Integrated Journal of Applied Technology and Informatics
Journal homepage: https://journal2.uinssc.ac.id/index.php/IJATI

Google Gemini-Based Service Chatbot for Indomaret Pekanbaru

Implementasi *Chatbot* Layanan Indomaret Pekanbaru Berbasis Google Gemini

Daffa Al Syaddad, Safa Salsabila, Eni Urbaningrum, Muhammad Rafi, Rasyidi Amara, Chairun Nas
Program Studi Sistem Informasi, Universitas Riau, Pekanbaru, 28293, Indonesia
daffa.al5922@student.unri.ac.id, safa.salsabila1608@student.unri.ac.id, eni.urbaningrum1305@student.unri.ac.id, muhammad.rafi4646@student.unri.ac.id, rasyidi.amara6345@student.unri.ac.id, chairun.nas@lecturer.unri.ac.id

ARTICLE INFO
Article history:
Received 2026-06-30
Revised 2026-06-30
Accepted 2026-06-30

CORRESPONDING AUTHOR
Chairun Nas
chairun.nas@lecturer.unri.ac.id

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International.

ABSTRACT
Modern retail customer service requires fast, consistent, and accessible information delivery. This study implements a text-based chatbot prototype for Indomaret Pekanbaru services using Google Gemini and semantic similarity retrieval. The dataset uses a question-answering structure that contains service contexts, user questions, and expected answers related to operating hours, payment methods, promotions, digital top-up services, bill payments, package delivery, and branch information. User input is normalized through preprocessing, converted into sentence embeddings with Sentence Transformer, and compared with dataset questions using cosine similarity to select the most relevant context. The selected context and user question are then sent to the Gemini API to generate a contextual response. Functional testing used five service-question scenarios. The chatbot produced appropriate answers in three scenarios and partially appropriate answers in two scenarios. The main limitation appears when requested information is too specific or unavailable in the dataset. The results indicate that semantic retrieval and generative language models can support flexible customer service automation, but dataset expansion and response validation remain necessary.
Keyword: *Chatbot, Google Gemini, Semantic Similarity, Customer Service, Retail Information System*

ABSTRAK
Pelayanan pelanggan ritel modern membutuhkan penyampaian informasi yang cepat, konsisten, dan mudah diakses. Penelitian ini mengimplementasikan prototipe *chatbot* berbasis teks untuk layanan Indomaret Pekanbaru dengan Google Gemini dan pencarian konteks berbasis *semantic similarity*. *Dataset* disusun dalam bentuk *question-answering* yang memuat konteks layanan, pertanyaan pengguna, dan jawaban yang diharapkan terkait jam operasional, metode pembayaran, promo, layanan *top up*, pembayaran tagihan, pengiriman paket, dan informasi cabang. *Input* pengguna dinormalisasi melalui *preprocessing*, diubah menjadi *sentence embedding* menggunakan *Sentence Transformer*, lalu dibandingkan dengan pertanyaan *dataset* menggunakan *cosine similarity* untuk memilih konteks paling relevan. Konteks terpilih dan pertanyaan pengguna dikirimkan ke API Gemini untuk menghasilkan respons kontekstual. Pengujian fungsional dilakukan pada lima skenario pertanyaan layanan. *Chatbot* memberikan jawaban sesuai pada tiga skenario dan kurang sesuai pada dua skenario. Keterbatasan utama muncul ketika informasi yang diminta terlalu spesifik atau belum tersedia pada *dataset*. Hasil ini menunjukkan bahwa pencarian semantik dan model bahasa generatif dapat mendukung otomasi layanan pelanggan, tetapi perluasan *dataset* dan validasi respons tetap diperlukan.
Keyword: *Chatbot*, Google Gemini, *Semantic Similarity*, Layanan Pelanggan, Sistem Informasi Ritel

1. PENDAHULUAN

Layanan pelanggan pada sektor ritel modern semakin bergantung pada kecepatan akses informasi dan konsistensi jawaban. Pelanggan minimarket membutuhkan informasi yang bersifat rutin, seperti jam operasional, metode pembayaran, promo, layanan *top up*, pembayaran tagihan, dan informasi cabang. Jika pertanyaan tersebut hanya ditangani secara manual, proses layanan dapat terhambat oleh antrean, keterbatasan waktu kerja, dan perbedaan kualitas jawaban antar petugas. *Natural Language Processing* (NLP) banyak digunakan untuk mendukung layanan pelanggan karena mampu memproses pertanyaan pengguna dalam bentuk bahasa alami dan menghubungkannya dengan jawaban yang relevan [1], [2]. Perkembangan *chatbot* berbasis AI juga menunjukkan peningkatan kebutuhan terhadap sistem percakapan yang lebih adaptif dalam layanan digital [3].

Model bahasa generatif memberi peluang baru bagi *chatbot* layanan pelanggan. Model tersebut tidak hanya memilih jawaban dari aturan tetap, tetapi juga menyusun respons berdasarkan konteks yang diberikan. Studi tentang implementasi *chatbot* menunjukkan bahwa sistem percakapan dapat mendukung efisiensi layanan dan kepuasan pelanggan jika didukung desain interaksi, basis pengetahuan, dan evaluasi respons yang memadai [4]. Kajian tentang *generative conversational* AI juga menunjukkan adanya peluang pada layanan digital, riset, dan praktik organisasi, tetapi tetap memunculkan tantangan akurasi, etika, dan tata kelola [5]. Dengan demikian, *chatbot* layanan pelanggan perlu menggabungkan kemampuan generatif dan pembatasan konteks agar jawaban tetap relevan.

Laporan awal penelitian ini menggunakan istilah model GPT berbasis Google Gemini. Dalam artikel ini, istilah tersebut disesuaikan menjadi model bahasa generatif berbasis Google Gemini karena Gemini bukan produk GPT OpenAI, meskipun sama-sama memanfaatkan pendekatan model bahasa besar berbasis arsitektur *Transformer*. Penelitian terkait Gemini-*powered chatbot* menunjukkan bahwa Gemini dapat digunakan untuk mendukung pemrosesan bahasa alami dan pembangkitan respons percakapan [6]. Sementara itu, laporan teknis Gemini menjelaskan bahwa Gemini mer*up*akan keluarga model multimodal yang dirancang untuk pemahaman dan penalaran lintas konteks [7], dan pengembangan Gemini 1.5 menekankan efisiensi serta kemampuan konteks panjang yang dapat mendukung aplikasi berbasis percakapan [8]. Penyesuaian istilah ini penting agar artikel lebih akurat secara teknis.

Salah satu tantangan utama *chatbot* generatif adalah risiko jawaban yang tidak didukung oleh data layanan yang tersedia. Pendekatan *retrieval-augmented generation* menjawab masalah ini dengan menggabungkan kemampuan generasi model bahasa dan pencarian dokumen atau konteks eksternal [9]. Evaluasi RAG menunjukkan bahwa kualitas *retrieval*, kemampuan menolak informasi yang tidak tersedia, serta integrasi konteks memengaruhi kualitas jawaban model [10]. Pada penelitian ini, konsep tersebut diterapkan dalam bentuk sederhana melalui pencarian konteks pada *dataset question-answering* sebelum model Google Gemini menghasilkan respons. Strategi ini dipilih agar jawaban *chatbot* tetap diarahkan pada informasi layanan Indomaret Pekanbaru yang tersedia pada *dataset*.

Penelitian ini bertujuan mengimplementasikan dan mengevaluasi prototipe *chatbot* layanan Indomaret Pekanbaru berbasis Google Gemini dengan metode *semantic similarity*. Fokus penelitian meliputi penyusunan *dataset question-answering*, *preprocessing* teks, pencarian konteks menggunakan *Sentence Transformer* dan *cosine similarity*, pembangkitan jawaban dengan API Gemini, serta pengujian fungsional terhadap beberapa skenario pertanyaan layanan. Artikel ini berkontribusi pada penerapan praktis *chatbot* generatif untuk layanan informasi ritel lokal. Kontribusi tersebut bersifat prototipe akademik dan tidak merepresentasikan sistem resmi perusahaan.

2. METODE

2.1. Desain Penelitian

Penelitian ini menggunakan pendekatan implementatif dengan model pengembangan prototipe. Tujuan utama penelitian bukan membangun sistem produksi penuh, melainkan membuktikan alur kerja *chatbot* generatif yang dapat menjawab pertanyaan layanan pelanggan berdasarkan *dataset* terbatas. Sistem dikembangkan menggunakan bahasa pemrograman Python pada Google Colaboratory. Komponen utama sistem terdiri dari *dataset* layanan, modul *preprocessing*, modul *semantic similarity*, *prompt construction*, API Google Gemini, dan antarmuka percakapan berbasis teks.

2.2. Dataset Question-Answering

Tabel 1. Contoh *Dataset Question-Answering* Layanan Indomaret Pekanbaru

Context	*Question*	*Answer*
Indomaret Panam buka 24 jam setiap hari.	Jam berapa Indomaret Panam buka?	24 jam setiap hari.
Indomaret menerima pembayaran QRIS.	Apakah bisa bayar menggunakan QRIS?	Ya, tersedia pembayaran QRIS.
Indomaret pertama di Pekanbaru berada di Jalan Sudirman.	Di mana lokasi Indomaret pertama di Pekanbaru?	Jalan Sudirman.

Dataset disusun dalam format *question-answering* yang memuat tiga elemen, yaitu *context*, *question*, dan *answer*. Elemen *context* berisi informasi layanan yang menjadi dasar jawaban. Elemen *question* berisi contoh pertanyaan pengguna yang mungkin diajukan. Elemen *answer* berisi jawaban ringkas yang diharapkan dari sistem. Informasi yang dimasukkan mencak*up* jam operasional, metode pembayaran, promo, layanan *top up*, pembayaran tagihan, pengiriman paket, dan informasi umum cabang. Struktur ini dipilih agar proses *retrieval* dapat menghubungkan pertanyaan pengguna dengan konteks yang paling dekat secara semantik.

2.3. Preprocessing Teks

Preprocessing dilakukan sebelum proses pencarian kemiripan. Tahap ini mencak*up* perubahan huruf menjadi huruf kecil, penghapusan tanda baca, pembersihan spasi berlebih, dan normalisasi bentuk teks sederhana. *Preprocessing* bertujuan mengurangi variasi penulisan yang tidak relevan terhadap makna pertanyaan. Dengan tahap ini, pertanyaan seperti 'Apakah bisa bayar pakai QRIS?' dan 'apakah bisa bayar menggunakan qris' diproses sebagai *input* yang lebih konsisten.

2.4. Pencarian Konteks Menggunakan Semantic Similarity

Pencarian konteks dilakukan dengan mengubah pertanyaan pengguna dan pertanyaan dalam *dataset* menjadi *sentence embedding*. *Sentence embedding* memungkinkan kalimat direpresentasikan sebagai vektor sehingga hubungan makna antarkalimat dapat dihitung. *Sentence*-BERT menjadi salah satu rujukan utama karena menghasilkan *embedding* kalimat yang dapat digunakan untuk menghitung kemiripan semantik secara efisien [11]. Studi tentang GPT *sentence embedding* menunjukkan bahwa representasi kalimat juga dapat digunakan untuk *semantic* search [12]. Kajian terbaru tentang peningkatan text *embedding* dengan model bahasa besar menunjukkan bahwa representasi kalimat yang kuat dapat meningkatkan kualitas pencarian semantik pada berbagai tugas temu balik informasi [13].

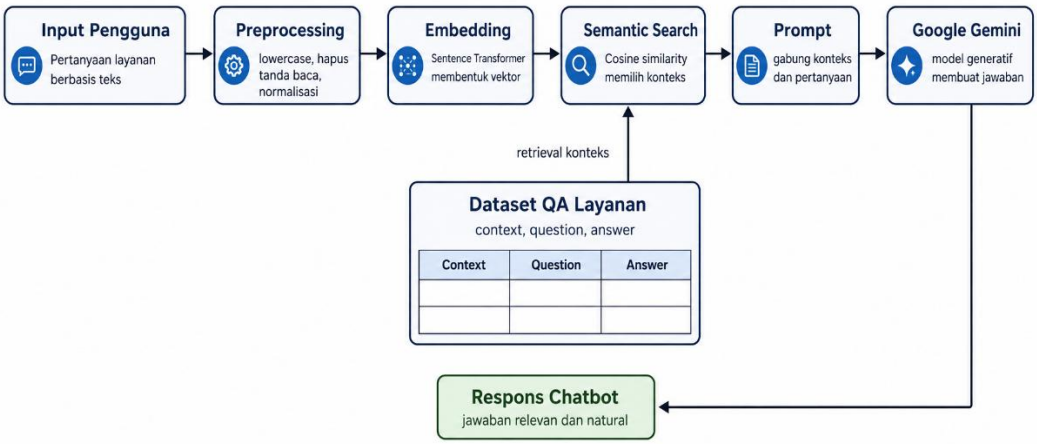
Setelah *embedding* terbentuk, sistem menghitung kemiripan antara vektor pertanyaan pengguna dan vektor pertanyaan pada *dataset*. Nilai kemiripan tertinggi digunakan untuk memilih konteks yang akan dikirimkan ke model Gemini. Rumus *cosine similarity* ditunjukkan pada Persamaan (1).

$$\cos(A,B) = (A \cdot B) / (||A||\ ||B||) \quad (1)$$

Pada Persamaan (1), A merepresentasikan vektor pertanyaan pengguna, sedangkan B merepresentasikan vektor pertanyaan pada *dataset*. Nilai *cosine similarity* yang mendekati 1 menunjukkan hubungan semantik yang semakin relevan. Konteks dengan nilai tertinggi dipilih sebagai dasar jawaban *chatbot*.

2.5. Pembangkitan Jawaban Menggunakan Google Gemini

Konteks terpilih digabungkan dengan pertanyaan pengguna dalam *prompt* yang dikirimkan ke API Google Gemini. *Prompt* disusun agar model menjawab berdasarkan konteks yang tersedia dan tidak memperluas jawaban ke informasi yang tidak ada pada *dataset*. Strategi ini diperlukan karena model bahasa besar dapat menghasilkan jawaban yang tampak meyakinkan, tetapi tidak selalu faktual jika tidak diarahkan oleh konteks yang jelas. Tinjauan tentang halusinasi LLM menunjukkan bahwa pembangkitan teks yang tidak didukung fakta masih menjadi risiko penting pada sistem generatif [16].



Gambar 1. Arsitektur Sistem *Chatbot* Layanan Indomaret Pekanbaru Berbasis Google Gemini

2.6. Pengujian Sistem

Pengujian dilakukan menggunakan pendekatan *black-box* terhadap lima skenario pertanyaan layanan. Pertanyaan pengujian mencak*up* jam operasional cabang, metode pembayaran QRIS, lokasi cabang pertama, layanan *top up* DANA, dan promo mingguan. Setiap keluaran *chatbot* dinilai secara manual dengan kategori sesuai atau kurang sesuai. Kategori sesuai diberikan jika jawaban menjawab pertanyaan dan konsisten dengan konteks *dataset*. Kategori kurang sesuai diberikan jika jawaban masih relevan secara umum, tetapi tidak menjawab informasi spesifik yang diminta. Evaluasi ini juga mempertimbangkan faktor kepercayaan pengguna terhadap sistem AI, karena desain dan akurasi respons memengaruhi penerimaan sistem berbasis AI [15].

2.7. Alur Algoritmik Sistem

Alur kerja sistem dirancang agar setiap pertanyaan pengguna melewati tahapan yang konsisten. Pertama, sistem menerima pertanyaan dalam bentuk teks. Kedua, sistem melakukan *preprocessing* agar variasi penulisan tidak mengganggu proses pencarian. Ketiga, sistem membentuk *embedding* dari pertanyaan pengguna dan membandingkannya dengan *embedding* pertanyaan pada *dataset*. Keempat, sistem memilih konteks dengan nilai *cosine similarity* tertinggi. Kelima, konteks dan pertanyaan disusun ke dalam *prompt* yang dikirimkan ke Google Gemini untuk menghasilkan jawaban akhir.

Alur tersebut mendukung prinsip *grounding* karena model generatif tidak langsung menjawab tanpa konteks. Model diberi batas informasi dari *dataset* yang telah dipilih melalui proses *retrieval*. Desain ini tidak menghilangkan seluruh risiko kesalahan, tetapi membantu mengurangi kemungkinan jawaban yang terlalu bebas. Pada implementasi berikutnya, *threshold* dapat ditambahkan agar sistem hanya menjawab ketika nilai *similarity* memenuhi batas minimal. Jika nilai *similarity* berada di bawah ambang, sistem dapat mengembalikan respons *fallback*.

Tabel 2. *Pseudocode* Alur Kerja *Chatbot* Layanan Indomaret Pekanbaru.

Langkah	Proses
1	Terima pertanyaan pengguna melalui antarmuka teks.
2	Lakukan *preprocessing*: *lowercase*, hapus tanda baca, dan normalisasi spasi.
3	Bangun *embedding* pertanyaan pengguna menggunakan *Sentence Transformer*.
4	Hitung *cosine similarity* terhadap seluruh pertanyaan pada *dataset* QA.
5	Pilih konteks dengan nilai *similarity* tertinggi.
6	Susun *prompt* yang memuat instruksi, konteks terpilih, dan pertanyaan pengguna.
7	Kirim *prompt* ke API Google Gemini.
8	Tampilkan respons *chatbot* kepada pengguna dan simpan hasil untuk evaluasi.

2.8. Kriteria Evaluasi Respons

Evaluasi respons dilakukan secara manual karena jumlah skenario pengujian masih terbatas. Penilaian tidak hanya melihat kemiripan kata, tetapi juga kecocokan maksud pertanyaan, ketepatan informasi, dan kejelasan jawaban. Pendekatan ini sesuai untuk tahap awal pengembangan prototipe karena pengembang perlu melihat pola kesalahan sebelum menyusun pengujian kuantitatif yang lebih besar. Kriteria evaluasi yang jelas juga membantu membedakan jawaban yang benar-benar sesuai dari jawaban yang hanya relevan secara umum.

Kategori sesuai diberikan ketika jawaban memuat informasi inti yang diminta pengguna dan tidak bertentangan dengan *dataset*. Kategori kurang sesuai diberikan ketika jawaban masih berada pada *top*ik yang sama, tetapi kurang spesifik atau tidak memberikan detail yang diminta. Kategori tidak sesuai dapat digunakan pada pengembangan berikutnya jika sistem menghasilkan jawaban yang keluar dari konteks, bertentangan dengan *dataset*, atau tidak menjawab pertanyaan pengguna.

Tabel 3. Kriteria Evaluasi Respons *Chatbot*

Kategori	Kriteria
Sesuai	Jawaban memuat informasi yang diminta, konsisten dengan konteks *dataset*, dan dapat dipahami pengguna.
Kurang sesuai	Jawaban masih relevan dengan *top*ik, tetapi belum menjawab detail spesifik yang diminta pengguna.
Tidak sesuai	Jawaban keluar dari konteks, bertentangan dengan *dataset*, atau tidak menjawab pertanyaan pengguna.

3. HASIL DAN PEMBAHASAN

3.1. Implementasi Prototipe

Prototipe *chatbot* berhasil diimplementasikan pada Google Colaboratory. Pengguna mengirim pertanyaan melalui antarmuka teks, kemudian sistem melakukan *preprocessing* terhadap *input* tersebut. Setelah itu, sistem membentuk *embedding* pertanyaan dan menghitung *cosine similarity* terhadap pertanyaan yang tersimpan di *dataset*. Konteks dengan nilai kemiripan tertinggi dikirim ke Gemini untuk menghasilkan jawaban. Alur kerja ini membuat *chatbot* lebih fleksibel dibandingkan *chatbot* berbasis aturan karena sistem tidak hanya mencocokkan kata kunci, tetapi juga mempertimbangkan kedekatan makna antarkalimat [2], [14].

Penggunaan *semantic similarity* membantu sistem menangani variasi pertanyaan pengguna. Misalnya, pertanyaan tentang pembayaran QRIS dapat muncul dalam bentuk 'bisa bayar QRIS?', 'apakah menerima QRIS?', atau 'pembayaran pakai QRIS tersedia?'. Dengan *embedding*, variasi tersebut dapat diarahkan pada konteks yang sama jika *dataset* memuat informasi layanan QRIS. Studi penggunaan *sentence transformer* pada pencocokan teks menunjukkan bahwa model ini dapat mendukung pencarian semantik pada data yang memiliki variasi bahasa dan konteks [12], [13].

3.2. Hasil Pengujian Fungsional

Hasil pengujian menunjukkan bahwa *chatbot* memberikan jawaban sesuai pada tiga dari lima skenario. Jawaban yang sesuai muncul pada pertanyaan tentang jam operasional Indomaret Panam, pembayaran menggunakan QRIS, dan lokasi Indomaret pertama di Pekanbaru. Dua skenario lain memperoleh kategori kurang sesuai, yaitu pertanyaan tentang *top up* DANA dan promo mingguan. Keduanya kurang sesuai karena sistem menjawab secara umum, tetapi belum memberikan rincian yang diminta pengguna.

Temuan tersebut menunjukkan bahwa kualitas jawaban sangat bergantung pada cak*up*an *dataset*. Jika informasi yang diminta tersedia secara eksplisit pada *dataset*, *semantic retrieval* dapat menemukan konteks yang relevan dan Gemini mampu menyusun jawaban yang *natural*. Jika informasi tidak tersedia atau terlalu umum, model cenderung memberikan jawaban yang masih terkait *top*ik, tetapi belum spesifik. Kondisi ini sejalan dengan temuan penelitian RAG yang menekankan bahwa model generatif membutuhkan *retrieval* yang kuat dan konteks yang tepat untuk menghasilkan jawaban yang akurat [9], [10].

Tabel 4. Hasil Pengujian Fungsional *Chatbot* Layanan Indomaret Pekanbaru

No.	Pertanyaan Pengguna	Jawaban *Chatbot*	Hasil
1	Jam operasional cabang Indomaret Panam?	24 jam nons*top* setiap hari.	Sesuai
2	Apakah bisa bayar menggunakan QRIS?	Ya, Indomaret menyediakan pembayaran menggunakan QRIS dan beberapa *e-wallet*.	Sesuai
3	Di mana lokasi Indomaret pertama di Pekanbaru?	Indomaret pertama di Pekanbaru berada di Jalan Sudirman pada tahun 1995.	Sesuai
4	Apakah tersedia layanan *top up* DANA?	Indomaret menyediakan layanan pembayaran digital pada beberapa cabang tertentu.	Kurang sesuai
5	Promo apa yang tersedia minggu ini?	Tersedia promo akhir pekan untuk beberapa produk pilihan di Indomaret.	Kurang sesuai

Tabel 5. Ringkasan hasil pengujian

Kategori	Jumlah Skenario	Persentase	Interpretasi
Sesuai	3	60%	Jawaban menjawab pertanyaan dan sesuai dengan konteks *dataset*.
Kurang sesuai	2	40%	Jawaban masih terkait *top*ik, tetapi belum spesifik terhadap informasi yang diminta.
Total	5	100%	Pengujian awal menunjukkan kebutuhan perluasan *dataset* dan validasi respons.

3.3. Pembahasan

Integrasi *semantic similarity* dan Google Gemini memberi dua keuntungan utama. Pertama, *semantic similarity* membantu sistem menemukan konteks berdasarkan makna, bukan hanya kesamaan kata. Kedua, Gemini membantu menyusun respons yang lebih *natural* dan mudah dipahami pengguna. Kombinasi ini membuat *chatbot* lebih adaptif terhadap variasi pertanyaan pelanggan dibandingkan pendekatan *rule-based* yang sangat bergantung pada pola pertanyaan tetap [3], [6], [11].

Namun, penerapan model generatif pada layanan pelanggan tetap membutuhkan pembatasan dan validasi. Model dapat menghasilkan respons yang baik secara bahasa, tetapi jawaban tersebut belum tentu benar jika konteks yang diberikan tidak memadai. Karena itu, *dataset* harus diperluas, diperbarui, dan diuji secara berkala. Sistem juga perlu mekanisme *fallback*, seperti jawaban 'informasi belum tersedia' atau pengalihan ke petugas manusia jika pertanyaan berada di luar cak*up*an *dataset*.

Dari sisi penggunaan, prototipe ini relevan untuk layanan informasi ritel yang memiliki pertanyaan berulang. Layanan seperti jam buka, metode pembayaran, promo, dan layanan digital dapat dijawab secara otomatis jika data tersedia. Dari sisi pengembangan, model ini dapat diperluas menjadi aplikasi *web* atau *mobile* dengan *backend* terintegrasi basis data layanan. Dari sisi tata kelola, penggunaan nama dan informasi merek harus mengikuti izin dan kebijakan pemilik merek jika sistem akan digunakan secara publik.

3.4. Keterbatasan

Penelitian ini memiliki beberapa keterbatasan. Pertama, *dataset* masih terbatas sehingga jawaban pada pertanyaan spesifik belum selalu akurat. Kedua, evaluasi masih menggunakan lima skenario pengujian sehingga belum cuk*up* untuk menyimpulkan performa sistem secara menyeluruh. Ketiga, sistem masih bergantung pada koneksi internet dan ketersediaan API Gemini. Keempat, artikel ini tidak mengevaluasi metrik kuantitatif seperti *accuracy*, *precision*, *recall*, F1-score, *response time*, dan *user satisfaction*. Keterbatasan tersebut perlu diperbaiki pada penelitian berikutnya agar sistem lebih siap untuk diuji dalam lingkungan operasional.

3.5. Kesiapan Data dan Kualitas Pengetahuan

Kualitas *dataset* menjadi faktor teknis yang paling menentukan performa *chatbot*. *Dataset* yang hanya memuat contoh pertanyaan terbatas dapat membuat sistem gagal ketika pengguna memakai istilah berbeda, bertanya dengan ejaan tidak baku, atau meminta rincian yang belum tersedia. Oleh karena itu, *dataset* layanan perlu disusun sebagai basis pengetahuan yang memiliki cak*up*an *top*ik, variasi pertanyaan, dan jawaban terverifikasi. Setiap informasi layanan juga perlu dilengkapi status pembaruan agar sistem tidak menjawab berdasarkan data yang sudah kedaluwarsa.

Dalam konteks layanan ritel, pembaruan data bersifat penting karena promo, metode pembayaran, dan layanan digital dapat berubah secara berkala. Sistem harus membedakan informasi stabil, seperti jenis layanan umum, dan informasi dinamis, seperti promo mingguan. Informasi dinamis sebaiknya dihubungkan dengan sumber data yang dapat diperbarui, bukan hanya teks statis di dalam *file* JSON. Pemisahan ini membantu menurunkan risiko jawaban yang benar secara bahasa tetapi salah secara faktual.

3.6. Analisis Kesalahan Respons

Dua skenario pengujian memperoleh kategori kurang sesuai karena sistem memberikan jawaban umum. Pada pertanyaan *top up* DANA, *chatbot* menjawab bahwa layanan pembayaran digital tersedia pada beberapa cabang tertentu, tetapi belum memastikan ketersediaan *top up* DANA secara spesifik. Pada pertanyaan promo mingguan, *chatbot* menjawab adanya promo akhir pekan, tetapi tidak menunjukkan daftar produk, periode promo, atau syarat promosi. Pola kesalahan ini menunjukkan bahwa model generatif mampu mempertahankan konteks umum, tetapi tidak dapat mengganti kekurangan data spesifik pada basis pengetahuan.

Kesalahan tersebut tidak semata-mata berasal dari model Gemini. Pada arsitektur yang menggunakan *retrieval*, jawaban akhir sangat dipengaruhi oleh konteks yang dipilih sebelum proses generasi. Jika konteks yang dipilih terlalu umum, jawaban yang dihasilkan juga akan cenderung umum. Jika konteks yang dipilih tidak memuat fakta yang diminta, model dapat menghasilkan respons yang tampak wajar tetapi tidak menjawab kebutuhan pengguna. Karena itu, evaluasi sistem perlu memeriksa dua hal secara terpisah, yaitu ketepatan *retrieval* dan ketepatan generasi jawaban.

3.7. Implikasi Implementasi

Hasil penelitian ini menunjukkan bahwa *chatbot* berbasis Google Gemini dapat digunakan sebagai prototipe layanan informasi ritel. Sistem cocok untuk pertanyaan berulang yang memiliki jawaban stabil, seperti jam operasional, metode pembayaran, dan informasi layanan umum. Untuk pertanyaan yang berubah cepat, sistem membutuhkan integrasi data yang lebih kuat. Integrasi tersebut dapat ber*up*a basis data promo, API internal, atau *dashboard* admin yang memungkinkan petugas memperbarui informasi tanpa mengubah kode program.

Dari sudut pandang pengalaman pengguna, *chatbot* perlu memberikan jawaban singkat, jelas, dan dapat ditindaklanjuti. Jika sistem tidak memiliki data, jawaban yang aman adalah menyampaikan bahwa informasi belum tersedia dan menyarankan pengguna menghubungi petugas atau cabang terkait. Strategi ini lebih baik daripada menghasilkan jawaban spekulatif. Dengan demikian, prototipe perlu ditingkatkan dengan *confidence threshold*, *fallback response*, dan log percakapan untuk analisis perbaikan *dataset*.

3.8. Validasi, Etika, dan Tata Kelola

Penelitian ini tidak menggunakan data pribadi pelanggan dan tidak mengakses sistem internal perusahaan. Data yang digunakan disusun untuk kebutuhan prototipe akademik. Namun, jika sistem akan diterapkan secara publik, pengembang perlu memperhatikan izin penggunaan merek, akurasi informasi, perlindungan data pengguna, dan transparansi bahwa pengguna sedang berinteraksi dengan *chatbot*. Prinsip tersebut penting karena kesalahan informasi pada layanan pelanggan dapat memengaruhi keputusan pengguna.

Tabel 6. Keterbatasan sistem dan rekomendasi pengembangan

Aspek	Keterbatasan	Rekomendasi
Dataset	Cak*up*an pertanyaan dan jawaban masih terbatas.	Menambah variasi pertanyaan, sinonim, dan konteks layanan yang diverifikasi.
Informasi dinamis	Promo dan layanan cabang dapat berubah cepat.	Menghubungkan *chatbot* dengan basis data atau *dashboard* pembaruan informasi.
Retrieval	Konteks yang dipilih dapat terlalu umum.	Menambahkan *confidence threshold* dan evaluasi *similarity* score.
Generasi jawaban	Model dapat menjawab secara *natural* tetapi kurang spesifik.	Membatasi *prompt* agar menjawab hanya dari konteks dan memakai *fallback response*.
Evaluasi	Pengujian masih terbatas pada lima skenario.	Melakukan uji kuantitatif dengan lebih banyak pertanyaan dan melibatkan pengguna nyata.

Tata kelola sistem juga perlu mencakup audit respons. Audit dapat dilakukan dengan menyimpan pertanyaan pengguna, konteks yang dipilih, jawaban model, dan penilaian kualitas jawaban. Catatan tersebut membantu pengembang

mengidentifikasi *top*ik yang sering gagal dijawab. Selain itu, audit memungkinkan perbaikan *dataset* secara bertahap dan membantu memastikan bahwa model tidak menjawab di luar cak*up*an layanan.

4. KESIMPULAN

Penelitian ini berhasil mengimplementasikan prototipe *chatbot* layanan Indomaret Pekanbaru berbasis Google Gemini dengan dukungan pencarian konteks menggunakan *semantic similarity*. Sistem memproses pertanyaan pengguna melalui *preprocessing*, membentuk *sentence embedding*, menghitung *cosine similarity* terhadap *dataset question-answering*, memilih konteks paling relevan, dan menghasilkan jawaban menggunakan model Gemini. Hasil pengujian fungsional pada lima skenario menunjukkan tiga jawaban sesuai dan dua jawaban kurang sesuai. Temuan ini menunjukkan bahwa integrasi *retrieval* semantik dan model generatif dapat mendukung otomasi layanan pelanggan berbasis teks.

Kualitas respons *chatbot* sangat dipengaruhi oleh cak*up*an dan kualitas *dataset*. Pertanyaan yang informasinya tersedia pada *dataset* menghasilkan jawaban yang lebih sesuai, sedangkan pertanyaan yang terlalu spesifik atau belum memiliki konteks lengkap menghasilkan jawaban yang kurang tepat. Penelitian berikutnya perlu memperluas *dataset*, menambahkan pengujian kuantitatif, mengukur waktu respons, melibatkan pengguna nyata, dan mengembangkan antarmuka *web* atau *mobile*. Sistem juga perlu mekanisme validasi, pembaruan data, dan *fallback* ke petugas manusia agar lebih aman digunakan pada layanan publik.

DAFTAR PUSTAKA

[1] M. Mashaabi, A. Alotaibi, H. Qudaih, R. Alnashwan, and H. Al-Khalifa, "Natural language processing in customer service: A systematic review," arXiv preprint arXiv:2212.09523, 2022, doi: 10.48550/arXiv.2212.09523.

[2] M. S. Nurmansyah and H. Zakaria, "Implementasi metode Natural Language Processing (NLP) pada chatbot dalam sebuah aplikasi pesan berbasis web untuk meningkatkan layanan pelanggan (Studi Kasus: Reicomp)," LOGIC: Jurnal Ilmu Komputer dan Pendidikan, vol. 2, pp. 1032-1042, Oct. 2024.

[3] Kholisatunnisa and D. N. Misidawati, "Peran teknologi AI dalam pengembangan chatbot layanan pelanggan," Jurnal Sistem Informasi, Manajemen dan Teknologi Informasi, vol. 2, pp. 92-97, Jul. 2024.

[4] N. A. Wahyudivanti, A. D. Armunanto, and A. D. Arrayan, "Analisis dampak implementasi chatbot terhadap kepuasan pelanggan dan efisiensi operasional bisnis: Studi literatur berbasis grounded theory," Jurnal Sistem Informasi (TEKNOFILE), vol. 3, pp. 879-892, Dec. 2025.

[5] Y. K. Dwivedi et al., "So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy," International Journal of Information Management, vol. 71, Art. no. 102642, 2023, doi: 10.1016/j.ijinfomgt.2023.102642.

[6] S. AlGhozali and S. Mukminatun, "Natural Language Processing of Gemini Artificial Intelligence Powered Chatbot," Balangkas, vol. 1, no. 1, pp. 41-48, Jul. 2024, doi: 10.66317/688570.

[7] Gemini Team et al., "Gemini: A family of highly capable multimodal models," arXiv preprint arXiv:2312.11805, 2023, doi: 10.48550/arXiv.2312.11805.

[8] Gemini Team et al., "Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context," arXiv preprint arXiv:2403.05530, 2024, doi: 10.48550/arXiv.2403.05530.

[9] Y. Gao et al., "Retrieval-augmented generation for large language models: A survey," arXiv preprint arXiv:2312.10997, 2023, doi: 10.48550/arXiv.2312.10997.

[10] J. Chen, H. Lin, X. Han, and L. Sun, "Benchmarking large language models in retrieval-augmented generation," arXiv preprint arXiv:2309.01431, 2023, doi: 10.48550/arXiv.2309.01431.

[11] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence embeddings using Siamese BERT-networks," in Proc. 2019 Conf. Empirical Methods in Natural Language Processing and 9th Int. Joint Conf. Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 2019, pp. 3982-3992, doi: 10.18653/v1/D19-1410.

[12] N. Muennighoff, "SGPT: GPT sentence embeddings for semantic search," arXiv preprint arXiv:2202.08904, 2022, doi: 10.48550/arXiv.2202.08904.

[13] L. Wang, N. Yang, X. Huang, L. Yang, R. Majumder, and F. Wei, "Improving text embeddings with large language models," arXiv preprint arXiv:2401.00368, 2024, doi: 10.48550/arXiv.2401.00368.

[14] D. Biesner, M. Pielka, R. Ramamurthy, T. Dilmaghani, B. Kliem, R. Loitz, and R. Sifa, "Zero-shot text matching for automated auditing using sentence transformers," arXiv preprint arXiv:2211.07716, 2022, doi: 10.48550/arXiv.2211.07716.

[15] T. A. Bach, A. Khan, H. Hallock, G. Beltrao, and S. Sousa, "A systematic literature review of user trust in AI-enabled systems: An HCI perspective," arXiv preprint arXiv:2304.08795, 2023, doi: 10.48550/arXiv.2304.08795.

[16] L. Huang et al., "A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions," arXiv preprint arXiv:2311.05232, 2023, doi: 10.48550/arXiv.2311.05232.