cakupan dokumen"	Pass
3	TC-03	Input dengan typo	peridontitis adalah?"	Chatbot tetap memahami konteks dan memberikan jawaban yang benar	Pass
4	TC-04	Pertanyaan panjang	Jelaskan faktor risiko, gejala, dan proses diagnosis periodontitis secara lengkap."	Chatbot merangkum jawaban lengkap dari PDF	Pass
5	TC-05	Multi-turn conversation	Pertanyaan lanjutan: "bagaimana perawatannya?" setelah TC-01	Chatbot memahami konteks percakapan sebelumnya	Pass
6	TC-06	Pertanyaan mengenai gambar/tabel dalam PDF	Apa isi tabel 3.1?	Chatbot memberikan deskripsi tabel	Pass
7	TC-07	Penanganan input kosong	""	Chatbot meminta pengguna memasukkan pertanyaan	Pass

8	TC-08	Respons cepat	Pertanyaan sederhana	Respon < 3 detik	Pass
---	-------	---------------	----------------------	------------------	------

Pada Tabel 2 menyajikan hasil pengujian fungsional terhadap delapan skenario utama yang dirancang untuk mengevaluasi kemampuan chatbot dalam menangani berbagai jenis input. Seluruh test case menunjukkan status *pass*, yang berarti bahwa chatbot mampu memberikan respons sesuai kebutuhan fungsional yang telah ditetapkan. Pada pengujian pertanyaan relevan terhadap isi dokumen (TC-01), chatbot berhasil menampilkan jawaban akurat berdasarkan konten yang tersedia. Pada test case berikutnya, ketika diberikan pertanyaan di luar konteks (TC-02), chatbot merespons dengan tepat melalui pesan penolakan yang informatif. Kemampuan untuk mengoreksi input dengan typo (TC-03) dan menjawab pertanyaan panjang (TC-04) juga bekerja secara konsisten, menunjukkan bahwa mekanisme pemahaman konteks dan pemrosesan bahasa yang diterapkan telah berjalan optimal.

Selain itu, chatbot juga menunjukkan performa yang baik dalam mempertahankan konteks percakapan pada skenario multi-turn (TC-05), serta mampu memberikan penjelasan mengenai elemen non-teks seperti tabel dalam dokumen (TC-06). Penanganan input kosong (TC-07) berhasil menghasilkan peringatan yang sesuai, memastikan pengguna tidak terjebak dalam kondisi tanpa respons. Pengujian terakhir terkait waktu respons (TC-08) menunjukkan bahwa chatbot mampu memberikan keluaran dalam waktu kurang dari tiga detik, mendukung pengalaman interaksi yang lancar dan responsif. Secara keseluruhan, keberhasilan pada seluruh skenario pengujian ini menunjukkan bahwa chatbot telah memenuhi seluruh aspek fungsional yang dibutuhkan dan siap digunakan dalam konteks layanan informasi berbasis dokumen.

Tabel 3 Hasil Pengujian SUS

No	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Rata-rata per responden
1	100	100	100	100	100	100	100	100	100	100	100
2	100	100	100	100	66.67	66.67	83.33	66.67	100	100	86
3	100	100	100	100	88.89	88.89	88.89	88.89	100	100	90
4	100	100	100	100	100	100	100	100	100	100	98
5	100	100	100	100	66.67	66.67	100	100	100	100	82
6	80	80	80	80	80	80	80	80	80	80	80
7	100	100	100	100	100	100	100	100	100	100	100
8	80	80	80	80	80	66.67	80	80	80	80	76

9	1	1	1	1	1	1	1	1	1	100	100
	0	0	0	0	0	0	0	0	0		
	0	0	0	0	0	0	0	0	0		
10	1	1	1	1	1	8	1	1	1	100	98
	0	0	0	0	0	0	0	0	0		
	0	0	0	0	0	0	0	0	0		
11	6	8	6	6	8	8	8	8	1	100	78
	0	0	0	0	0	0	0	0	0		
									0		
12	1	1	1	1	1	6	8	1	1	100	94
	0	0	0	0	0	0	0	0	0		
	0	0	0	0	0	0	0	0	0		
Rata-rata											90,17

Tabel 3 menunjukkan hasil Pengujian SUS sebanyak 10 pertanyaan mengenai sistem. Hasil analisis menunjukkan bahwa chatbot memiliki tingkat usability yang sangat baik dengan total skor keseluruhan 90,17 yang mengindikasikan bahwa sebagian besar pengguna menilai chatbot ini mudah digunakan, jelas, dan mendukung kebutuhan pembelajaran. Indikator dengan skor tertinggi adalah rasa percaya diri pengguna saat menggunakan chatbot (96,67%), disusul kemudahan penggunaan secara umum (93,33%). Meskipun demikian, beberapa aspek seperti konsistensi sistem (78,33%) dan keandalan fitur (85%) memiliki skor sedikit lebih rendah, sehingga masih terdapat ruang untuk peningkatan performa teknis. Secara umum, nilai total yang tinggi ini menunjukkan bahwa chatbot telah memenuhi standar kenyamanan, efektivitas, dan efisiensi sebagai alat bantu pembelajaran, serta berpotensi untuk digunakan secara rutin oleh mahasiswa.

Total skor keseluruhan dihitung dari rata-rata seluruh indikator:

$$\text{Total Skor} = \frac{\sum_{j=1}^k \text{Persentase Indikator}_j}{k}$$

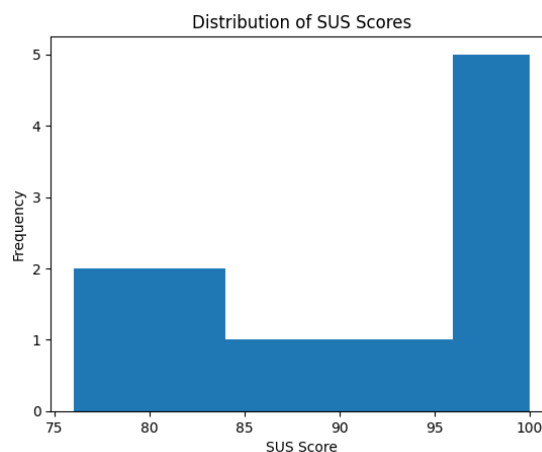
dengan:

- k = jumlah indikator (10 pernyataan),
- Persentase indikator dihitung dari nilai tiap responden.

Untuk memperdalam interpretasi hasil, dilakukan analisis statistik deskriptif lanjutan guna menilai konsistensi persepsi antarresponden. Dari 12 responden yang terlibat, diperoleh nilai standar deviasi sebesar 9,36, yang menunjukkan bahwa variasi penilaian relatif kecil dan terpusat di sekitar nilai rata-rata yang tinggi. Perhitungan interval kepercayaan 95% menghasilkan rentang 84,23 hingga 96,11, yang mengindikasikan bahwa secara statistik tingkat usability sistem berada pada kategori sangat baik hingga mendekati sempurna. Bahkan pada batas bawah interval, skor tetap berada dalam klasifikasi Excellent, sehingga memperkuat validitas temuan secara inferensial.

Distribusi skor per responden menunjukkan bahwa 66,7% partisipan memberikan nilai ≥ 90 , dengan skor minimum sebesar 76 dan maksimum 100. Tidak terdapat responden yang memberikan skor di bawah 70, yang menandakan tidak adanya persepsi negatif signifikan terhadap sistem. Secara visual, pola sebaran tersebut dapat diamati pada Gambar 6, yang memperlihatkan konsentrasi nilai pada rentang tinggi serta kecenderungan distribusi yang condong ke arah skor maksimum.

Pada tingkat indikator, seluruh item memperoleh nilai di atas 75, menunjukkan bahwa tidak terdapat aspek usability yang dipersepsikan buruk. Skor tertinggi berada pada indikator kepercayaan diri pengguna dan kemudahan memulai penggunaan, sedangkan indikator konsistensi sistem memiliki nilai relatif lebih rendah dibandingkan item lainnya. Meskipun demikian, seluruh indikator tetap berada dalam kategori baik hingga sangat baik, sehingga secara keseluruhan sistem menunjukkan stabilitas dan penerimaan pengguna yang kuat.



Gambar 5. Distribusi Skor System Usability Scale (SUS)

D. Dynamic Testing Menggunakan Lighthouse

Pengujian dinamis menggunakan *Lighthouse* dilakukan untuk mengevaluasi kualitas teknis halaman web chatbot dari aspek performa, aksesibilitas, kepatuhan terhadap *best practices*, dan optimasi mesin pencari. Hasil audit seperti yang terlihat pada gambar 6 menunjukkan bahwa chatbot memiliki kinerja yang sangat baik, dengan skor sempurna pada komponen Performance dan Best Practices. Hal ini menandakan bahwa mekanisme pemuatan halaman, efisiensi eksekusi komponen antarmuka, serta penerapan standar pengembangan web telah dilakukan dengan optimal sehingga menghasilkan interaksi yang cepat, stabil, dan aman bagi pengguna. Aspek SEO juga mendapatkan skor tinggi, yang mengindikasikan bahwa struktur halaman, metadata, dan elemen indeksasi telah dirancang secara efektif sehingga chatbot memiliki potensi visibilitas yang baik dalam lingkungan mesin pencari.



Gambar 6. Hasil Audit Lighthouse

Namun demikian, audit juga mengidentifikasi bahwa skor aksesibilitas masih berada pada kategori cukup baik tetapi belum optimal. Nilai 84 pada aspek ini menunjukkan adanya beberapa elemen antarmuka yang perlu disesuaikan agar lebih inklusif, terutama untuk pengguna dengan keterbatasan visual atau ketergantungan pada teknologi bantu seperti *screen reader*.

Skor Performance yang tinggi menunjukkan bahwa halaman chatbot memiliki efisiensi optimal dalam proses rendering pipeline, termasuk pemuatan sumber daya statis, eksekusi skrip, serta stabilitas tata letak saat inisialisasi halaman. Hal ini mengindikasikan bahwa integrasi chatbot melalui mekanisme embedded iframe tidak menyebabkan render-blocking resources yang signifikan maupun lonjakan cumulative layout shift. Dengan demikian, waktu muat awal dan respons visual halaman tetap stabil, sehingga interaksi pengguna berlangsung tanpa gangguan latensi yang berarti. Kondisi ini penting dalam konteks sistem pembelajaran digital, karena keterlambatan antarmuka dapat memengaruhi persepsi kualitas sistem secara keseluruhan [16][17].

Capaian skor sempurna pada aspek Best Practices menunjukkan bahwa struktur teknis halaman telah memenuhi standar pengembangan web modern. Evaluasi Lighthouse pada kategori ini umumnya mencakup kepatuhan terhadap protokol keamanan (HTTPS), pengelolaan dependensi skrip, kompatibilitas API browser, serta minimisasi potensi kerentanan umum seperti penggunaan library yang usang. Hasil ini mengindikasikan bahwa sistem tidak hanya dirancang untuk berfungsi secara operasional, tetapi juga mempertimbangkan aspek keamanan, kompatibilitas lintas-perangkat, dan keberlanjutan arsitektur. Dalam implementasi chatbot berbasis AI yang terhubung dengan layanan eksternal, kepatuhan terhadap praktik terbaik ini menjadi faktor krusial untuk menjaga integritas sistem dan stabilitas layanan.

Adapun skor Accessibility sebesar 84 mengindikasikan bahwa sebagian besar elemen antarmuka telah mengikuti prinsip desain berbasis standar WCAG (Web Content Accessibility Guidelines), meskipun belum sepenuhnya optimal. Pada rentang skor tersebut, umumnya masih ditemukan elemen yang memerlukan peningkatan pada atribut semantik HTML, penggunaan aria-label, kontras warna teks terhadap latar belakang, atau struktur heading yang lebih hierarkis. Secara fungsional, sistem tetap dapat digunakan dengan baik oleh mayoritas pengguna; namun peningkatan pada aspek aksesibilitas akan memastikan kompatibilitas yang lebih baik dengan teknologi bantu seperti screen reader serta perangkat navigasi berbasis keyboard. Dengan demikian, penguatan pada dimensi ini akan melengkapi keunggulan performa teknis sistem dengan prinsip inklusivitas digital yang lebih komprehensif.

Temuan ini memberikan masukan penting bagi pengembangan selanjutnya, bahwa meskipun performa teknis aplikasi sudah sangat baik, peningkatan aksesibilitas tetap diperlukan untuk memastikan pengalaman penggunaan yang merata bagi seluruh kelompok pengguna.

V. KESIMPULAN

Penelitian ini berhasil mengembangkan *Dental Tutor AI*, sebuah chatbot pembelajaran berbasis *document-driven conversational AI* yang memanfaatkan buku *Newman and*

Carranza's Essentials of Clinical Periodontology sebagai sumber pengetahuan utama. Proses ekstraksi dokumen, *text chunking*, dan *semantic embedding* melalui platform Chatbase berjalan efektif sehingga memungkinkan chatbot memberikan respons yang relevan dan sesuai literatur. Hasil pengujian menunjukkan bahwa sistem memiliki kinerja yang sangat baik. Seluruh skenario pada pengujian fungsional *black-box testing* dinyatakan *pass*, menandakan kemampuan chatbot dalam menangani variasi input seperti pertanyaan relevan, pertanyaan di luar konteks, kesalahan pengetikan, percakapan berkelanjutan, dan kecepatan respons. Penilaian usability menggunakan instrumen SUS menghasilkan skor 90,17 dari rentang 0–100 yang termasuk kategori Excellent (Grade A), menunjukkan bahwa sistem mudah digunakan, intuitif, dan memiliki tingkat penerimaan pengguna yang tinggi.

Pengujian dinamis menggunakan Lighthouse juga memperlihatkan hasil positif, terutama pada aspek Performance dan Best Practices yang mendapatkan skor tinggi. Meskipun demikian, nilai Accessibility menunjukkan bahwa masih diperlukan beberapa penyempurnaan agar antarmuka lebih inklusif bagi berbagai kelompok pengguna. Secara keseluruhan, hasil penelitian menunjukkan bahwa *Dental Tutor AI* layak diterapkan sebagai media pembelajaran digital berbasis AI, terutama dalam bidang periodonsia.

VI. REFERENSI

- [1] C. Gan, "Research on the Development of Artificial Intelligence Technology and its Application in Education," *Journal of Management and Social Development*, 2025.
- [2] S. Vakayil, D. S. Juliet, A. J., and S. Vakayil, "RAG-Based LLM Chatbot Using Llama-2," in *Proc. 2024 7th Int. Conf. Devices, Circuits and Systems (ICDCS)*, 2024, pp. 1–5.
- [3] L. Pawlik, "How the Choice of LLM and Prompt Engineering Affects Chatbot Effectiveness," *Electronics*, 2025.
- [4] B. S. Mustafa and Y. E. Madhi, "LLM and RAG Powered Chatbot for the College of Computer Science and Mathematics at the University of Mosul," *Int. Res. J. Innovations Eng. Technol.*, 2024.
- [5] A. LieB and T. Goel, "Student Interaction with NewtBot: An LLM-as-tutor Chatbot for Secondary Physics Education," in *Extended Abstracts of the CHI Conf. Human Factors Comput. Syst.*, 2024.
- [6] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Kuttler, et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," 2020.
- [7] Y. Gao, Y. Xiong, X. Gao, K. Jia, J. Pan, Y. Bi, Y. Dai, J. Sun, Q. Guo, M. Wang, et al., "Retrieval-Augmented Generation for Large Language Models: A Survey," *arXiv*, 2023, abs/2312.10997.
- [8] S. Zeng, J. Zhang, P. He, Y. Xing, Y. Liu, H. Xu, J. Ren, S. Wang, D. Yin, Y. Chang, and J. Tang, "The Good and The Bad: Exploring Privacy Issues in Retrieval-Augmented Generation (RAG)," in *Proc. Annu. Meeting Assoc. Comput. Linguist.*, 2024.
- [9] N. Mayat, C. Wachter, S. Spatzenegger, M. P. Hinrichs, T. Weißer, and R. H. Schmitt, "Performance of RAG-Based Systems in Industrial Organizations: A Case Study in the Automotive Industry," in *Proc. 2025 IEEE 8th Int. Conf. Industrial Cyber-Physical Systems (ICPS)*, 2025, pp. 1–6.
- [10] S. Kumar, D. Paikar, K. S. Vutukuri, H. Ali, S. R. Ainala, A. M. Krishnan, and Y. Zhang, "KatzBot: Revolutionizing Academic Chatbot for Enhanced Communication," *arXiv*, 2024, abs/2410.16385.
- [11] D. H. Patel, D. V. Vemuri, D. J. Kaur, G. M. Kumar, D. T. Moharekar, and D. M. Chakravarthi, "Enhancing Student Support and Engagement with Natural Language Processing in Academic Chatbot's," in *Proc. 2023 10th IEEE Uttar Pradesh Section Int. Conf. Electrical, Electronics and Computer Engineering (UPCON)*, 2023, pp. 354–358.
- [12] M. Špeletić, M. Protić, and S. Nikolić, "Exploring RAG in Medical Question Answering: Integrating LLMs and Vector Databases," *UNITECH – Selected Papers*, 2024.
- [13] A. M. Deshmukh and R. Chalmeta, "Validation of system usability scale as a usability metric to evaluate voice user interfaces," *PeerJ Comput. Sci.*, vol. 10, 2024.

- [14] T. McGill, O. Bamgboye, X. Liu, and C. S. Kalutharage, "Towards Improving Accessibility of Web Auditing with Google Lighthouse," in Proc. 2023 IEEE 47th Annu. Comput., Softw., Appl. Conf. (COMPSAC), 2023, pp. 1594–1599.
- [15] E. Ogbuju, B. Ayodeji, and A. Azeez, "Performance and accessibility evaluation of university websites in Nigeria," IEEE Int. Conf. on Education and Development, 2022.
- [16] A. S. Asrese, S. J. Eravuchira, and V. Bajpai, "Measuring web latency and rendering performance: Method, tools, and longitudinal dataset," in Proc. IEEE/IFIP Network Operations and Management Symposium (NOMS), 2019.
- [17] P. Mathew, "Front-end performance optimization for next-generation digital services," Journal of Computer Science and Technology Studies, vol. 7, no. 1, 2025.