BAB I
PENDAHULUAN

A. Latar Belakang Masalah

Perkembangan industri otomotif di Indonesia menunjukkan pertumbuhan yang signifikan dalam beberapa tahun terakhir. Meningkatnya kebutuhan mobilitas masyarakat, pertumbuhan kelas menengah, serta kemudahan akses pembiayaan kendaraan mendorong tingginya permintaan mobil sebagai sarana transportasi utama (Gaikindo, 2026). Dalam proses pengambilan keputusan pembelian, konsumen tidak hanya mempertimbangkan harga, tetapi juga spesifikasi teknis seperti performa mesin, kapasitas penumpang, fitur keselamatan, serta teknologi dan kenyamanan yang disematkan. Sebagaimana dikutip dari buku karya Kotler & Keller (2016), keputusan pembelian produk bernilai tinggi seperti kendaraan roda empat termasuk dalam kategori *high-involvement decision* (keputusan dengan keterlibatan tinggi), yang menuntut konsumen untuk melakukan pertimbangan rasional dan emosional secara simultan sebelum mengambil keputusan final.

Dalam merespons perilaku konsumen yang kritis dan selektif tersebut, produsen otomotif dituntut untuk menyediakan pilihan produk yang mampu memenuhi berbagai kriteria spesifik. Toyota berhasil menjawab tantangan tersebut dan mengukuhkan posisinya sebagai pemimpin pasar otomotif nasional dengan menguasai 31,06% pangsa pasar ritel pada tahun 2025, mencatatkan total penjualan sebesar 258.923 unit (Gaikindo, 2026). Sebagai pemimpin pasar, Toyota menawarkan beragam lini produk, mulai dari *Low Cost Green Car* (LCGC), *Multi Purpose Vehicle* (MPV), *Sport Utility Vehicle* (SUV), hingga kendaraan elektrifikasi. Setiap model memiliki beberapa varian dengan perbedaan spesifikasi teknis dan fitur yang cukup kompleks (Motor, 2026). Berdasarkan data *pricelist* resmi, variasi harga dan tipe kendaraan dapat berbeda tergantung wilayah distribusi. Kompleksitas lini produk yang sangat masif inilah yang pada akhirnya menghadirkan tantangan operasional di jaringan *dealer* resmi, seperti Auto2000 Rantauprapat, terutama dalam proses pencocokan ratusan spesifikasi kendaraan dengan ekspektasi calon pembeli.

UIN IMAM BONJOL PADANG

Tantangan pencocokan spesifikasi tersebut menjadi semakin rumit karena dalam praktiknya, konsumen jarang mengungkapkan kebutuhan kendaraan dalam bentuk parameter teknis yang kaku dan terstruktur. Sebaliknya, kebutuhan tersebut lebih sering disampaikan menggunakan narasi bahasa alami (Gao dkk., 2021), seperti: "Saya butuh mobil keluarga 7 penumpang, yang irit dan nyaman untuk perjalanan jauh dengan budget sekitar 300 juta". Sistem pencarian berbasis filter konvensional gagal memfasilitasi hal ini karena hanya mampu memproses input berdasarkan kategori statis atau parameter numerik, sehingga tidak dapat memahami konteks semantik dari permintaan tersebut secara menyeluruh (Jannach dkk., 2021). Di sisi lain, proses pencocokan spesifikasi secara manual oleh tenaga pemasar memakan waktu dan berisiko tinggi mengabaikan alternatif kendaraan yang lebih tepat akibat besarnya volume data dan beban kognitif yang harus diproses (Rangarajan dkk., 2026). Oleh karena itu, masalah utama yang mendasari penelitian ini adalah ketiadaan sistem asisten cerdas di Auto2000 Rantauprapat yang mampu memetakan kebutuhan bahasa alami konsumen ke dalam pencarian spesifikasi teknis yang kompleks secara kontekstual, serta menekan risiko halusinasi informasi agar respons yang diberikan tetap berbasis fakta (faktual).

UIN IMAM BONJOL PADANG

Dari sudut pandang operasional dealer, kondisi tersebut menimbulkan konsekuensi bisnis yang nyata. Proses pencocokan spesifikasi yang masih bergantung sepenuhnya pada tenaga pemasar secara manual menyita waktu kerja yang seharusnya dapat dialokasikan untuk aktivitas penjualan bernilai tinggi, seperti negosiasi dan penutupan transaksi (Rangarajan dkk., 2026). Pada kondisi volume pertanyaan calon konsumen yang tinggi, keterbatasan kapasitas tenaga pemasar dalam merespons secara cepat dan akurat berpotensi memperlambat waktu pelayanan (*response time*), yang pada gilirannya dapat mendorong calon konsumen beralih mempertimbangkan dealer kompetitor dengan layanan informasi yang lebih responsif. Risiko ini menjadi krusial mengingat keputusan pembelian kendaraan tergolong *high-involvement decision* (Kotler & Keller, 2016), sehingga kualitas informasi dan kecepatan

respons pada tahap konsultasi produk turut memengaruhi kepercayaan calon konsumen terhadap dealer sebelum mengambil keputusan pembelian. Dengan demikian, ketiadaan asisten digital yang mampu menjawab kebutuhan spesifikasi secara cepat dan konsisten akurat bukan sekadar persoalan teknis, melainkan juga persoalan operasional yang berdampak langsung pada efisiensi kerja tenaga pemasar dan potensi kehilangan prospek penjualan (*lost lead*) di Auto2000 Rantauprapat.

Dalam ranah akademis, beberapa penelitian telah mencoba mengatasi kelemahan sistem pencarian konvensional tersebut. Sebagai contoh, Hibatullah & Baizal (2024) mengembangkan sistem rekomendasi interaktif yang menggabungkan riwayat pencarian pengguna dengan pemetaan kategori spesifikasi kendaraan secara terstruktur. Meskipun sistem ini cukup rapi dalam menyusun informasi spesifikasi dan mampu mencapai akurasi 85,39% , pendekatan ini memiliki kelemahan karena sifatnya yang masih kaku. Sistem tersebut sangat bergantung pada aturan pengelompokan data yang harus didefinisikan secara manual , mengharuskan pengguna melewati proses tanya-jawab yang panjang dan berulang , serta kesulitan memberikan rekomendasi kepada pengguna baru yang belum memiliki riwayat pencarian sama sekali (kendala ini dikenal dengan istilah *cold-start*).

UIN IMAM BONJOL PADANG

Seiring berkembangnya teknologi kecerdasan buatan, *Large Language Model* (LLM) mulai banyak digunakan untuk membuat sistem dapat memahami pertanyaan pengguna dengan bahasa yang lebih alami. Salah satu pendekatan yang dapat digunakan adalah *fine-tuning*, yaitu melatih kembali model menggunakan data tertentu agar lebih sesuai dengan kebutuhan sistem. Penelitian Wirawan & Baizal (2025), misalnya, menggunakan LLM yang telah dilatih ulang (*fine-tuning)* untuk membangun sistem rekomendasi mobil. Pendekatan ini dapat membantu AI memahami bahasa pengguna dan memberikan rekomendasi sesuai dengan data yang dipelajari. Namun, *fine-tuning* membutuhkan sumber daya komputasi yang besar dan tidak praktis untuk data yang sering berubah. Ketika terdapat perubahan harga atau penambahan varian mobil, model perlu dilatih kembali agar dapat

menggunakan informasi terbaru (Y. Gao dkk., 2024) . Selain itu, LLM tetap dapat menghasilkan informasi yang tidak sesuai dengan fakta atau disebut halusinasi karena sebagian jawaban dihasilkan berdasarkan pengetahuan yang tersimpan di dalam model, bukan berdasarkan data terbaru (Ji dkk., 2023).

Masalah tersebut menjadi penting dalam sistem rekomendasi mobil karena informasi seperti harga dan spesifikasi kendaraan dapat berubah ubah. AI yang hanya mengandalkan pengetahuan dari hasil pelatihan berisiko memberikan informasi yang sudah tidak sesuai dengan kondisi sebenarnya. Salah satu pendekatan yang dapat digunakan untuk mengatasi masalah tersebut adalah *Retrieval-Augmented Generation* (RAG). Penelitian Nishisako dkk. (2025) menunjukkan bahwa penggunaan informasi dari sumber eksternal dapat membantu mengurangi halusinasi pada model bahasa. Pada RAG, model AI tidak hanya mengandalkan pengetahuan yang telah dipelajarinya, tetapi terlebih dahulu mengambil informasi yang relevan dari sumber data yang tersedia. Informasi tersebut kemudian diberikan kepada model sebagai konteks untuk menghasilkan jawaban. Dengan cara ini, jawaban AI dapat lebih terarah pada data yang digunakan oleh sistem.

Pendekatan tersebut kemudian menjadi dasar pengembangan sistem rekomendasi mobil pada Auto2000 Rantauprapat. Dalam konteks implementasi industri, Auto2000 Rantauprapat telah merespons kebutuhan akan asisten digital dengan mengembangkan *website* resmi yang terintegrasi (toyotarantauprapat.vercel.app). Namun, fitur kecerdasan buatan pada sistem tersebut masih menggunakan pendekatan *static prompt-based knowledge injection*, yaitu memasukkan seluruh informasi mengenai spesifikasi dan harga mobil secara langsung ke dalam *system prompt*. Data katalog kendaraan dan contoh rekomendasi disimpan dalam satu teks yang panjang. Ketika pengguna memberikan pertanyaan, sistem hanya menambahkan informasi seperti *budget*, kategori kendaraan, dan kebutuhan penggunaan ke dalam teks tersebut sebelum dikirimkan ke model AI. Pendekatan ini menjadi kurang efektif ketika jumlah varian kendaraan dan data harga terus bertambah. Semakin banyak informasi yang dimasukkan ke dalam *prompt*, semakin besar penggunaan *token* dan

UIN IMAM BONJOL PADANG

semakin sulit bagi model untuk mempertahankan seluruh informasi dengan baik. Kondisi tersebut juga dapat meningkatkan risiko AI memberikan informasi yang tidak sesuai dengan data sebenarnya. Dari perspektif bisnis, risiko ini menimbulkan kerugian ganda: alih-alih memangkas beban kerja tenaga pemasar, kesalahan informasi yang dihasilkan justru menuntut tenaga pemasar melakukan verifikasi ulang secara manual, sekaligus berisiko menurunkan kepercayaan calon konsumen terhadap kredibilitas dealer.

Untuk mengatasi keterbatasan tersebut, penelitian ini menerapkan *Hybrid Retrieval-Augmented Generation* (*Hybrid* RAG) pada fitur asisten digital di *website* Auto2000 Rantauprapat. Pendekatan ini menggabungkan dua cara pencarian informasi, yaitu pencarian data terstruktur menggunakan SQL dan pencarian berdasarkan kemiripan makna menggunakan *Vector Search*. Penggabungan ini memungkinkan sistem menyesuaikan metode pencarian dengan kebutuhan pengguna. Sebagai contoh, ketika pengguna mencari mobil dengan harga di bawah Rp500 juta, sistem dapat menggunakan SQL untuk menyaring kendaraan berdasarkan harga. Sebaliknya, ketika pengguna bertanya mengenai mobil yang cocok untuk keluarga atau nyaman digunakan untuk perjalanan jauh, *Vector Search* dapat digunakan untuk menemukan data yang memiliki makna paling sesuai dengan pertanyaan tersebut. Hasil dari kedua pencarian kemudian digabungkan dan diperingkat ulang (*re-ranking*) untuk memperoleh kandidat kendaraan yang paling relevan. Kandidat tersebut selanjutnya digunakan sebagai konteks bagi Google Gemini dalam menghasilkan rekomendasi.

Penerapan Hybrid RAG membutuhkan sistem penyimpanan yang mampu mendukung kedua jenis pencarian tersebut. Pada umumnya, data terstruktur seperti harga dan spesifikasi kendaraan disimpan pada *database* biasa, sedangkan data vektor disimpan pada *database* khusus untuk melakukan pencarian berdasarkan kemiripan makna. Penggunaan dua *database* yang terpisah dapat membuat pengelolaan data menjadi lebih kompleks dan berisiko menimbulkan ketidaksesuaian ketika data diperbarui (PingCAP, 2024; Taipalus, 2024). Oleh karena itu, penelitian ini menggunakan TiDB sebagai

UIN IMAM BONJOL PADANG

tempat penyimpanan sekaligus pencarian data. TiDB mendukung SQL dan *Vector Search* dalam satu sistem *database*, sehingga data kendaraan seperti nama model, varian, harga, dan spesifikasi dapat disimpan sebagai data terstruktur, sementara data yang telah melalui proses *embedding* dapat digunakan untuk pencarian berdasarkan kemiripan makna. Dengan menggunakan TiDB, proses penyimpanan, pencarian, dan pembaruan data dapat dilakukan secara lebih terpusat.

Berdasarkan rancangan tersebut, sistem yang dikembangkan akan menggantikan metode *static prompt-based knowledge injection* (penyisipan pengetahuan statis ke dalam *prompt*) yang digunakan pada sistem sebelumnya dengan arsitektur *Hybrid* RAG. Dalam sistem ini, Google Gemini berperan dalam memahami pertanyaan dan menghasilkan jawaban, sedangkan TiDB berperan dalam menyimpan dan menyediakan data kendaraan yang dibutuhkan. Ketika pengguna memberikan pertanyaan, sistem akan menentukan data yang perlu dicari, mengambil informasi yang relevan melalui SQL dan *Vector Search*, kemudian menggunakan hasil pencarian tersebut sebagai konteks bagi Google Gemini. Dengan alur ini, pembaruan harga maupun spesifikasi kendaraan dapat dilakukan pada data di TiDB tanpa harus mengubah atau melatih kembali model AI. Bagi konsumen, pendekatan ini diharapkan dapat memberikan rekomendasi yang lebih relevan dan berdasarkan data kendaraan yang tersedia, sekaligus mengurangi risiko informasi yang tidak sesuai atau halusinasi.

UIN IMAM BONJOL PADANG

Dari sisi bisnis, hal ini diharapkan dapat mendukung efisiensi kerja tenaga pemasar dalam menangani pertanyaan repetitif, mempercepat waktu pelayanan, serta menjaga kredibilitas informasi yang disampaikan kepada calon konsumen selama proses konsultasi produk berlangsung.

B. Rumusan Masalah

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

Bagaimana mengembangkan asisten digital pada website Auto2000 Rantauprapat dari yang sebelumnya menggunakan metode *static prompt-based*

knowledge injection (penyisipan pengetahuan statis ke dalam prompt) menjadi sistem yang menerapkan arsitektur *Retrieval-Augmented Generation* (RAG) terintegrasi dengan basis data TiDB dan model Google Gemini, guna mendukung efisiensi proses konsultasi produk melalui rekomendasi mobil yang faktual berdasarkan basis data serta menekan tingkat halusinasi informasi?

C. Batasan Masalah

Agar implementasi dan pengembangan sistem ini lebih terarah, fokus, dan tidak menyimpang dari tujuan awal, maka ditetapkan beberapa batasan masalah sebagai berikut:

1. Data *knowledge base* pada sistem RAG dibatasi pada teks spesifikasi resmi (brosur) dan daftar harga *On The Road* (OTR) mobil baru Toyota wilayah Auto2000 Labuhanbatu per Mei 2026. Sistem hanya memproses data mobil baru dan tidak mencakup analisis harga mobil bekas atau depresiasi nilai jual.
2. Pengembangan difokuskan pada perombakan mesin AI menggunakan LLM Google Gemini dengan infrastruktur TiDB (*Hybrid Search*) untuk menggantikan metode *static prompt-based knowledge injection* (penyisipan pengetahuan statis ke dalam prompt). Sistem diintegrasikan sebagai modul pelengkap pada website Auto2000 Rantauprapat tanpa mengubah arsitektur *frontend* utama secara radikal.
3. Layanan AI dibatasi pada interaksi tanya-jawab serta rekomendasi kendaraan berdasarkan spesifikasi dan kebutuhan pengguna. Sistem dikembangkan sebagai prototipe menggunakan *Gemini API Free Tier*, sehingga tidak menangani proses transaksional (seperti kalkulasi kredit, promo *up to date*) dan dibatasi pada skenario pengguna tunggal (*single user*).
4. Pengujian difokuskan pada fungsionalitas antarmuka (Next.js) melalui *Black Box Testing* dan *System Usability Scale* (SUS), serta evaluasi arsitektur RAG menggunakan *DeepEval* dengan metrik *retrieval* (*Contextual Precision, Contextual Recall, Contextual Relevancy*) dan generatif (*Answer Relevancy, Faithfulness*). Penelitian ini tidak mencakup

UIN IMAM BONJOL PADANG

load testing, analisis mendalam UI/UX, maupun pengukuran dampak sistem terhadap efisiensi kerja tenaga pemasar, seperti waktu respons dan beban kerja sebelum dan sesudah implementasi.

D. Tujuan Penelitian

Tujuan umum dari penelitian ini adalah untuk mengembangkan dan mengimplementasikan sistem rekomendasi kendaraan Toyota berbasis arsitektur *Retrieval-Augmented Generation* (RAG) yang mampu memahami kebutuhan pengguna dalam bentuk bahasa alami serta menghasilkan rekomendasi yang relevan dan berbasis data aktual pada lingkungan Auto2000 Rantauprapat.

Secara khusus, rincian tujuan penelitian ini adalah sebagai berikut:

1. Mengembangkan asisten digital pada website Auto2000 Rantauprapat (*toyotarantauprapat.vercel.app*) dari yang sebelumnya menggunakan metode *static prompt-based knowledge injection* (penyisipan pengetahuan statis ke dalam prompt) menjadi sistem yang menerapkan arsitektur *Hybrid Retrieval-Augmented Generation* (RAG) terintegrasi dengan basis data TiDB dan model Google Gemini, sebagai upaya mendukung efisiensi proses konsultasi produk pada tenaga pemasar Auto2000 Rantauprapat.
2. Mengevaluasi tingkat efektivitas dan akurasi sistem yang dikembangkan dalam menghasilkan rekomendasi mobil yang faktual serta menekan tingkat halusinasi informasi, menggunakan kerangka kerja pengujian algoritma (*DeepEval*).

UIN IMAM BONJOL PADANG

E. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan kontribusi yang positif, baik dari segi akademis maupun praktis. Adapun rincian manfaat dari penelitian ini dijabarkan sebagai berikut:

1. Manfaat Teoretis:

   Penelitian ini berkontribusi mengisi *research gap* pada sistem rekomendasi cerdas di domain otomotif yang masih terbatas dibandingkan domain umum lainnya. Kontribusi utamanya terletak pada penerapan Hybrid RAG (gabungan pencarian SQL dan vektor) yang dispesifikasikan

untuk rekomendasi mobil. Selain itu, penelitian ini memvalidasi penggunaan DeepEval sebagai instrumen evaluasi objektif untuk mengukur faktualitas AI secara matematis.

2. Manfaat Praktis:
   a. Bagi Auto2000 Rantauprapat

      Penelitian ini menghasilkan solusi pemodelan asisten digital cerdas yang memiliki kapabilitas untuk melayani kueri pelanggan secara otomatis tanpa batasan waktu. Sebagai sebuah prototipe, sistem ini mendemonstrasikan bagaimana analisis instruksi pelanggan dapat dilakukan secara instan untuk menyajikan rekomendasi varian mobil yang faktual sesuai spesifikasi resmi. Implementasi konsep ini diharapkan dapat memberikan gambaran nyata dalam memangkas beban kerja tenaga penjual (*sales*) dalam merespons pertanyaan yang bersifat repetitif, sekaligus mempercepat waktu pelayanan (*response time*). Selain itu, prototipe ini juga bermanfaat bagi perusahaan sebagai instrumen edukasi produk yang valid dan berbasis data, sehingga mampu menekan risiko penyampaian informasi keliru kepada calon pembeli sebelum mereka diarahkan untuk melakukan interaksi lanjutan atau *test drive* secara langsung di *dealer* Auto2000 Rantauprapat.

   b. Bagi Pengguna Sistem (*End-User*)

      Sistem ini mempermudah calon pembeli dalam menavigasi kompleksitas varian dan spesifikasi kendaraan Toyota melalui antarmuka interaktif yang mendukung instruksi bahasa alami. Calon konsumen dapat langsung memperoleh rekomendasi mobil yang paling relevan dengan preferensi mereka berdasarkan anggaran (*budget*), kapasitas penumpang, jenis bahan bakar, serta fitur teknis yang diinginkan, secara cepat dan mandiri, dengan informasi yang faktual serta minim risiko halusinasi kecerdasan buatan.

   c. Bagi Penulis

UIN IMAM BONJOL PADANG

Penelitian ini menjadi wadah pembuktian untuk mengimplementasikan keilmuan Rekayasa Perangkat Lunak dan Kecerdasan Buatan tingkat lanjut pada skala industri nyata (*production environment*). Pengalaman merancang arsitektur *Hybrid RAG*, melakukan rekayasa *Vector Database* (TiDB), serta mengintegrasikan *Large Language Model* ke dalam kerangka kerja *Full-stack* (*Next.js*) ini secara praktis akan mengasah dan meningkatkan kompetensi analitis dan teknis peneliti di bidang pengembangan perangkat lunak (*Software Engineering*) dan arsitektur sistem.

UIN IMAM BONJOL
PADANG