

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kemunculan *Large Language Models* (LLM) telah membawa pergeseran besar dalam perkembangan *Natural Language Processing* (NLP). Perkembangan ini ditandai dengan hadirnya model bahasa yang tidak lagi bergantung pada pencocokan kata kunci, tetapi mampu memahami konteks bahasa, hubungan antarkalimat, serta menghasilkan respons yang lebih alami berdasarkan pola yang dipelajari dari data berskala besar [1]. Kemampuan tersebut mendorong pemanfaatan LLM pada berbagai aplikasi, seperti chatbot, asisten virtual, hingga sistem rekomendasi berbasis percakapan [2]. Meskipun demikian, penerapan LLM pada sistem yang membutuhkan informasi faktual masih menghadapi tantangan berupa *hallucination*. *Hallucination* merupakan kondisi ketika model menghasilkan informasi yang tidak didukung oleh sumber pengetahuan yang valid, tetapi disampaikan seolah-olah benar sehingga berpotensi menyesatkan pengguna. Permasalahan ini menjadi perhatian karena respons yang tampak meyakinkan dapat mengurangi kepercayaan pengguna terhadap sistem apabila informasi yang diberikan tidak akurat [3].

Sebagai solusi dari permasalahan tersebut, pendekatan *Retrieval-Augmented Generation* (RAG) kini banyak diadopsi untuk membangun sistem cerdas berbasis pengetahuan. Berbeda dengan LLM konvensional yang hanya mengandalkan pengetahuan hasil pelatihan model, RAG menambahkan proses *retrieval* untuk mengambil informasi dari basis pengetahuan eksternal sebelum jawaban dihasilkan. Dokumen yang ditemukan digunakan sebagai konteks bagi LLM dalam menyusun respons sehingga informasi yang dihasilkan menjadi relevan dan memiliki dasar yang jelas. Pendekatan ini juga lebih fleksibel karena basis pengetahuan dapat diperbarui tanpa perlu melakukan *fine-tuning* ulang terhadap model bahasa [4]. Dalam implementasinya, RAG umumnya didukung oleh *semantic retrieval* berbasis *vector embedding* untuk menemukan dokumen yang memiliki kemiripan makna dengan pertanyaan pengguna, bukan hanya kesamaan

kata [5], [6]. Selain itu, pada sistem rekomendasi modern, integrasi LLM dan RAG memungkinkan sistem tidak hanya menghasilkan daftar rekomendasi, tetapi juga memberikan penjelasan mengenai alasan pemilihan rekomendasi sehingga meningkatkan transparansi kepada pengguna [7]–[10].

Meskipun RAG mampu meningkatkan kualitas informasi yang diberikan LLM, proses *retrieval* saja belum sepenuhnya menjamin bahwa seluruh respons akan tetap berada pada ruang lingkup yang diinginkan. Oleh karena itu, pengembangan sistem berbasis LLM juga memerlukan mekanisme pengendalian (*guardrails*) agar respons tetap konsisten dan sesuai dengan domain yang telah ditentukan. *Guardrails* berfungsi melakukan validasi masukan, membatasi domain pembahasan, menyaring keluaran, serta menyediakan *fallback response* ketika informasi yang dibutuhkan tidak tersedia [11]. Kombinasi antara RAG dan *guardrails* banyak diterapkan pada domain yang membutuhkan tingkat akurasi informasi tinggi karena mampu mengurangi potensi *hallucination* sekaligus menjaga konsistensi respons sistem [12].

Salah satu domain yang membutuhkan informasi akurat adalah rekomendasi produk *skincare*. Pada domain ini, proses pemilihan produk dipengaruhi oleh berbagai faktor, seperti jenis kulit, kondisi kulit, kandungan bahan aktif, serta kecocokan bahan terhadap kebutuhan pengguna [13], [14]. Oleh karena itu, rekomendasi yang tidak didukung oleh informasi yang tepat berpotensi menyebabkan pemilihan produk yang kurang sesuai [15]. Selain itu, pengguna di Indonesia umumnya menyampaikan keluhan menggunakan bahasa sehari-hari yang bersifat naratif dan informal sehingga sistem perlu memahami makna semantik dari pertanyaan pengguna, bukan sekadar melakukan pencocokan kata kunci [16].

Berbagai penelitian telah menunjukkan bahwa penerapan *Retrieval-Augmented Generation* (RAG) mampu meningkatkan kualitas respons LLM dengan memanfaatkan dokumen hasil *retrieval* sebagai sumber informasi [12]. Di sisi lain, penelitian pada domain rekomendasi *skincare* masih banyak menggunakan pendekatan *content-based filtering* maupun aturan statis sehingga memiliki keterbatasan dalam memahami pertanyaan pengguna yang disampaikan

dalam bahasa alami [17]. Selain itu, implementasi yang mengintegrasikan RAG, *semantic retrieval* berbasis *vector embedding*, *guardrails*, serta pemrosesan bahasa Indonesia informal pada domain rekomendasi *skincare* masih relatif terbatas [11], [16]. Oleh karena itu, penelitian ini mengimplementasikan sistem rekomendasi *skincare* berbasis *Retrieval-Augmented Generation* (RAG) yang mengintegrasikan *semantic retrieval* menggunakan FAISS, *Large Language Model* (LLM), dan mekanisme *guardrails* untuk menghasilkan rekomendasi yang lebih relevan serta mengurangi potensi *hallucination*. Pengembangan sistem dilakukan menggunakan metodologi CRISP-DM sebagai kerangka kerja implementasi [18].

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang yang telah dijelaskan, maka rumusan masalah dalam penelitian ini adalah:

1. Bagaimana mengimplementasikan sistem rekomendasi *skincare* berbasis *Retrieval-Augmented Generation* (RAG) dengan integrasi mekanisme *guardrails* untuk memahami keluhan pengguna berbahasa Indonesia informal dan istilah *skincare* umum?
2. Bagaimana kinerja sistem rekomendasi RAG dalam menghasilkan rekomendasi produk yang relevan jika diukur menggunakan metrik *Precision*, *Recall*, *F1-Score*, dan *MRR*, serta bagaimana kualitas respons sistem berdasarkan validasi ahli dan evaluasi pengguna (*user study*)?

1.3 Batasan Masalah

Agar penelitian terarah dan fokus, batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Dataset terdiri dari 400 produk *skincare* populer dari platform Female Daily dengan empat kategori utama, yaitu *facial wash*, *toner*, *serum*, dan *moisturizer* (masing-masing 100 produk). Setiap produk direpresentasikan melalui tiga ulasan pengguna (total 1.200 ulasan). Basis pengetahuan dibentuk melalui proses serialisasi yang menggabungkan nama produk, merk, kategori, deskripsi, ulasan, harga, serta tautan produk ke dalam unit informasi.
2. Sebagai pengetahuan pendukung terkait alasan pemilihan produk, digunakan

tiga referensi dermatologi dari dokumen PDF yang telah diekstraksi dan dipotong (*chunking*) menjadi teks terstruktur dalam format CSV.

3. Input sistem berupa teks keluhan pengguna berbahasa Indonesia, termasuk bahasa informal dan tidak memproses input berupa gambar kondisi kulit.
4. Pencarian informasi dokumen menggunakan *semantic retrieval* berbasis vektor melalui FAISS dan *Sentence-Transformers (all-MiniLM-L6-v2)*, yang dikombinasikan dengan *rule-based guardrails* untuk menangani batasan spesifik seperti budget dan kategori produk.
5. Model generatif yang diuji mencakup lima varian LLM berbasis teks (*text-to-text*). yaitu *LlaMA 3.1 8B*, *LlaMA 3.3 70B*, *OpenAI GPT-OSS 20B*, *OpenAI GPT-OSS 120B*, dan *Qwen 3 32B* yang diakses melalui API Groq.
6. Implementasi akhir pada sistem prototipe menggunakan satu model terbaik berdasarkan hasil pengujian awal. Kriteria pemilihan didasarkan pada tiga aspek utama, yaitu kecepatan respons (*inference latency*), kepatuhan terhadap instruksi (*instruction following*), serta konsistensi jawaban terhadap konteks (*hallucination mitigation*).
7. Metode utama yang digunakan adalah *Retrieval-Augmented Generation (RAG)*. Penelitian ini tidak menggunakan metode *Collaborative Filtering* atau data riwayat personal pengguna lainnya.
8. Evaluasi dilakukan secara bertingkat, yaitu evaluasi kinerja *retrieval* menggunakan metrik *Precision*, *Recall*, *F1-Score*, dan *MRR* serta penilaian kualitas respons generatif (*relevansi*, *explainability*, *usability*, *context awareness*, dan kepuasan pengguna) lewat *user study*.
9. Sistem dikembangkan dalam bentuk prototipe web menggunakan Streamlit, tanpa integrasi *e-commerce*, penyimpanan riwayat percakapan (*history log*), maupun aplikasi *mobile*.
10. Output sistem menyajikan rekomendasi produk terstruktur yang mencakup harga, kandungan utama, manfaat, serta tautan produk. Sistem ini murni berfungsi sebagai alat bantu rekomendasi dan bukan untuk diagnosis medis profesional.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

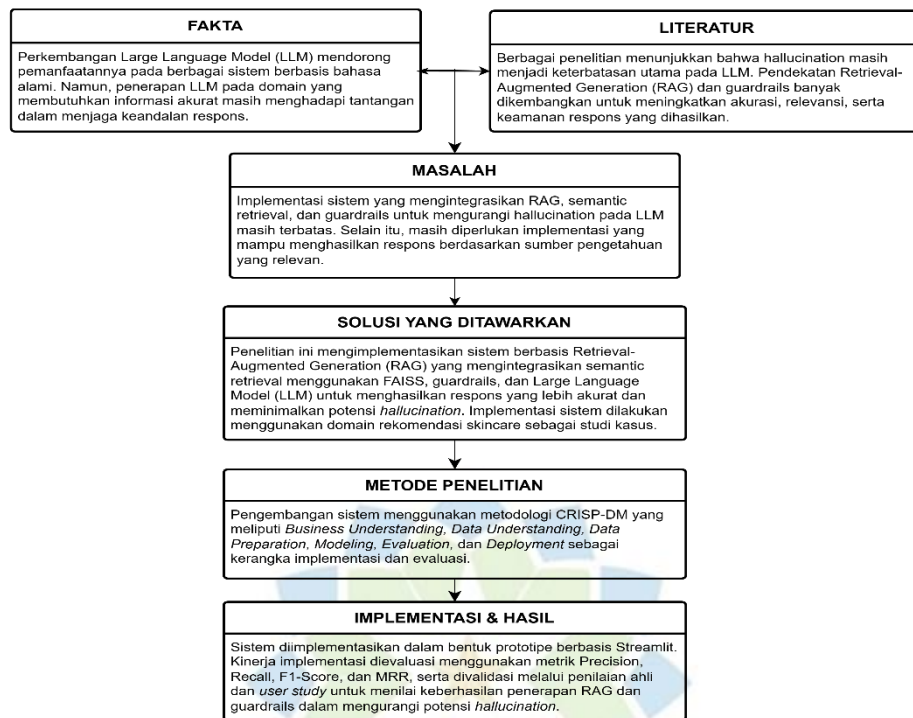
1. Mengimplementasikan sistem rekomendasi produk *skincare* berbasis *Retrieval-Augmented Generation* (RAG) yang terintegrasi dengan mekanisme *guardrails* untuk memahami keluhan pengguna berbahasa Indonesia informal dan istilah *skincare* umum.
2. Mengevaluasi kinerja sistem dalam menghasilkan rekomendasi produk yang relevan menggunakan metrik *Precision*, *Recall*, *F1-Score*, *Mean Reciprocal Rank* (MRR), serta menilai kualitas respons generatif melalui validasi ahli dan evaluasi pengguna (*user study*).

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Memberikan kontribusi terhadap pengembangan metode *Retrieval-Augmented Generation* (RAG) pada pemrosesan bahasa alami (*Natural Language Processing* (NLP) bahasa Indonesia, khususnya dalam memahami keluhan pengguna yang disampaikan secara informal serta pencarian informasi berbasis *semantic retrieval*. Penelitian ini juga dapat menambah kajian penerapan AI generatif pada domain *skincare*.
2. Menjadi referensi bagi penelitian selanjutnya dalam pengembangan sistem rekomendasi berbasis *Artificial Intelligence* (AI) yang mengintegrasikan model generatif, *semantic retrieval*, validasi berbasis aturan, dan referensi dermatologi dalam satu sistem.
3. Membantu pengguna memilih produk *skincare* yang lebih sesuai dengan kondisi kulit dan kebutuhan mereka melalui rekomendasi yang disertai penjelasan informatif serta didukung oleh referensi yang relevan.
4. Menyediakan model solusi bagi platform digital *skincare* untuk menghadirkan layanan rekomendasi otomatis yang lebih informatif dan kontekstual, sekaligus membantu meminimalkan risiko *hallucination* pada sistem AI generatif.

1.6 Kerangka Penelitian



Gambar 1.1 Kerangka Penelitian

Berdasarkan Gambar 1.1 alur penelitian dimulai dari fakta mengenai perkembangan *Large Language Model* (LLM) dan meningkatnya pemanfaatannya pada berbagai aplikasi berbasis bahasa alami. Berdasarkan kajian literatur, diketahui bahwa LLM masih memiliki keterbatasan berupa *hallucination*, sehingga pendekatan *Retrieval-Augmented Generation* (RAG) dan *guardrails* banyak dikembangkan untuk meningkatkan akurasi dan kejelasan respons. Berdasarkan fakta dan literatur dirumuskan permasalahan penelitian, yaitu masih terbatasnya implementasi sistem yang mengintegrasikan RAG, *semantic retrieval*, dan *guardrails* untuk mengurangi *hallucination* pada domain rekomendasi skincare. Sebagai solusi, penelitian ini mengimplementasikan sistem rekomendasi berbasis RAG yang memanfaatkan *semantic retrieval* menggunakan FAISS, LLM, dan mekanisme *guardrails*. Proses pengembangan sistem menggunakan metode CRISP-DM, kemudian diimplementasikan dalam prototipe Streamlit dan dievaluasi menggunakan metrik *Precision*, *Recall*, *F1-Score*, dan *MRR*, serta divalidasi melalui penilaian ahli dan *user study*.