

BAB II

LANDASAN TEORI

2.1 Penelitian Terdahulu

Penelitian Perkembangan chatbot berbasis *Large Language Model (LLM)* menunjukkan bahwa model bahasa mampu menjadi fondasi awal dalam membangun sistem percakapan yang dapat memahami konteks dan merespons pertanyaan pengguna secara natural, terutama pada domain kesehatan. Dalam literatur yang telah di rangkum menegaskan bahwa chatbot berbasis *LLM* dapat memberikan dampak nyata terhadap perilaku pengguna[23], sementara studi lain memperlihatkan bahwa *LLM* memang efektif untuk mendukung interaksi kesehatan dasar[17], [23], [24]. Namun, pada beberapa temuan juga terlihat bahwa *LLM* murni masih memiliki keterbatasan karena tidak selalu terikat pada sumber faktual saat menjawab, sehingga akurasi belum cukup aman untuk konteks layanan informasi yang menuntut keandalan tinggi. Karena itu, walaupun *LLM* membuka peluang besar, pendekatan ini masih perlu dilengkapi dengan mekanisme *grounding* agar jawaban yang dihasilkan tidak hanya natural, tetapi juga benar dan dapat ditelusuri sumbernya.

Keterbatasan tersebut kemudian dijawab oleh pendekatan *Retrieval-Augmented Generation (RAG)*, yang dalam studi literatur muncul sebagai salah satu tema paling kuat. Studi menunjukkan bahwa *RAG* dirancang untuk mengatasi kelemahan *LLM* yang bergantung pada pengetahuan internal dan rentan hallucination[25], [26]. Di beberapa temuan juga ada temuan penting bahwa integrasi *RAG* pada chatbot domain-spesifik dapat meningkatkan akurasi respons hingga 85,55%, yang menunjukkan bahwa mekanisme retrieval memang memberi kontribusi nyata pada kualitas jawaban. Penelitian lain memperkuat bahwa *RAG* cocok digunakan pada domain berbasis dokumen karena jawaban lebih terikat pada basis pengetahuan yang valid[27], [28]. Secara teoritis, ini penting bagi penelitian ini karena layanan PMI membutuhkan jawaban yang bukan hanya cepat, tetapi juga akurat, relevan, dan konsisten dengan dokumen rujukan organisasi.

Perkembangan berikutnya adalah pergeseran dari RAG linear menuju *Agentic AI* dan *multi-agent RAG*, yaitu pendekatan yang membagi tugas ke beberapa agen agar proses retrieval, reasoning, dan generation bisa lebih terstruktur. Dalam studi literatur yang telah dibuat, arah ini terlihat pada studi multi-agent berbasis *LangGraph* yang kamu rangkum di bagian penelitian terdahulu[29], [30]. Penelitian lain bahkan menunjukkan bahwa pendekatan *multi-agent RAG* dapat meningkatkan akurasi hingga 94,3% dibandingkan pendekatan *single-LLM* tradisional[8], sedangkan penelitian lain menegaskan bahwa integrasi *multi-agent AI* dengan RAG meningkatkan reliability dan contextual accuracy pada sistem virtual assistant kesehatan[29]. meski begitu, beberapa temuan juga menyinggung keterbatasan penting dari sistem *agentic*, yaitu adanya behavior gap antara agen AI dan manusia menunjukkan skor yang masih rendah, yaitu *F1-score* 0,464 pada dialog acts dan 0,139 pada penggunaan tool[11]. Ini memperlihatkan bahwa meskipun *multi-agent RAG* lebih kuat secara reasoning, desain orkestrasi tetap harus hati-hati agar sistem tidak terlalu kompleks, tidak kehilangan stabilitas, dan tetap adaptif terhadap konteks pertanyaan pengguna.

Selain dari sisi arsitektur, studi literatur yang telah dibuat juga menunjukkan bahwa penelitian chatbot kesehatan masih banyak yang menilai sistem dari performa teknis, tetapi belum cukup kuat pada evaluasi dari perspektif pengguna. Studi menegaskan bahwa evaluasi chatbot kesehatan perlu lebih dari sekadar akurasi model; sistem juga harus dinilai dari kualitas informasi, manfaat, dan pengalaman pengguna[19], [31]. Karena itu, penggunaan *IS Success Model* dalam penelitian ini menjadi relevan, sebab model ini memungkinkan evaluasi terhadap *System Quality*, *Information Quality*, *Service Quality*, *Use*, *User Satisfaction*, dan *Net Benefits* secara lebih menyeluruh. Dengan pendekatan ini, penelitianmu tidak hanya menguji apakah chatbot PMI Klaten dapat berfungsi, tetapi juga apakah sistem tersebut benar-benar berguna, mudah dipakai, memberi informasi yang berkualitas, dan menghasilkan manfaat nyata bagi pengguna. Hubungan antarartikel terdahulu di sini menjadi jelas *LLM* memberi kemampuan generatif, RAG memberi *grounding*, *agentic AI* memberi orkestrasi dan *reasoning*, sedangkan

IS Success Model memberi kerangka evaluasi yang menilai keberhasilan sistem dari sisi teknis sekaligus sisi pengguna.

Tabel 2. 1 Temuan Penelitian Terdahulu

No	Peneliti	Tahun	Metode	Temuan
1	Farías et al.[23]	2025	LLM Chatbot + Self-RAG	Chatbot meningkatkan kepatuhan pasien terhadap terapi
2	Njei et al.[30]	2026	AI Agents Healthcare Review	AI agents memiliki potensi besar dalam sistem kesehatan
3	Althaf et al.[8]	2025	Multi-Agent RAG	Akurasi sistem mencapai 94.3%
4	Tanasa et al.[29]	2026	Multi-Agent Healthcare Assistant	Meningkatkan <i>reliability</i> dan <i>contextual reasoning</i>
5	Liu et al.[26]	2025	RAG Survey	RAG mengurangi hallucination pada LLM

2.2 Teori yang berkaitan

2.2.1 Retrieval-Augmented Generation (RAG)

RAG merupakan paradigma arsitektur yang menggabungkan kemampuan generatif *Large Language Model (LLM)* dengan mekanisme pengambilan data eksternal (*retrieval*) dari korpus dokumen terverifikasi. Teori ini diusulkan untuk mengatasi keterbatasan *LLM* dalam hal keterbatasan pengetahuan (*cut-off data*) dan kecenderungan fabrikasi informasi[9], [27]. Dalam alur kerja RAG, kueri pengguna tidak langsung diproses oleh model, melainkan digunakan sebagai kunci pencarian untuk menarik potongan teks (*chunks*) yang paling relevan dari basis data vektor .

Proses augmentasi data dalam RAG melibatkan tiga tahap utama: *Indexing* (representasi dokumen ke ruang vektor), *Retrieval* (pencarian kemiripan semantik), dan *Generation* (penyusunan jawaban akhir). Penggunaan teknik ini

menjamin bahwa respons yang dihasilkan chatbot memiliki landasan faktual (*grounded*) pada dokumen referensi, yang secara sistematis meningkatkan skor *faithfulness* sistem[27], [32]. Penelitian terbaru menunjukkan bahwa sistem RAG yang dioptimalkan dengan strategi *reranking* dapat memberikan peningkatan akurasi hingga 10% dibandingkan model bahasa standar[28].

Kebutuhan akan RAG sangat krusial bagi organisasi seperti PMI karena protokol kesehatan dan informasi tanggap darurat bersifat dinamis. Dengan menambahkan jawaban pada basis data internal, risiko kesalahan fatal dalam pemberian saran medis awal atau *triase* mandiri dapat diminimalisir secara signifikan. RAG memungkinkan chatbot bertindak sebagai antarmuka cerdas yang tidak hanya mengolah teks tetapi juga mereferensikan sumber otoritatif secara transparan kepada pengguna.

2.2.2 Agentic AI dan Orkestrasi LangGraph

Agentic AI merepresentasikan pergeseran dari sistem chatbot pasif menuju agen otonom yang mampu melakukan perencanaan (*planning*), evaluasi mandiri (*self-reflection*), dan penggunaan alat (*tool use*) secara dinamis. Berbeda dengan RAG konvensional yang bekerja secara linear satu arah, *Agentic RAG* memungkinkan model untuk memutuskan secara iteratif apakah data yang ditemukan sudah cukup untuk menjawab pertanyaan atau memerlukan pencarian ulang[27], [32]. Pendekatan ini menyerupai pola pikir sistem 2 manusia yang bersifat deliberatif dan analitis.

LangGraph hadir sebagai kerangka kerja orkestrasi yang mendefinisikan alur kerja agen dalam bentuk graf berarah (*Directed Graph*)[10]. Setiap simpul (*node*) dalam graf merepresentasikan unit fungsi tertentu, seperti klasifikasi niat pengguna atau eksekusi kueri *SQL*, sementara tepi (*edge*) mengatur logika transisi antar simpul. Arsitektur *stateful* pada *LangGraph* memastikan bahwa informasi konteks dan riwayat pesan tetap tersimpan dan dapat diakses oleh setiap simpul sepanjang sesi percakapan berlangsung[10].

Penerapan pola desain agen seperti *self-correction loop* pada *LangGraph* terbukti mampu meningkatkan keandalan sistem dalam domain kesehatan.

Agen validator dapat memeriksa apakah jawaban yang dihasilkan mengandung halusinasi dengan membandingkannya kembali terhadap dokumen referensi sebelum dikirimkan ke pengguna[10]. Pola ini mengurangi tingkat kegagalan respons hingga 80% dan memberikan kemampuan triase yang lebih cerdas dibandingkan chatbot berbasis pohon keputusan tradisional.

2.3 Framework/Algoritma yang Digunakan

Penelitian ini menggunakan *DeLone and McLean Information System Success Model* sebagai kerangka evaluasi sistem *chatbot PMI* yang dikembangkan. Model ini merupakan salah satu framework yang paling banyak digunakan dalam penelitian sistem informasi untuk mengevaluasi keberhasilan implementasi sistem informasi. Model ini pertama kali diperkenalkan oleh *DeLone dan McLean* pada tahun 1992 dan kemudian diperbarui pada tahun 2003 dengan menambahkan beberapa variabel penting seperti penggunaan sistem dan manfaat yang dihasilkan oleh sistem informasi. framework ini terdiri dari beberapa dimensi utama yaitu *System Quality*, *Information Quality*, *Use*, *User Satisfaction*, dan *Net Benefits*. *System Quality* mengukur kualitas teknis sistem seperti kecepatan respon, stabilitas sistem, dan kemudahan penggunaan sistem. *Information Quality* mengukur kualitas informasi yang dihasilkan oleh sistem seperti akurasi, relevansi, dan kelengkapan informasi yang diberikan kepada pengguna. Kedua dimensi ini sangat penting dalam sistem chatbot yang memberikan informasi kesehatan karena kualitas sistem dan informasi yang rendah dapat mengurangi kepercayaan pengguna terhadap sistem.

Dimensi *Use* dan *User Satisfaction* menggambarkan bagaimana pengguna berinteraksi dengan sistem serta tingkat kepuasan pengguna setelah menggunakan sistem tersebut. Sistem dengan kualitas yang baik cenderung meningkatkan tingkat penggunaan sistem serta meningkatkan kepuasan pengguna. Dalam konteks chatbot layanan informasi kesehatan, kepuasan pengguna dapat dipengaruhi oleh kemudahan penggunaan sistem serta kemampuan sistem dalam memberikan jawaban yang relevan terhadap pertanyaan pengguna. dimensi terakhir dalam model ini adalah *Net Benefits* yang menggambarkan manfaat yang dihasilkan oleh

sistem bagi pengguna maupun organisasi. Dalam penelitian ini, *Net Benefits* berkaitan dengan manfaat penggunaan *chatbot PMI* dalam meningkatkan akses masyarakat terhadap informasi kesehatan dasar serta membantu masyarakat memahami langkah awal dalam menangani gejala kesehatan ringan.

2.4 Tools/Software yang Digunakan

Pengembangan *chatbot PMI* dalam penelitian ini memanfaatkan berbagai tools dan software yang mendukung implementasi sistem berbasis *Large Language Models* dan *Retrieval-Augmented Generation*. Tools yang digunakan dipilih berdasarkan kemampuannya dalam mendukung pengembangan sistem AI yang modular, scalable, dan mudah diintegrasikan dengan berbagai layanan lainnya. *framework* utama yang digunakan dalam pengembangan chatbot adalah LangChain dan LangGraph[10]. LangChain digunakan untuk mengintegrasikan model bahasa dengan berbagai sumber data seperti database, API, dan dokumen eksternal. Sementara itu, LangGraph digunakan untuk mengelola alur kerja sistem berbasis agent yang memungkinkan koordinasi antara beberapa agen dalam memproses permintaan pengguna. Pendekatan multi-agent ini terbukti mampu meningkatkan efisiensi sistem dalam menangani tugas yang kompleks.

Selain *framework* pengembangan AI, penelitian ini juga menggunakan PostgreSQL dengan ekstensi FAISS sebagai sistem penyimpanan data. PostgreSQL digunakan untuk menyimpan data relasional seperti log percakapan pengguna serta metadata dokumen. Sementara itu, FAISS digunakan untuk menyimpan *embedding* vektor yang dihasilkan oleh model bahasa sehingga memungkinkan sistem melakukan pencarian dokumen berbasis kemiripan semantik. model bahasa yang digunakan dalam penelitian ini dijalankan melalui Ollama yang memungkinkan deployment model bahasa secara lokal tanpa bergantung pada layanan API eksternal[27], [32]. Penggunaan model lokal memberikan keuntungan dalam hal keamanan data serta mengurangi ketergantungan terhadap layanan cloud pihak ketiga.

Tabel 2. 2 Tools yang Digunakan

Tools	Fungsi
LangChain	Integrasi LLM dengan sistem
LangGraph	Orkestrasi multi-agent workflow
PostgreSQL	Database relasional
FIASS	Penyimpanan embedding vektor
Ollama	Menjalankan model LLM secara lokal
Next.js	Antarmuka chatbot

2.5 Pemetaan Gap Analytic

Tabel 2. 3 Pemetaan Gap Analisis

Aspek	Referensi Utama	Gap	Solusi Penelitian
LLM Healthcare Chatbot	[17], [23]	Informasi belum ter-grounding.	Integrasi RAG.
Retrieval-Augmented Generation	[25], [26]	Reasoning masih terbatas.	Agentic RAG.
Multi-Agent AI	[8], [29]	Belum diterapkan pada PMI.	Implementasi pada PMI Klaten.
Evaluasi AI Chatbot	[11], [19]	Evaluasi masih dominan teknis.	IS Success Model.
AI Healthcare Agent	[30]	Implementasi organisasi kemanusiaan masih minim.	Studi kasus PMI Klaten.

Perkembangan chatbot kesehatan modern diawali oleh pemanfaatan *Large Language Models (LLMs)* yang mampu memahami bahasa alami dan menghasilkan respons yang lebih kontekstual. Penelitian menunjukkan bahwa chatbot berbasis *LLM* dapat meningkatkan keterlibatan pengguna dalam program terapi[23]. Di sisi

lain, penelitian menunjukkan bahwa *NLP* berbasis *LLM* mampu membantu penyampaian informasi medis dasar secara lebih interaktif[17]. Penelitian juga menegaskan bahwa *LLM* memiliki potensi besar dalam mendukung diagnosis dan pemahaman awal terhadap kondisi kesehatan[24]. Namun, ketiga penelitian tersebut masih memperlihatkan keterbatasan mendasar, yaitu ketergantungan pada pengetahuan internal model yang tidak selalu terhubung dengan sumber fakta eksternal, sehingga jawaban tetap berisiko kurang akurat ketika menghadapi pertanyaan yang spesifik atau berada di luar data pelatihan.

Keterbatasan *LLM* murni mendorong perkembangan pendekatan *Retrieval-Augmented Generation (RAG)* yang menggabungkan kemampuan generatif model bahasa dengan mekanisme *retrieval* terhadap dokumen eksternal. Penelitian menjadi salah satu fondasi utama yang memperkenalkan konsep *grounded generation* pada tugas-tugas yang membutuhkan pengetahuan intensif[25]. Selanjutnya, penelitian menunjukkan bahwa *RAG* dapat mengurangi hallucination dan meningkatkan *factual consistency* pada jawaban model[26]. Dalam temuan juga mencatat bahwa integrasi *RAG* pada chatbot domain-spesifik dapat meningkatkan akurasi respons hingga 85,55%, sedangkan penelitian memperlihatkan bahwa pendekatan *RAG* lokal dapat diterapkan secara lebih ringan dan fleksibel pada sistem berbasis dokumen[27]. Secara teoritis, temuan ini menegaskan bahwa *retrieval* menjadi komponen kunci untuk menjamin jawaban chatbot tetap relevan, dapat ditelusuri, dan tidak semata-mata bergantung pada prediksi bahasa model.

Perkembangan berikutnya adalah munculnya *Agentic AI* dan *Multi-Agent RAG*, yang berupaya mengatasi keterbatasan *RAG* linear melalui pembagian peran antar agen. Penelitian menunjukkan bahwa orkestrasi beberapa agen dapat meningkatkan *reliability* dan *contextual reasoning* pada sistem asisten kesehatan[29]. Hal serupa juga tampak pada penelitian yang melaporkan akurasi sistem mencapai 94,3% melalui koordinasi agen spesialis[8]. Pada level teoretis, pendekatan ini memberi ruang bagi *retrieval agent*, *reasoning agent*, *validator agent*, dan *response generator* untuk bekerja secara lebih terstruktur dibandingkan pendekatan *single-*

LLM. Namun, penelitian menegaskan bahwa penerapan AI agent di bidang kesehatan masih berkembang dan belum sepenuhnya matang untuk skenario dunia nyata. Ini menunjukkan bahwa *agentic RAG* memang menjanjikan, tetapi masih memerlukan orkestrasi yang ketat agar tidak menambah kompleksitas berlebihan, latensi, atau ketidakstabilan respons[30].

Selain pengembangan arsitektur, hasil SLR juga menunjukkan bahwa evaluasi chatbot kesehatan masih lebih banyak berfokus pada performa teknis dibandingkan pengalaman pengguna. Penelitian menunjukkan bahwa agen *LLM* masih memiliki kesenjangan perilaku dibanding manusia, dengan capaian *F1-score* 0,464 pada *dialog acts* dan 0,139 pada penggunaan *tools*[11]. Angka ini menegaskan bahwa kemampuan *agentic system* belum otomatis selaras dengan ekspektasi interaksi manusia. Di sisi lain, penelitian menyoroti perlunya evaluasi chatbot kesehatan yang tidak hanya menilai performa model, tetapi juga kualitas sistem, kualitas informasi, dan manfaat pengguna[19]. Penelitian juga memperlihatkan bahwa *retrieval* pada konteks klinis membutuhkan evaluasi yang lebih terstruktur dan realistis[31]. Secara teoritis, rangkaian studi ini mendukung penggunaan *IS Success Model* karena model tersebut memungkinkan penilaian sistem dari sisi *System Quality*, *Information Quality*, *Service Quality*, *Use*, *User Satisfaction*, dan *Net Benefits*, bukan hanya dari sisi akurasi teknis semata.

Berdasarkan keseluruhan literatur, masih terdapat tiga gap utama yang relevan dengan penelitian ini. Pertama, sebagian besar chatbot kesehatan masih berhenti pada *LLM* atau *RAG* linear, sementara pendekatan *Agentic RAG* dengan orkestrasi berbasis *LangGraph* masih relatif terbatas. Kedua, evaluasi sistem masih dominan menggunakan metrik teknis seperti akurasi *retrieval* atau kualitas generasi, padahal keberhasilan sistem informasi seharusnya juga diukur dari perspektif pengguna. Ketiga, implementasi chatbot berbasis *LLM*, *RAG*, maupun *multi-agent* masih sangat jarang diterapkan pada konteks organisasi kemanusiaan seperti Palang Merah Indonesia (PMI) yang memiliki kebutuhan informasi cepat, konsisten, dan berbasis dokumen resmi. Oleh karena itu, penelitian ini memposisikan diri sebagai pengembangan chatbot PMI Klaten berbasis *Agentic Retrieval-Augmented*

Generation yang dievaluasi menggunakan *IS Success Model*, sehingga kontribusinya tidak hanya pada aspek teknis pengembangan sistem, tetapi juga pada penilaian keberhasilan sistem informasi secara lebih komprehensif.



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA