

BAB I PENDAHULUAN

I.1 Latar Belakang

Dalam era globalisasi dan perkembangan teknologi informasi, metode bisnis mengalami perubahan yang signifikan. Salah satu inovasi dari perkembangan teknologi dalam dunia bisnis adalah *e-commerce*. Sektor *e-commerce* telah menjadi salah satu pendorong utama dalam perekonomian dunia. Berdasarkan data yang dipublikasikan oleh Badan Pusat Statistik (BPS) pada tahun 2023, pertumbuhan volume transaksi *online* telah meningkat pesat setiap tahunnya, terutama didorong oleh perubahan perilaku konsumen yang semakin terbiasa berbelanja secara digital. Hal ini mencerminkan semakin meningkatnya preferensi konsumen terhadap belanja daring melalui *e-commerce*. Proyeksi pendapatan *e-commerce* diperkirakan akan mencapai US\$4,18 triliun pada tahun 2024 (Roumeliotis dkk., 2024). Data-data tersebut menunjukkan betapa besarnya pengaruh *e-commerce* dalam transformasi dunia bisnis, serta dampaknya yang semakin kuat terhadap ekonomi global di masa depan (Roumeliotis dkk., 2024).

Semua *platform e-commerce* menyajikan hampir jutaan produk yang dilengkapi dengan detail terkait yang memberikan informasi berharga bagi konsumen (Srinivas dkk., 2024). Kekurangan dan kelengkapan informasi produk yang ada di *e-commerce* sangat penting, di mana kepercayaan konsumen bergantung pada transparansi informasi tersebut (Srinivas dkk., 2024). Deskripsi produk yang tidak memadai dan kurangnya informasi dapat meragukan kredibilitas perusahaan dan mengurangi kemungkinan terjadinya transaksi pembelian yang sukses (Srinivas dkk., 2024).

Seiring bertambahnya jumlah produk yang ditawarkan di platform *e-commerce*, kompleksitas pengelolaan data produk juga meningkat secara signifikan. Setiap produk biasanya memiliki atribut yang beragam seperti nama, deskripsi, spesifikasi teknis, harga, stok, ulasan pengguna, dan atribut pendukung lainnya. Kondisi ini menciptakan tantangan besar bagi penyedia layanan dalam memastikan bahwa informasi yang disampaikan kepada pelanggan tetap akurat, konsisten, dan selalu diperbarui sesuai perubahan data terbaru. Permasalahan ini semakin kompleks karena deskripsi produk sering kali terlalu panjang, beragam formatnya, serta tidak

selalu menggunakan istilah yang sama dengan istilah yang digunakan pelanggan. Di sisi lain, konsumen *modern* menuntut informasi produk yang tidak hanya lengkap, tetapi juga disajikan secara ringkas dan mudah dipahami agar dapat mendukung pengambilan keputusan pembelian yang cepat. Permintaan konsumen terhadap informasi yang detail, akurat, dan mudah diakses terus meningkat seiring pesatnya pertumbuhan jumlah produk di platform *e-commerce*. Kondisi ini mendorong sistem layanan informasi untuk dapat menangani volume data yang besar sekaligus memahami kebutuhan pengguna yang beragam. Pendekatan tradisional yang hanya mengandalkan pencarian berbasis kata kunci semakin sulit memenuhi harapan tersebut karena terbatas dalam menangkap konteks dan makna dari kebutuhan ataupun permintaan pengguna. Akibatnya, pelanggan sering kali masih kesulitan menemukan informasi produk yang benar-benar sesuai dengan kebutuhan mereka, yang pada akhirnya dapat memengaruhi kepuasan dan kepercayaan mereka terhadap platform *e-commerce*.

Penting untuk menekankan bahwa kepuasan pelanggan tetap menjadi fokus utama dalam setiap aspek yang terkait dengan *e-commerce* (Roumeliotis dkk., 2024). Namun, untuk memahami perasaan dan tingkat kepuasan pelanggan, diperlukan sumber daya manusia yang besar serta analisis data yang mendalam dan terus-menerus (Roumeliotis dkk., 2024). Untuk meningkatkan tingkat kepercayaan dan keyakinan pelanggan, bisnis *online* akan cenderung untuk menggunakan sistem *Electronic Customer Relationship Management* (e-CRM) agar dapat memberikan kesan yang positif kepada pelanggan (Tan, 2023). E-CRM adalah strategi yang menggunakan teknologi digital untuk mengelola dan memperkuat hubungan antara perusahaan dan pelanggan, dengan tujuan meningkatkan loyalitas dan kepuasan pelanggan melalui interaksi yang lebih personal dan relevan (Nicolas dkk., 2024).

Salah satu alat komunikasi e-CRM yang populer adalah *live chat*, yang memungkinkan pelanggan untuk berinteraksi secara langsung dengan *customer service* (Tan, 2023). Namun, sistem *live chat* memiliki kekurangan karena memerlukan kehadiran dari seseorang untuk mengelola komunikasi tersebut (Tan, 2023). Selain itu, *live chat* tidak tersedia 24 jam, yang menyebabkan ketidaknyamanan bagi pelanggan yang membutuhkan informasi atau umpan balik secara cepat (Tan, 2023). Di sisi lain, sistem *live chat* juga menyebabkan bisnis

mengalami peningkatan biaya karena memerlukan lebih banyak karyawan untuk menangani masalah pelanggan dengan cepat dan *responsive* (Tan, 2023). Permasalahan ini membuat *live chat* kurang ideal untuk memenuhi kebutuhan pelanggan di era digital yang serba cepat, di mana pelanggan mengharapkan respons instan dan layanan yang tidak terbatas dengan waktu.

Sebagai alternatif, banyak bisnis kini beralih menggunakan *chatbot* berbasis AI untuk mengatasi keterbatasan teknologi *live chat*. Penggunaan *chatbot* memungkinkan otomatisasi komunikasi dengan pelanggan, sehingga meningkatkan kemampuan layanan pelanggan dan menghilangkan kebutuhan untuk melibatkan karyawan dalam merespons pertanyaan pelanggan atau mengarahkan permintaan mereka (Angelov & Lazarova, 2019). *Chatbot* awalnya dibuat hanya untuk menangani tugas sederhana melalui komunikasi berbasis teks (Yun & Park, 2022). Namun, seiring perkembangan teknologi, *chatbot* kini mampu menjalankan tugas yang lebih kompleks, seperti memberikan rekomendasi produk, mengatur pemesanan awal, dan melaksanakan layanan berbasis lokasi. Hal ini memberikan kemudahan, aksesibilitas, serta penghematan biaya bagi pengguna (Yun & Park, 2022). *Chatbot* dapat merespons pertanyaan, memberikan informasi, dan mengarahkan permintaan pelanggan secara cepat dan akurat, tanpa keterlibatan langsung karyawan, sehingga meningkatkan pengalaman pelanggan sekaligus efisiensi operasional (Angelov & Lazarova, 2019).

Meskipun *chatbot* telah terbukti membantu meningkatkan kecepatan layanan dan mengurangi beban operasional, tantangan baru tetap muncul seiring meningkatnya kompleksitas kebutuhan terhadap informasi. Pelanggan tidak lagi hanya menanyakan pertanyaan sederhana, tetapi juga mengharapkan jawaban yang lebih mendalam, kontekstual, dan berbasis data terbaru. Dalam praktiknya, *chatbot* tradisional yang hanya mengandalkan *rule-based* atau *template* jawaban sering kali tidak cukup responsif dalam menangani pertanyaan kompleks yang memerlukan pemahaman semantik atau penarikan informasi dari basis data produk yang besar dan dinamis. Tantangan inilah yang mendorong perlunya pendekatan lebih *modern*, agar *chatbot* mampu menghasilkan jawaban yang relevan, faktual, dan sesuai konteks, bahkan ketika berhadapan dengan pertanyaan yang lebih variatif.

Untuk menjawab tantangan tersebut, muncul kebutuhan akan teknologi yang lebih *modern* dan adaptif. Salah satu pendekatan yang banyak digunakan adalah penerapan *chatbot* berbasis *Large Language Model* (LLM). LLM merupakan model bahasa berskala besar yang dilatih untuk memahami konteks, menangkap makna semantik, dan menghasilkan teks secara lebih natural dan koheren. LLM menawarkan potensi untuk menggantikan keterbatasan *chatbot* tradisional, karena mampu merespons pertanyaan yang bervariasi dengan gaya bahasa yang lebih alami, sekaligus memproses konteks percakapan secara lebih menyeluruh. Pada penelitian ini akan digunakan *pretrained* LLM karena model yang telah dilatih sebelumnya memiliki kemampuan untuk memahami dan menghasilkan teks dengan lebih efisien dibandingkan memulai pelatihan dari awal. Dengan memanfaatkan *pretrained* LLM, penelitian dapat menghemat sumber daya komputasi dan waktu, karena model ini memiliki pengetahuan awal yang relevan untuk tugas NLP.

Model *pretrained* LLM yang dipilih adalah LLaMA (*Large Language Model Meta AI*), sebuah model bahasa berskala besar yang dikembangkan oleh Meta dan dirancang untuk efisiensi serta performa tinggi pada berbagai tugas NLP. LLaMA dilatih menggunakan dataset berskala besar yang mencakup data publik dari berbagai sumber dengan jumlah parameter yang lebih kecil dibandingkan model lain sekelasnya (Touvron dkk., 2023). Meskipun *pretrained* model seperti LLaMA memiliki kemampuan awal yang sangat baik dalam memahami bahasa alami, model ini tidak cukup langsung digunakan untuk memecahkan masalah spesifik yang dihadapi dalam penelitian ini. Hal ini dikarenakan *pretrained* model dilatih pada data yang umum dan mungkin tidak memiliki pengetahuan atau kemampuan spesifik yang diperlukan untuk konteks atau domain tertentu, seperti *e-commerce*.

Dalam beberapa tahun terakhir, *Large Language Models* (LLM) telah mengalami perkembangan pesat dan menunjukkan kapabilitas luar biasa dalam pemahaman bahasa alami, generasi teks, serta penyelesaian tugas-tugas berbasis pengetahuan secara kontekstual. Meskipun LLM bersifat *general-purpose* dan mampu memberikan jawaban untuk berbagai topik, model ini tetap memiliki keterbatasan dalam menjawab pertanyaan domain-spesifik, seperti pertanyaan mengenai produk dalam platform *e-commerce* terutama ketika informasi yang dibutuhkan tidak terdapat dalam parameter model itu sendiri.

Untuk mengatasi keterbatasan tersebut, pendekatan *Retrieval-Augmented Generation (RAG)* dikembangkan sebagai solusi yang menggabungkan kemampuan sistem pencarian dokumen berbasis *vector retrieval* dengan kemampuan generatif LLM. Pendekatan ini dirancang agar *chatbot* tidak hanya menghasilkan respons generatif dari parameter internal model, tetapi juga dapat didukung oleh data aktual yang diambil secara *real-time* dari *knowledge base*. Melalui mekanisme *retrieval* yang memanfaatkan pencarian semantik dengan *vektor embedding*, *chatbot* dapat menemukan dokumen atau informasi produk yang paling relevan dengan pertanyaan pengguna. Pendekatan ini bekerja dengan mengambil dokumen relevan dari *knowledge base* eksternal menggunakan metode *dense retrieval*, lalu memberikan dokumen tersebut sebagai konteks tambahan ke dalam model LLM generatif untuk menghasilkan jawaban yang lebih akurat, relevan, dan berbasis fakta (Lewis dkk., 2020). Dengan demikian, pendekatan RAG tidak hanya membantu meminimalkan risiko jawaban asuntif atau spekulatif yang kerap muncul pada model generatif murni, tetapi juga memastikan bahwa jawaban yang diberikan selalu relevan dan selaras dengan data produk terbaru yang tersedia di platform *e-commerce*.

Dalam konteks *e-commerce*, penggunaan RAG memungkinkan pengembangan *chatbot* yang dapat menjawab pertanyaan calon pembeli secara otomatis berdasarkan data produk aktual, seperti deskripsi, spesifikasi, harga, dan kondisi barang. Alih-alih melatih ulang model dengan *fine-tuning*, pendekatan RAG memanfaatkan kemampuan *retrieval* untuk menyediakan informasi terkini dan relevan dari sumber eksternal, sehingga mengurangi kebutuhan komputasi serta meningkatkan fleksibilitas sistem terhadap perubahan data produk. Sistem ini juga lebih efisien secara operasional karena memungkinkan pemrosesan berbasis dokumen dinamis tanpa perlu memperbarui parameter internal model. Pendekatan ini secara fundamental mengatasi tantangan utama dalam domain *e-commerce* yaitu kebutuhan untuk memberikan jawaban yang selalu relevan dan berdasarkan data produk yang terus berubah, seperti ketersediaan stok, harga, dan promosi terbaru. Kemampuan adaptasi secara *real-time* inilah yang menjadikan RAG sebagai solusi strategis untuk menciptakan asisten belanja virtual yang andal di tengah pasar digital yang sangat kompetitif.

Penelitian ini berfokus pada pengembangan sistem *chatbot* dalam domain *e-commerce* yang didukung oleh LLM dan menggunakan pendekatan RAG sebagai solusi untuk menghasilkan jawaban otomatis yang akurat dan kontekstual terhadap pertanyaan pengguna. Dalam arsitektur ini, data produk seperti nama, deskripsi, spesifikasi teknis, harga, dan informasi relevan lainnya terlebih dahulu diproses menjadi representasi vektor menggunakan model *embedding* dan kemudian disimpan dalam bentuk indeks vektor di dalam *vector storage* berupa FAISS, yang memungkinkan pencarian semantik berbasis kedekatan makna antar kalimat. Ketika pengguna mengajukan pertanyaan, sistem tidak serta-merta mengandalkan pengetahuan internal dari LLM, melainkan melakukan *retrieval* terhadap sejumlah dokumen atau potongan informasi produk yang paling relevan dari basis data eksternal tersebut. Hasil pencarian ini kemudian digabungkan dengan pertanyaan pengguna dalam satu konteks *prompt*, yang selanjutnya diberikan kepada model generatif untuk menghasilkan respons yang lebih informatif, relevan, dan berbasis fakta.

Dengan mengintegrasikan pendekatan RAG ke dalam sistem *chatbot e-commerce*, penelitian ini menawarkan solusi yang secara teoritis mampu menjawab tantangan tingginya kompleksitas data produk serta kebutuhan informasi yang relevan dan selalu diperbarui. RAG memungkinkan *chatbot* untuk tidak hanya menghasilkan jawaban berbasis pengetahuan internal model LLM, tetapi juga memperkaya respons dengan data produk aktual yang diambil secara dinamis dan *real-time* dari *knowledge base*. Bagi pengguna, sistem ini menjanjikan pengalaman pencarian informasi yang lebih superior, di mana mereka dapat memperoleh jawaban yang akurat, instan, dan relevan tanpa harus menavigasi deskripsi produk yang panjang. Selain itu, otomatisasi yang dihasilkan juga dapat menekan biaya operasional karena mengurangi ketergantungan pada intervensi manual dalam layanan pelanggan. Hal ini diyakini dapat meningkatkan akurasi, relevansi, serta kepercayaan pengguna terhadap informasi yang diberikan, sekaligus mengurangi risiko jawaban yang tidak faktual atau spekulatif. Dengan demikian, pendekatan RAG menjadi landasan yang masuk akal dan strategis untuk mendukung terciptanya layanan asisten virtual yang responsif, informatif, dan adaptif terhadap perubahan data di ekosistem *e-commerce* yang kompetitif dan dinamis.

I.2 Perumusan Masalah

Rumusan masalah yang mendasari penelitian ini adalah:

- a. Bagaimana efektivitas penerapan sistem RAG dalam menghasilkan respons *chatbot* yang faktual serta relevan dengan konteks produk pada platform *e-commerce*?
- b. Bagaimana komponen *retriever* dalam sistem RAG dapat secara akurat memahami permintaan pengguna dan melakukan pencarian data produk yang relevan sesuai dengan permintaan tersebut?
- c. Bagaimana komponen LLM dalam sistem RAG dapat menghasilkan jawaban yang tidak hanya sesuai konteks hasil pencarian data produk, tetapi juga disampaikan secara natural dan mudah dipahami oleh pengguna?

I.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dijabarkan sebelumnya, penelitian ini bertujuan untuk:

- a. Menganalisis efektivitas keseluruhan sistem RAG dalam menghasilkan respons yang faktual dan relevan melalui serangkaian pengujian berbasis skenario pengguna pada platform *e-commerce*.
- b. Mengidentifikasi peran komponen *retriever* dalam sistem RAG untuk memahami permintaan pengguna serta melakukan pencarian data produk yang relevan sesuai dengan permintaan tersebut.
- c. Mengukur kualitas respons yang dihasilkan oleh komponen LLM berdasarkan kesesuaian jawaban terhadap konteks hasil pencarian data produk, serta struktur penyampaian yang terdengar natural sehingga mudah dipahami oleh pengguna.

I.4 Batasan Penelitian

Batasan masalah dalam penelitian ini meliputi:

- a. Penelitian ini berfokus pada pengembangan chatbot berbasis *Large Language Model* (LLM) dengan pendekatan *Retrieval-Augmented Generation* (RAG) untuk mendukung *automated query resolution* dalam platform *e-commerce*, khususnya pada pertanyaan yang berkaitan dengan informasi produk.

- b. Sistem hanya menggunakan teknik *dense retrieval* berbasis *vector embedding* untuk mengambil dokumen dari *knowledge base* yang disusun dari data produk *e-commerce*, tanpa melakukan *fine-tuning* ulang terhadap model LLM yang digunakan.
- c. Model LLM yang digunakan dalam penelitian ini adalah Llama-3.3-70B-Instruct, yang dilakukan *quantization 4-bit* menggunakan metode PEFT dengan parameter dan konfigurasi yang disesuaikan agar dapat dijalankan pada perangkat dengan keterbatasan sumber daya komputasi.
- d. Penelitian ini tidak mencakup pengembangan antarmuka pengguna secara menyeluruh maupun penerapan sistem komunikasi *chat* secara *real-time* dengan *WebSocket*, karena fokus utama adalah pada pembuatan model *chatbot* berbasis LLM yang terintegrasi dengan RAG.

I.5 Manfaat Penelitian

Manfaat penelitian ini:

- 1. Bagi Pengguna atau Pelanggan *e-commerce* : Penelitian ini memberikan manfaat langsung bagi pelanggan dengan menyediakan sistem *chatbot* yang mampu memberikan jawaban yang valid, akurat, dan relevan terhadap pertanyaan mereka mengenai produk atau layanan. Pelanggan akan lebih mudah berinteraksi dengan sistem *e-commerce*, menghemat waktu, dan mengurangi kemungkinan kesalahan dalam komunikasi terkait produk.
- 2. Bagi Pemilik Bisnis *E-commerce (Stakeholder)* : Penelitian ini dapat membantu *platform e-commerce* meningkatkan efisiensi dan akurasi dalam *automated query resolution* dan *order processing*. Dengan penerapan *Large Language Model (LLM)* dengan mekanisme RAG, *platform e-commerce* dapat memberikan respons yang lebih tepat dan cepat terhadap pertanyaan pelanggan, meningkatkan pengalaman pengguna, serta meningkatkan tingkat kepuasan pelanggan dan konversi penjualan.
- 3. Bagi Universitas Telkom: Sistem ini dapat meningkatkan citra institusi sebagai pelopor transformasi digital, terutama dalam menyediakan solusi yang mendukung efisiensi transaksi penjualan di *e-commerce*. Penelitian ini dapat meningkatkan reputasi universitas sebagai institusi yang mendukung

pengembangan teknologi berbasis *Artificial Intelligence*. Hasil penelitian ini juga dapat menjadi bukti kontribusi universitas dalam memberikan solusi inovatif untuk permasalahan industri.

4. Bagi Mahasiswa: Penelitian ini dapat menjadi studi kasus dan pembelajaran bagi mahasiswa untuk memperkuat kompetensi mereka di bidang teknologi informasi, khususnya dalam memahami konsep dan teknologi *Large Language Models* (LLM), *Natural Language Processing* (NLP), dan *quantized* model, beserta implementasinya. Selain itu, penelitian ini juga dapat menjadi panduan bagi mahasiswa yang ingin berkontribusi pada penelitian lain yang berkaitan dengan *Artificial Intelligence*, sekaligus memberikan wawasan praktis dalam mengembangkan dan menerapkan teknologi ini dalam berbagai solusi industri.
5. Bagi Bidang Sistem Informasi: Penelitian ini memberikan kontribusi penting bagi perkembangan bidang Sistem Informasi, terutama dalam integrasi teknologi AI ke dalam solusi bisnis. Selain memperkaya wawasan teknis dalam integrasi sistem, penelitian ini juga membuka peluang inovasi lebih lanjut untuk mengembangkan solusi cerdas berbasis AI yang dapat diimplementasikan di berbagai sektor dalam lingkup Sistem Informasi.

I.6 Sistematika Penulisan

Sistematika penulisan dalam penelitian ini, disusun sebagai berikut:

BAB I Pendahuluan

Pada bab ini memaparkan latar belakang penelitian, rumusan masalah, tujuan penelitian, batasan atau ruang lingkup penelitian, manfaat penelitian, serta penjelasan tentang struktur penulisan laporan ini.

BAB II Kajian Literatur

Pada bab ini berisi pembahasan mengenai literatur terkait yang relevan dengan penelitian. Di dalamnya, disertakan teori-teori pendukung yang mendasari masalah penelitian serta hubungan antara penelitian ini dengan studi lain yang serupa.

BAB III Metode Penelitian

Pada bab ini menguraikan pendekatan metodologi yang digunakan dalam penelitian, termasuk model konseptual dan langkah-langkah penyelesaian masalah. Langkah tersebut meliputi tahap identifikasi, pengumpulan data, analisis data, hingga pengambilan keputusan.

BAB IV Analisis dan Perancangan

Pada bab ini memuat analisis kebutuhan sistem serta proses perancangan solusi yang diusulkan dalam penelitian. Pembahasan mencakup identifikasi masalah, analisis kebutuhan fungsional dan non-fungsional, serta perancangan arsitektur sistem, alur kerja, dan komponen utama yang digunakan. Bab ini menjadi dasar dalam membangun sistem *chatbot* berbasis RAG yang sesuai dengan tujuan penelitian.

BAB V Implementasi dan Evaluasi

Pada bab ini menjelaskan penerapan hasil perancangan sistem ke dalam bentuk implementasi teknis pada sistem *retrieval* dan *generation*. Selain itu, bab ini juga memaparkan proses pengujian dan evaluasi terhadap kinerja sistem *chatbot* yang dibangun. Evaluasi dilakukan dengan menggunakan metrik tertentu untuk mengukur relevansi, akurasi, serta efektivitas sistem dalam menjawab pertanyaan pengguna.

BAB VI Kesimpulan dan Saran

Pada bab ini berisi ringkasan hasil penelitian yang mencakup kesimpulan utama dari analisis, perancangan, implementasi, dan evaluasi sistem yang telah dilakukan. Selain itu, rekomendasi diberikan sebagai masukan untuk pengembangan lebih lanjut, baik dari segi teknis, penggunaan, maupun potensi penelitian lanjutan di masa mendatang.