BAB II
TINJAUAN PUSTAKA

Chatbot adalah salah satu inovasi teknologi yang banyak digunakan dalam berbagai sektor seperti layanan kesehatan, pendidikan, perbankan, hingga e-*commerce*. Dengan kemajuan dalam pemrosesan bahasa alami (*Natural Language Processing*/NLP) dan model bahasa besar (*Large Language Models*/LLM), *chatbot* kini mampu memahami konteks secara lebih dalam dan memberikan respons yang lebih manusiawi. (tiap penelitian tambah metode testing)

Salah satu pendekatan yang banyak digunakan adalah *Retrieval-Augmented Generation* (RAG), yang menggabungkan pengambilan informasi eksternal dengan kemampuan generatif model LLM [4]. Berdasarkan survei komprehensif oleh Gao et al. (2023), efektivitas sistem RAG umumnya diuji menggunakan metrik seperti *Exact Match (EM)*, *F1-score*, *retrieval accuracy*, evaluasi *faithfulness* terhadap dokumen sumber, serta kombinasi evaluasi otomatis dan *human judgment* untuk menilai konsistensi faktual dan *grounding* terhadap basis pengetahuan eksternal.

Selain itu, *framework* LangGraph mulai digunakan secara luas untuk mengatur alur percakapan yang kompleks melalui representasi eksplisit berbasis graf, mendukung alur *multi-agent* dan pemanggilan *tool* secara modular. Dalam bagian ini, penulis akan mengulas beberapa penelitian terkait penggunaan *chatbot* RAG dan LangGraph, baik dalam konteks medis maupun non-medis.

Penelitian oleh Montes-Torres, dkk. (2025) mengembangkan *chatbot* berbasis LLM untuk mengekstraksi informasi dari *electronic health records* (EHR) berbahasa Spanyol, khususnya untuk pasien onkologi. Studi ini menguji tujuh model berbasis Ollama serta GPT-4.5 menggunakan pendekatan *in-context learning* terhadap 100 rekam medis klinis. Metode testing dilakukan dengan membandingkan output JSON model terhadap *gold standard* yang disusun oleh spesialis kanker payudara, menggunakan metrik *Precision*, *Recall*, dan *F1-score*. Hasil menunjukkan GPT-4.5 memperoleh *F1-score* tertinggi (0,88), sementara model *open-source* menunjukkan performa lebih rendah namun dapat ditingkatkan

melalui *post-processing* teradaptasi. Artinya *chatbot* dapat menyederhanakan data kompleks dan mempercepat proses pengambilan keputusan medis, memperlihatkan potensi integrasi LLM dalam sistem informasi rumah sakit [5].

Sementara itu, Bahaadinbeigy (2025) melakukan tinjauan sistematis terkait penggunaan *chatbot* dalam kesehatan ibu (*maternal health*). *Chatbot* yang dikaji difungsikan untuk memberikan edukasi kehamilan, pemantauan gizi, serta deteksi risiko dini [6]. Metode testing dalam studi ini bersifat metodologis, yaitu melalui evaluasi kualitas studi yang disertakan, analisis desain penelitian, serta sintesis hasil terkait efektivitas *chatbot* terhadap edukasi kehamilan, pemantauan gizi, dan deteksi risiko dini. *Outcome* yang dianalisis mencakup peningkatan pengetahuan, perubahan perilaku kesehatan, tingkat keterlibatan pengguna, dan kepuasan. Dengan demikian, pengujian dilakukan melalui evaluasi komparatif antar studi dan analisis kualitas metodologis.

Dalam konteks edukasi kesehatan, Fetrati, dkk. (2025) mengembangkan sistem *multi-chatbot* untuk kampanye edukasi penyakit menular seksual (STI). Sistem ini menggabungkan model generatif dan *rule-based* untuk memberikan informasi yang persuasif dan kontekstual. Metode *testing* dilakukan melalui desain eksperimental berbasis pengguna aplikasi *mobile*, menggunakan pendekatan *between-subject design*. Evaluasi dilakukan dengan *pre-test* dan *post-test* untuk mengukur *knowledge gain*, pengukuran *engagement metrics* (durasi interaksi dan frekuensi penggunaan), serta survei persepsi berbasis skala Likert. Evaluasi menunjukkan adanya peningkatan pemahaman dan keterlibatan pengguna, serta efektivitas pendekatan *multi-chatbot* dalam menyampaikan pesan kesehatan secara adaptif [7].

Pendekatan RAG dalam *chatbot* juga banyak dieksplorasi di luar bidang medis. Meyer, dkk. (2023) meneliti penerapan *chatbot* RAG dalam layanan pelanggan [8]. Evaluasi dilakukan melalui pengukuran *user satisfaction score*, *task completion rate*, relevansi jawaban, serta waktu penyelesaian masalah (*resolution time*). Analisis statistik menunjukkan bahwa integrasi retrieval meningkatkan kualitas respons kontekstual dan kepuasan pengguna dibanding sistem statis.

Dalam bidang pendidikan, Qiu, dkk. (2025) melakukan studi sistematis

tentang penggunaan *chatbot* dalam proses belajar, termasuk penerapan teknik *Retrieval-Augmented Generation* (RAG) [9]. Metode testing dilakukan melalui analisis sistematis terhadap artikel terindeks dengan mengkaji indikator *learning gain, student engagement,* dan *user perception*. Beberapa studi dalam tinjauan tersebut menggunakan data log interaksi sistem, skor akademik sebelum dan sesudah penggunaan *chatbot*, serta survei persepsi siswa. Sintesis hasil menunjukkan bahwa chatbot berbasis RAG meningkatkan relevansi informasi dan partisipasi aktif siswa dalam pembelajaran.

Selain RAG, LangGraph kini mulai banyak digunakan untuk membangun *chatbot* dengan alur logika kompleks dan modular. Okimoto, dkk. (2025) mengulas kerangka kerja seperti LangGraph dalam pengembangan *chatbot multi-agent* yang mampu mengatur orkestrasi alur kerja dan pemanggilan *tool* secara fleksibel [10]. Metode evaluasi dalam studi ini berupa analisis komparatif arsitektural, membandingkan fleksibilitas orkestrasi *tool*, modularitas sistem, dan efisiensi *workflow* antar pendekatan.

Pérez (2025) merancang *chatbot* cerdas untuk mendukung peneliti genetika dalam proses rekayasa genetic [11]. Metode *testing* dilakukan melalui *task-based evaluation***,** yaitu pengujian kemampuan sistem dalam menyelesaikan tugas teknis seperti perhitungan genom, pemanggilan kalkulator eksternal, dan eksekusi alur logika multi-step. Evaluasi mencakup pengujian ketepatan hasil (*task correctness*), reliabilitas *workflow*, serta validasi output oleh pakar domain. Selain itu, dilakukan pengujian terhadap *robustness* sistem dalam menangani pertanyaan kompleks dan multi-langkah.

Secara keseluruhan, tinjauan pustaka menunjukkan bahwa pendekatan *Retrieval-Augmented Generation* (RAG) dan arsitektur modular seperti LangGraph memainkan peran penting dalam pengembangan *chatbot* modern. RAG memperkuat kemampuan *chatbot* dalam memberikan jawaban yang relevan berdasarkan sumber eksternal, sementara LangGraph mendukung orkestrasi percakapan kompleks yang bersifat dinamis dan *multi-agent*. Dalam konteks layanan kesehatan seperti di Siloam Hospitals, penggabungan kedua pendekatan ini memberikan peluang besar dalam membangun *chatbot* yang tidak hanya informatif,

tetapi juga adaptif terhadap data *real-time* (API), preferensi pengguna (*session/user*), serta integrasi fitur kompleks seperti pencarian dokter, *booking*, dan FAQ layanan. Oleh karena itu, pengembangan Siloam Virtual Assistant (Silvia) dengan kombinasi RAG, LangGraph, dan LLM merupakan langkah strategis untuk mewujudkan sistem layanan digital kesehatan yang cerdas dan responsif.

Berdasarkan tabel perbandingan 2.1, dapat disimpulkan bahwa masing-masing studi memiliki fokus dan kontribusi spesifik, namun belum ada yang mengintegrasikan secara utuh pendekatan *Retrieval-Augmented Generation* (RAG), arsitektur percakapan modular dengan LangGraph, serta penerapan *Large Language Model* (LLM) dalam konteks layanan kesehatan berbasis sistem nyata. Penelitian seperti yang dilakukan oleh Montes-Torres, dkk. (2025) dan ShojaeiBaghini & Bahaadinbeigy (2025) berfokus pada domain medis, namun terbatas pada ekstraksi data atau edukasi, tanpa dukungan interaksi dinamis maupun integrasi sistem operasional.

Sementara itu, studi lain seperti oleh Meyer, dkk. (2023) dan Qiu, dkk. (2025) menampilkan keunggulan RAG dalam layanan pelanggan dan pendidikan, tetapi tidak melibatkan data *real-time*, *vector database*, atau personalisasi berbasis sesi pengguna. Skripsi ini menjadi satu-satunya yang menggabungkan penggunaan LangGraph untuk mengatur alur percakapan kompleks, pemanfaatan LLM OpenAI untuk pemahaman dan respons adaptif, integrasi dengan API waktu nyata, serta pengelolaan data pengguna dalam basis data sementara. Gabungan tersebut diimplementasikan dalam sistem *chatbot* layanan rumah sakit (Siloam Hospitals), menjadikan skripsi ini mengisi celah penelitian yang belum dijelajahi sebelumnya sekaligus memberikan kontribusi orisinal dalam pengembangan *chatbot* cerdas di sektor layanan kesehatan Indonesia.

Tabel 2. 1. Perbandingan Penelitian

No	Peneliti	Domain	Teknologi Utama	Fitur *Chatbot*	Data Input	Evaluasi Pengguna	Kelebihan	Gap / Kekurangan
1	Montes-Torres, dkk. (2025) [5]	Medis (Onkologi)	LLM (*data extraction*)	Ekstraksi data EHR	*Unstructured* EHR (Spanish)	Perbandingan output model dengan *gold standard* ahli; metrik Precision, Recall, F1-score pada 100 rekam medis	Akurasi tinggi dalam ekstraksi klinis	Tidak menyentuh interaksi pasien, tidak ada LangGraph atau fitur percakapan
2	ShojaeiBaghini & Bahaadinbeigy (2025) [6]	Medis (Kehamilan)	AI *chatbot (rule-based)*	Edukasi kehamilan, pemantauan gizi	Data edukatif, reminder	Protokol PRISMA; evaluasi kualitas studi; analisis outcome (knowledge gain, engagement, usability)	Berguna untuk daerah minim fasilitas	Tidak mendukung fitur kompleks (*real-time*)
3	Fetrati, dkk. (2025) [7]	Edukasi Kesehatan (STI)	*Multi-chatbot (rule-based* + generatif)	Edukasi interaktif STI	Informasi kesehatan	Desain eksperimen *pre-test/post-test*; pengukuran knowledge gain, engagement metrics, survei Likert	Gabungkan 2 model (*generatif* + *rule*)	Tidak cocok untuk domain layanan kesehatan nyata, tidak ada *retrieval*
4	Meyer, dkk. (2025) [8]	*Customer Service*	RAG + *Deep Learning*	Jawaban kontekstual untuk pelanggan	Informasi statis & dokumen	A/B testing; user satisfaction score, task completion rate, response relevance, resolution time	RAG meningkatkan kepuasan pengguna	Tidak diimplementasikan di sistem dengan API/data dinamis
5	Qiu, dkk. (2025) [9]	Pendidikan	*Chatbot* berbasis RAG	Mendukung proses belajar	Materi belajar + pertanyaan	Systematic review; analisis learning gain, engagement, perception; data log	Peningkatan *engagement* siswa	Tidak menyentuh domain medis atau fitur sistem layanan

						interaksi dan skor akademik		
6	Okimoto, dkk. (2025) [10]	Teknikal (LLM *survey*)	LangGraph + *Multi-agent* + RAG	Orkestrasi agen (*survey*)	Informasi eksterior multi-model	Analisis komparatif arsitektur; evaluasi modularitas, workflow efficiency, dan orkestrasi tool	*Framework* survey mendalam	Tidak berfokus pada aplikasi konkret, tidak spesifik pada *chatbot* operasional
7	Aragón Pérez (2025) [11]	Genetika / Ilmiah	LangGraph + GPT-4o	Jawaban teknis + *tool calling*	*Natural language query + tool*	Task-based evaluation; uji ketepatan tugas teknis, validasi pakar domain, dan uji robustnes workflow	Mendukung peneliti teknis	Tidak cocok untuk layanan publik, tidak *real-time*
8	Penelitian Penulis (2025)	Medis (RS Siloam)	RAG + LangGraph + LLM (OpenAI)	Pencarian dokter, *booking*, *reschedule*, FAQ, rekomendasi	Vector DB + *real-time* API + *session* DB	Survey	Gabungan sistem nyata + *retrieval* + personalisasi	Belum ada penelitian lain yang menggabungkan LangGraph, RAG, dan *real-time service* dalam konteks rumah sakit Indonesia