

## IMPLEMENTASI RAG MULTIMODAL UNTUK REKOMENDASI TEKS DAN GAMBAR DI E-COMMERCE

**Bagas Wahyu Herdiansyah<sup>1\*</sup>, Mukhammad Andri Setiawan<sup>2</sup>**

Universitas Islam Indonesia, Indonesia

\*Corresponding author: [21523013@students.uii.ac.id](mailto:21523013@students.uii.ac.id)

**Abstract:** This study evaluates the effectiveness of Multimodal Retrieval-Augmented Generation (RAG) as a recommendation system in e-commerce platforms. Traditional search systems, whether keyword-based or semantic, often fail to capture the intent, context, and visual aspects of user needs, which negatively impacts the quality of the online shopping experience. Although Multimodal RAG promises more holistic interactions, empirical studies on user perceptions of its advantages remain limited. This research aims to design and implement a functional Multimodal RAG system, as well as evaluate user experience qualitatively compared to traditional search methods. An e-commerce web application was developed with four search features (keyword, text semantic, image semantic, and Multimodal RAG) using the SigLIP model for embeddings and Gemini 2.5 Pro for response generation. Qualitative evaluation with a task scenario design and the Think-Aloud Protocol shows that the Multimodal RAG feature consistently became the participants' primary choice, especially in complex and ambiguous search scenarios. Participants assessed that this system was able to provide justified, adaptive, and consultative recommendations, thereby overcoming the limitations of conventional approaches.

**Keywords:** Interactive Video, Neck Injury, Learning Media, Rapid Prototyping Model, Adolescents

**Abstrak:** Penelitian ini mengevaluasi efektivitas Retrieval-Augmented Generation (RAG) Multimodal sebagai sistem rekomendasi di platform e-commerce. Sistem pencarian tradisional, baik berbasis kata kunci maupun semantik, seringkali gagal menangkap niat, konteks, dan aspek visual dari kebutuhan pengguna, yang berdampak negatif pada kualitas pengalaman belanja daring. Meskipun RAG Multimodal menjanjikan interaksi yang lebih holistik, studi empiris mengenai persepsi pengguna terhadap keunggulannya masih terbatas. Penelitian ini bertujuan merancang dan mengimplementasikan sistem RAG Multimodal fungsional, serta mengevaluasi pengalaman pengguna secara kualitatif dibandingkan dengan metode pencarian tradisional. Sebuah aplikasi web e-commerce dikembangkan dengan empat fitur pencarian (kata kunci, semantik teks, semantik gambar, dan RAG Multimodal) menggunakan model SigLIP untuk embedding dan Gemini 2.5 Pro untuk generasi respons. Evaluasi kualitatif dengan desain task scenario dan Think-Aloud Protocol menunjukkan bahwa fitur RAG Multimodal secara konsisten menjadi pilihan utama partisipan, terutama dalam skenario pencarian yang kompleks dan ambigu. Partisipan menilai sistem ini mampu memberikan rekomendasi yang terjustifikasi, adaptif, dan konsultatif, sehingga mengatasi keterbatasan pendekatan konvensional.

**Kata kunci:** Video Interaktif, Cedera Leher, Media Pembelajaran, Rapid Prototyping Model, Remaja

## PENDAHULUAN

Ekonomi digital Indonesia telah menunjukkan pertumbuhan yang luar biasa, didorong oleh peningkatan penetrasi internet yang masif. Jumlah pengguna internet di Indonesia melonjak signifikan, dari tingkat penetrasi 73,7% pada periode 2019-2020, menjadi 79,5% pada awal tahun 2024 (Erlina F. Santika, 2024). Angka ini setara dengan 221,5 juta jiwa yang telah terhubung ke internet, menjadikan Indonesia salah satu pasar digital terbesar di Asia Tenggara. Pertumbuhan ini secara langsung mengakselerasi ekspansi sektor e-commerce, yang secara konsisten menjadi pilar utama dan pendorong pertumbuhan ekonomi digital nasional (Pratiwi, 2022).

Peningkatan basis pengguna ini berkorelasi langsung dengan lonjakan nilai transaksi e-commerce. Nilai Gross Merchandise Value (GMV) yang pada tahun 2020 tercatat sebesar \$44 miliar terus menunjukkan pertumbuhan yang pesat (Google, 2025). Berbagai studi memproyeksikan nilai pasar e-commerce Indonesia akan mencapai antara \$82 miliar hingga \$124 miliar pada tahun 2025, menegaskan posisinya sebagai salah satu pasar terbesar di kawasan ini (Google, 2025). Pertumbuhan eksponensial ini menandakan pergeseran fundamental dalam perilaku konsumen dan lanskap ritel di Indonesia, di mana e-commerce menjadi enabler penting bagi inklusi ekonomi, terutama bagi UMKM.

Seiring dengan matangnya pasar yang ditandai dengan melambatnya laju akuisisi pengguna baru, fokus platform e-commerce bergeser dari akuisisi ke retensi dan peningkatan nilai seumur hidup pelanggan. Dalam lingkungan yang semakin kompetitif, kualitas interaksi dan kemudahan dalam menemukan produk menjadi diferensiator utama untuk mempertahankan loyalitas. Pengalaman pencarian produk yang buruk, yang menimbulkan friksi dan frustrasi, secara langsung berkontribusi pada tingkat konversi yang lebih rendah dan perpindahan pelanggan ke platform pesaing (Bai et al., 2020). Oleh karena itu, optimalisasi teknologi pencarian bukan lagi sekadar peningkatan teknis, melainkan sebuah keharusan strategis untuk pertumbuhan bisnis yang berkelanjutan.

Secara historis, teknologi pencarian produk di e-commerce telah berevolusi melalui beberapa paradigma utama. Tahap awal didominasi oleh pencarian berbasis kata kunci (keyword-based search), sebuah metode yang bekerja dengan mencocokkan secara harfiah istilah yang dimasukkan pengguna dengan data produk yang ada di dalam indeks (Rokon et al., 2024). Meskipun cepat dan efisien untuk kueri yang sangat spesifik seperti nomor seri atau nama produk yang pasti, pendekatan ini pada dasarnya bersifat kaku dan

rapuh (Rokon et al., 2024). Keterbatasan ini mendorong pengembangan teknologi yang lebih canggih untuk memahami maksud pengguna dengan lebih baik.

Paradigma berikutnya adalah pencarian semantik (semantic search), yang menandai lompatan signifikan dalam kemampuan sistem untuk memahami niat pengguna. Berbeda dengan pencocokan literal, pencarian semantik memanfaatkan Natural Language Processing (NLP) dan vector embeddings untuk memahami makna dan konteks di balik sebuah kueri (Rokon et al., 2024). Teknologi ini mampu menginterpretasikan hubungan antar konsep, mengenali sinonim, dan memahami maksud pengguna bahkan ketika kata-kata yang digunakan tidak sama persis dengan yang ada di deskripsi produk. Dengan demikian, pencarian semantik mulai menjembatani kesenjangan antara bahasa alami pengguna dan data terstruktur di dalam basis data e-commerce. Meskipun merupakan kemajuan, kedua paradigma pencarian sebelumnya memiliki keterbatasan fundamental dalam konteks e-commerce modern. Pencarian berbasis kata kunci sering kali gagal karena dua masalah utama: masalah sinonim, di mana sistem tidak dapat menghubungkan istilah yang berbeda tetapi bermakna sama seperti "pakaian" dan "busana"; dan masalah ambiguitas atau polisemi, di mana satu kata dapat memiliki banyak arti seperti "Apple" (merek vs. buah) (Rokon et al., 2024). Kegagalan ini menghasilkan interaksi yang membuat frustrasi, dengan hasil yang tidak relevan (false positives) atau tidak ditemukannya produk yang dicari (false negatives) (Rokon et al., 2024).

Pencarian semantik berbasis teks, meskipun mampu mengatasi masalah sinonim dan ambiguitas, memiliki kelemahan kritis yang disebut "kebutaan visual" (visual blindspot) (Liang et al., 2022). Dalam domain e-commerce, penampilan visual suatu produk seringkali menjadi faktor penentu utama dalam keputusan pembelian, bahkan lebih penting daripada deskripsi tekstualnya (Liang et al., 2022). Sistem pencarian yang hanya mengandalkan teks tidak mampu memproses atau memahami informasi kaya yang terkandung dalam gambar, seperti gaya, warna, pola, atau estetika keseluruhan, yang seringkali sulit diartikulasikan oleh pengguna dengan kata-kata.

Keterbatasan ini secara kolektif menciptakan apa yang dikenal sebagai "friksi pencarian" (search friction), yaitu hambatan yang mempersulit pengguna untuk menemukan produk yang mereka inginkan. Menurut Teori Beban Kognitif (Cognitive Load Theory), peningkatan friksi ini secara langsung meningkatkan beban kognitif ekstrinsik—upaya mental yang tidak perlu yang disebabkan oleh cara informasi disajikan

(Li et al., 2025). Ketika beban kognitif ekstrinsik terlalu tinggi, kapasitas mental pengguna untuk mengevaluasi produk (tugas utama) menjadi berkurang, yang pada akhirnya menyebabkan frustrasi, kelelahan dalam mengambil keputusan, dan potensi pengabaian sesi belanja (Li et al., 2025).

Perkembangan terkini dalam Kecerdasan Buatan (AI) telah melahirkan paradigma baru yang dikenal sebagai perdagangan percakapan (*conversational commerce*). Pendekatan ini mengubah interaksi belanja daring dari model kueri-respons yang statis menjadi dialog interaktif dua arah yang dinamis. Tujuannya adalah untuk mereplikasi pengalaman konsultatif saat berbelanja di toko fisik, di mana pelanggan dapat berdiskusi, mengajukan pertanyaan lanjutan, dan menerima rekomendasi yang dipersonalisasi dari seorang asisten penjualan (Jannach, 2021).

Teknologi kunci yang memungkinkan pergeseran ini adalah Large Language Models (LLM). Dengan kemampuan pemahaman dan generasi bahasa alami yang canggih, LLM dapat memfasilitasi interaksi yang menyerupai manusia dalam skala besar, memberikan respons yang relevan secara kontekstual dan dipersonalisasi. Penggunaan LLM membuka peluang untuk menciptakan interaksi pelanggan yang lebih menarik dan memuaskan, yang pada akhirnya dapat meningkatkan loyalitas dan tingkat konversi secara signifikan.

Meskipun LLM memiliki kemampuan percakapan yang kuat, penerapannya secara langsung dalam domain kritis seperti e-commerce menghadapi dua tantangan utama: halusinasi dan pengetahuan yang usang. LLM diketahui dapat menghasilkan informasi yang terdengar meyakinkan namun secara faktual salah (halusinasi), dan pengetahuannya terbatas pada data terakhir yang digunakan untuk pelatihannya (Gupta et al., 2024). Isu-isu ini membuat LLM yang berdiri sendiri tidak dapat diandalkan untuk memberikan informasi produk yang akurat dan terkini.

Untuk mengatasi masalah ini, arsitektur Retrieval-Augmented Generation (RAG) diperkenalkan sebagai kerangka kerja AI yang inovatif (Gupta et al., 2024). RAG meningkatkan keandalan LLM dengan menghubungkannya ke basis pengetahuan eksternal yang dapat diverifikasi. Proses RAG terdiri dari dua fase utama: pertama, fase Retrieval, di mana sistem mengambil informasi atau fakta yang relevan dari sumber data eksternal (misalnya, katalog produk); kedua, fase Generation, di mana LLM menggunakan fakta yang telah diambil tersebut sebagai konteks untuk menghasilkan

respons yang akurat dan relevan (Gupta et al., 2024). Dengan demikian, RAG memastikan bahwa jawaban yang diberikan didasarkan pada data yang faktual dan mutakhir, sehingga meningkatkan kepercayaan pengguna (Gupta et al., 2024).

Konvergensi dari kemajuan teknologi yang telah dibahas—perdagangan percakapan, LLM, dan RAG—melahirkan solusi generasi berikutnya: RAG Multimodal. Sistem ini tidak hanya mewarisi kemampuan percakapan dari LLM dan landasan faktual dari RAG, tetapi juga secara krusial menambahkan kemampuan untuk memahami dan memproses informasi visual, mengatasi "kebutaan visual" dari sistem sebelumnya. Dengan mengintegrasikan data dari berbagai modalitas seperti teks dan gambar, sistem RAG Multimodal dapat menjawab kueri yang kompleks dan bernuansa yang tidak mungkin ditangani oleh sistem unimodal (Mortaheb et al., 2025).

Untuk membangun sistem RAG Multimodal yang fungsional, penelitian ini akan memanfaatkan kombinasi model-model canggih. Pada lapisan retrieval, sistem akan menggunakan model embedding google/siglip-so400m-patch14-384. Model ini dipilih karena kemampuannya yang unggul dalam menciptakan ruang embedding bersama untuk input teks dan gambar, yang difasilitasi oleh inovasi utamanya pada penggunaan fungsi Sigmoid Loss (Zhai et al., 2023). Fungsi loss ini terbukti lebih efisien dan stabil dibandingkan dengan softmax-based contrastive loss yang digunakan oleh model-model sebelumnya, memungkinkan pencarian kesamaan semantik yang cepat dan akurat (Zhai et al., 2023).

Pada lapisan generasi, sistem akan menggunakan Google Gemini 2.5 Pro. Model ini dipilih karena tiga keunggulan utamanya: pertama, ia dirancang sebagai model yang natively multimodal, mampu memahami dan menalar berdasarkan konteks yang mengandung teks dan gambar secara bersamaan. Kedua, ia menggunakan arsitektur Mixture-of-Experts (MoE) yang efisien, memungkinkan kinerja model raksasa dengan biaya komputasi yang jauh lebih rendah. Ketiga, Gemini 2.5 Pro memiliki jendela konteks yang sangat besar hingga 1 juta token, memungkinkannya memproses sejumlah besar informasi yang diambil dari katalog produk dalam satu prompt tunggal, yang krusial untuk menghasilkan respons yang komprehensif dan bernuansa (Gemini Team, 2025).

Pendekatan ini mengubah fungsi pencarian dari sekadar "mesin pencari" menjadi "asisten belanja konsultatif" yang cerdas. Pengguna dapat mengajukan permintaan yang

menggabungkan teks dan gambar, misalnya, "Carikan saya gaun yang mirip dengan ini [gambar], tetapi dengan warna biru dan cocok untuk pernikahan di musim panas." Namun, meskipun potensi teknologinya sangat besar, evaluasi terhadap sistem RAG, terutama yang bersifat multimodal, sebagian besar masih berfokus pada metrik kuantitatif yang berorientasi pada sistem seperti akurasi, waktu penyelesaian tugas, dan relevansi jawaban.

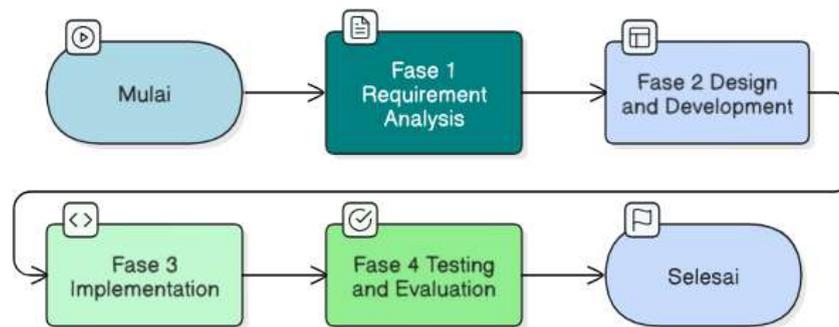
Fokus pada validasi teknis ini telah menciptakan kesenjangan yang signifikan dalam pemahaman tentang bagaimana pengguna manusia berinteraksi dengan dan merasakan sistem-sistem canggih ini. Terdapat kekurangan studi yang bergerak "melampaui akurasi" untuk mengeksplorasi aspek kualitatif dari interaksi pengguna (Bach et al., 2024). Metrik kuantitatif tidak dapat sepenuhnya menangkap persepsi subjektif seperti kemudahan penggunaan, relevansi yang dirasakan (*perceived relevance*), kepercayaan pada sistem, dan kepuasan secara keseluruhan, yang merupakan faktor-faktor penentu keberhasilan adopsi di dunia nyata (Bach et al., 2024).

Oleh karena itu, penelitian ini dirancang untuk mengisi kesenjangan tersebut. Dengan mengimplementasikan solusi teknis canggih yang menggabungkan kekuatan embedding dari SigLIP dan kemampuan penalaran multimodal dari Gemini 2.5 Pro, penelitian ini bertujuan untuk memberikan pemahaman empiris tentang interaksi pengguna dengan asisten belanja AI generasi baru. Analisis ini akan menyoroti nilai tambah yang dirasakan dari interaksi multimodal dibandingkan dengan metode pencarian tradisional, serta mengidentifikasi skenario di mana pendekatan konsultatif ini memberikan manfaat paling signifikan, sehingga memberikan kontribusi penting bagi pengembangan sistem AI yang lebih berpusat pada manusia.

Penelitian ini bertujuan untuk menganalisis dan membandingkan secara kualitatif pengalaman pengguna (*user experience*) saat berinteraksi dengan sistem RAG Multimodal versus metode pencarian tradisional (berbasis kata kunci, semantik, dan gambar), dengan evaluasi yang berfokus pada relevansi yang dirasakan (*perceived relevance*), kemudahan penggunaan, dan kemampuan sistem dalam menangani kueri yang kompleks serta bernuansa multimodal. Selanjutnya, penelitian ini akan mengidentifikasi skenario-skenario spesifik di mana sistem RAG Multimodal memberikan nilai tambah paling signifikan bagi pengguna, termasuk menganalisis faktor-faktor kualitatif yang membuat pendekatan percakapan dan multimodal ini terasa lebih

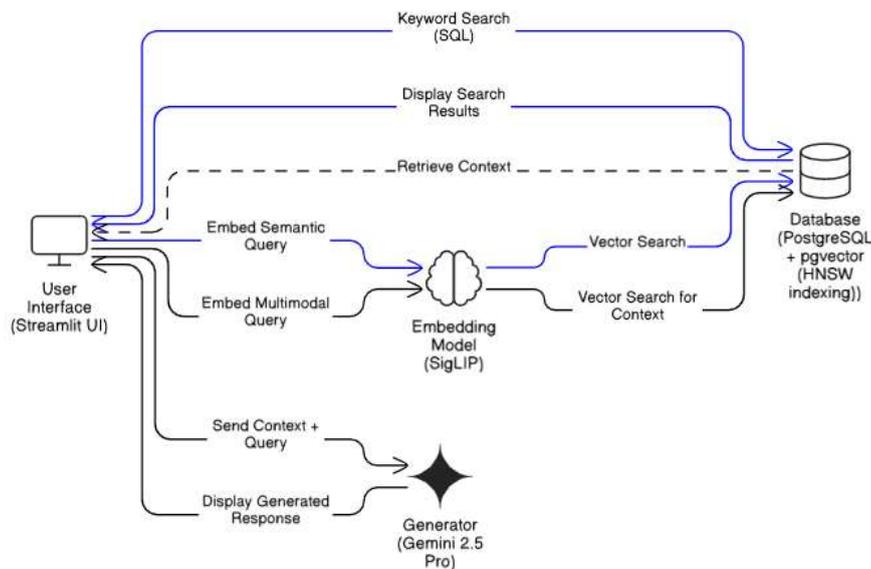
unggul. Berdasarkan temuan tersebut, tujuan akhirnya adalah untuk menghasilkan sebuah model konseptual awal dan rekomendasi berbasis bukti mengenai perilaku serta ekspektasi pengguna terhadap asisten belanja AI, yang diharapkan dapat menjadi referensi bagi para praktisi dalam merancang sistem AI yang lebih konsultatif, personal, dan berpusat pada manusia (*human-centered*).

## METODE



**Gambar 1.** Alur metode Incremental Model

Penelitian ini mengadopsi kerangka kerja pengembangan sistem *Incremental Model* sesuai dengan Gambar 1. Metodologi ini dipilih secara strategis karena sangat sesuai dengan sifat penelitian yang melibatkan pembangunan beberapa modul fungsional yang berbeda namun saling terkait (pencarian kata kunci, semantik, dan RAG). Dengan memecah proses pengembangan menjadi fase-fase sekuensial yaitu: *Requirement Analysis*, *Design & Development*, *Implementation*, dan *Testing & Evaluation*. pendekatan ini memungkinkan setiap fungsionalitas dapat dibangun, diuji secara internal, dan divalidasi secara terpisah (Jovanović et al., 2024). Alur kerja yang terstruktur ini secara efektif mengurangi risiko teknis dan memastikan setiap komponen, mulai dari fitur pencarian *baseline* hingga fitur RAG yang kompleks, berfungsi dengan baik sebelum diintegrasikan secara penuh.



**Gambar 2.** Flow Arsitektur Sistem

Sesuai dengan Gambar 2. Arsitektur dan Implementasi Sistem Sistem rekomendasi dibangun sebagai aplikasi web fungsional menggunakan *framework Streamlit* sebagai antarmuka pengguna. Fondasi data sistem berasal dari 11.604 data produk unik (kategori *footwear*) yang diperoleh melalui proses *web scraping* dari situs e-commerce Amazon UK. Data mentah ini kemudian melalui tahap pra-pemrosesan yang meliputi penghapusan data duplikat, penanganan nilai kosong, dan pembersihan kolom teks untuk menjaga kualitas data.

Inti dari sistem ini adalah arsitektur RAG Multimodal. Alur kerja dimulai dari antarmuka pengguna, di mana input (teks dan/atau gambar) diterima. Untuk pencarian semantik dan RAG, input ini diubah menjadi representasi vektor berdimensi 1152 oleh model *embedding google/siglip-so400m-patch14-384*. Vektor ini kemudian digunakan untuk melakukan pencarian kemiripan (*similarity search*) pada basis data PostgreSQL yang telah dilengkapi ekstensi *pgvector*. Untuk mengakselerasi proses pencarian pada jutaan vektor, indeks HNSW (*Hierarchical Navigable Small World*) diimplementasikan.

Pada alur RAG, hasil pencarian kemiripan (tahap *retrieval*) berupa sekumpulan produk yang paling relevan, digunakan sebagai konteks. Konteks ini, bersama dengan kueri asli pengguna, digabungkan ke dalam sebuah *prompt* terstruktur (tahap *augmentation*) dan dikirim ke model bahasa Gemini 2.5 Pro. Model kemudian menghasilkan respons percakapan yang informatif dan konsultatif (tahap *generation*) yang disajikan kembali kepada pengguna.

Prosedur Evaluasi Kualitatif Evaluasi kualitatif dirancang untuk memahami secara mendalam pengalaman pengguna. Partisipan dipilih menggunakan teknik *purposive sampling*, dengan kriteria utama adalah individu yang merupakan pengguna aktif platform e-commerce (melakukan pembelian alas kaki daring setidaknya 1-2 kali dalam 3 bulan terakhir) dan memiliki kemampuan verbal yang baik.

Setiap sesi evaluasi dimulai dengan pengenalan singkat mengenai empat fitur pencarian yang tersedia. Partisipan kemudian diminta untuk melakukan serangkaian tugas berbasis skenario yang dirancang untuk secara alami mendorong penggunaan fitur yang berbeda:

1. **Skenario Spesifik:** Menemukan produk ‘Running Shoes’ (target: Pencarian Kata Kunci).
2. **Skenario Konseptual:** Mencari ‘sepatu hiking yang bagus untuk pemula dan tahan air’ (target: Pencarian Semantik Teks).
3. **Skenario Visual:** Menemukan produk yang mirip dengan gambar sepatu boots unik tanpa mengetahui nama atau mereknya (target: Pencarian Semantik Gambar).
4. **Skenario Multimodal:** Mencari sepatu lari yang mirip dengan gambar yang diberikan, namun dengan spesifikasi tambahan ‘solnya lebih empuk dan cocok untuk lari jarak jauh’ (target: RAG).
5. **Skenario Ambigius & Konsultatif:** Meminta beberapa ide hadiah berupa ‘sneakers gaya vintage warna cerah’ untuk seorang sahabat (target: RAG).

Selama mengerjakan tugas, data dikumpulkan melalui metode *Think-Aloud Protocol*, di mana partisipan menyuarakan proses berpikir, perasaan, dan kebingungan mereka secara verbal. Sesi diakhiri dengan wawancara mendalam semi-terstruktur untuk menggali persepsi pengguna lebih jauh. Data transkrip yang terkumpul kemudian dianalisis menggunakan Analisis Tematik untuk mengidentifikasi pola, wawasan, dan tema-tema kunci.

## HASIL DAN PEMBAHASAN

### Hasil

#### Implementasi Sistem Berdasarkan Increment

##### 1. Increment 1: Fondasi Data dan Embedding

Dataset 11.604 produk alas kaki dari Amazon UK diperoleh melalui scraping (Oxylabs), lalu dipra-proses: penghapusan duplikat, pembersihan nilai kosong, ekstraksi fitur penting (contoh: breadcrumbs, features). Contoh kode preprocessing:

```
import pandas as pd
import os
import re

file_local_bersih = '/content/DATAFINAL_PREPROCESSED.csv'
file_input = '/content/DATAFILE/DATAFINAL - NOT CLEAN - stillduplicate_new.csv'

def extract_main_feature(feature_str):
    try:
        items = re.findall(r'["](.*?)["]', str(feature_str))
        if len(items) > 1:
            return items[1]
        Elif items:
            return items[0]
        return None
    Except:
        return None

try:
    print("--- Memulai Proses Baca Data ---")
    if not os.path.exists(file_input):
        raise FileNotFoundError(f"File input '{file_input}' tidak ditemukan. Pastikan path sudah benar.")
    df = pd.read_csv(file_input)
    print(f"Jumlah baris data awal: {len(df)}")

    columns_to_drop = ['url', 'asin', 'images_lis']
    columns_exist = [col for col in columns_to_drop if col in df.columns]
    df_processed = df.drop(columns=columns_exist)
    print(f"1. Kolom {columns_exist} telah dihapus.")

    If 'breadcrumb' in df_processed.columns:
        df_processed['breadcrumbs_clean'] = df_processed['breadcrumb'].str.replace('/', '>', regex=False)
        df_processed = df_processed.drop(columns=['breadcrumb'])
        print("2. Kolom 'breadcrumb' dibersihkan ('/' diganti dengan '>').")

    If 'features' in df_processed.columns:
        df_processed['features_clean'] = df_processed['features'].apply(extract_main_feature)
        df_processed = df_processed.drop(columns=['features'])
        print("3. Informasi diekstrak dari 'features' ke kolom baru 'features_clean'.")

    rows_before_na = len(df_processed)
    df_processed.dropna(inplace=True)
    print(f"4. Menghapus baris dengan nilai kosong. Baris berubah dari {rows_before_na} menjadi {len(df_processed)}.")

    rows_before_dup = len(df_processed)
    df_final = df_processed.drop_duplicates(subset=['title'], keep='first')
    print(f"5. Menghapus duplikat. Baris berubah dari {rows_before_dup} menjadi {len(df_final)}.")

    final_column_order = [
```

```

'title',
'brand',
'product_d',
'features_clean',
'breadcrumbs_clean',
'price'
]
final_column_order_exist = [col for col in final_column_order if col in df_final.columns]
df_final = df_final[final_column_order_exist]
print("6. Urutan kolom telah disesuaikan.")

output_dir = os.path.dirname(file_local_bersih)
if output_dir:
    os.makedirs(output_dir, exist_ok=True)
df_final.to_csv(file_local_bersih, index=False)
print(f"\n☑ SUKSES! Data yang sudah diproses berhasil disimpan ke '{file_local_bersih}'.")

Except FileNotFoundError as fnf_error:
    print(f"✘ Error: {fnf_error}")
Except Exception as e:
    print(f"✘ Terjadi kesalahan saat memproses file.")
    print(f"Error: {e}")

```

Tahapan embedding dilakukan menggunakan model **SigLIP** (google/siglip-so400m-patch14-384) untuk menghasilkan vektor dimensi 1152 dari data teks dan gambar. Agar representasi lebih robust, digunakan teknik chunking dan averaging pada data teks:

```

MAX_TOKENS = 60
OVERLAP = 10

def chunk_text_by_tokens(text, tokenizer, max_tokens=MAX_TOKENS, overlap=OVERLAP):
    input_ids = tokenizer.encode(text, add_special_tokens=False)
    chunks = []
    start = 0
    while start < len(input_ids):
        chunk_ids = input_ids[start:start+max_tokens]
        chunk_text = tokenizer.decode(chunk_ids, skip_special_tokens=True)
        chunks.append(chunk_text)
        start += max_tokens - overlap
    return chunks

```

Penyimpanan data dan vektor embedding menggunakan database PostgreSQL dengan ekstensi pgvector. Tabel utama dirancang sebagai berikut:

```

CREATE EXTENSION IF NOT EXISTS vector;
CREATE TABLE products (
    product_id SERIAL PRIMARY KEY,
    title TEXT,
    price TEXT,
    brand TEXT,
    product_details_clean TEXT,
    features_clean TEXT,

```

```

breadcrumbs_clean TEXT,
text_embedding vector(1152),
image_embedding vector(1152),
image_path TEXT,
combined_text TEXT
);
CREATE INDEX ON products USING hnsw (text_embedding vector_l2_ops);
CREATE INDEX ON products USING hnsw (image_embedding vector_l2_ops);

```

## 2. Increment 2: Implementasi Fitur Pencarian Baseline

Fitur pencarian baseline dikembangkan untuk menjadi titik perbandingan dengan fitur RAG Multimodal.

**Pencarian Kata Kunci** diimplementasikan menggunakan query SQL LIKE pada kolom title. Contoh kode:

```

like_query = "SELECT * FROM products WHERE LOWER(title) LIKE %s LIMIT 20"
search_term = f"%{{selected_shoe_name.lower()}}%"
cur.execute(like_query, (search_term,))

```

**Pencarian Semantik Teks** memanfaatkan embedding query teks (SigLIP) dan similarity search pada kolom text\_embedding.

**Pencarian Semantik Gambar** menggunakan embedding gambar sebagai query pada image\_embedding.

**Pencarian Multimodal** (late fusion): vektor query merupakan rata-rata antara text\_vector dan image\_vector.

## 3. Increment 3: Implementasi Fitur RAG Multimodal

Fitur Asisten Chat RAG dikembangkan dengan tiga tahap utama:

### Retrieval:

```

if text_vector is not None and image_vector is not None:
    query_vector = (text_vector + image_vector) / 2
    search_query = "SELECT * FROM products ORDER BY image_embedding <=> %s::vector
LIMIT 10;"
Elif image_vector is not None:
    query_vector = image_vector
    search_query = "SELECT * FROM products ORDER BY image_embedding <=> %s::vector
LIMIT 10;"
Elif text_vector is not None:
    query_vector = text_vector
    search_query = "SELECT * FROM products ORDER BY text_embedding <=> %s::vector LIMIT
10;"

```

**Augmentation:** Prompt builder secara otomatis menggabungkan data produk hasil retrieval (format JSON) dengan instruksi spesifik dan kueri asli ke dalam satu prompt terstruktur.

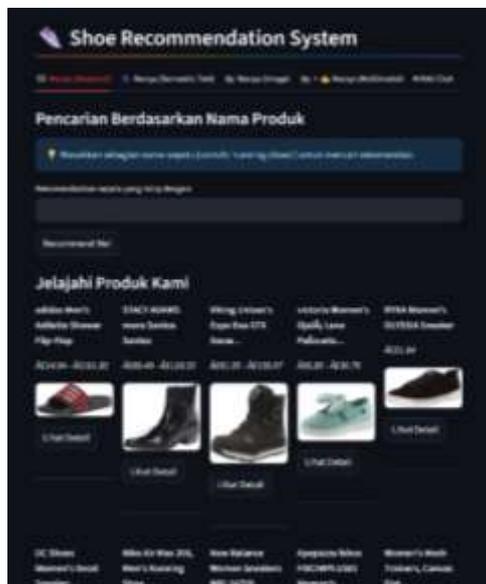
**Generation:** Prompt yang sudah diperkaya dikirim ke model Gemini 2.5 Pro melalui API Python. Respons berupa teks konsultatif dalam format JSON diolah dan ditampilkan ke pengguna melalui UI.

#### 4. Increment 4: Implementasi Antarmuka Pengguna

Aplikasi dikembangkan dengan framework Streamlit, terdiri dari lima tab utama: pencarian kata kunci, semantik teks, semantik gambar, pencarian multimodal, dan asisten chat RAG. Setiap tab terintegrasi dengan backend dan menampilkan hasil sesuai logika fitur terkait.

a. Recsys (Keyword): Pencarian Berdasarkan Nama Produk

Tab pertama memungkinkan pengguna untuk melakukan pencarian produk sepatu berdasarkan kata kunci atau nama spesifik.

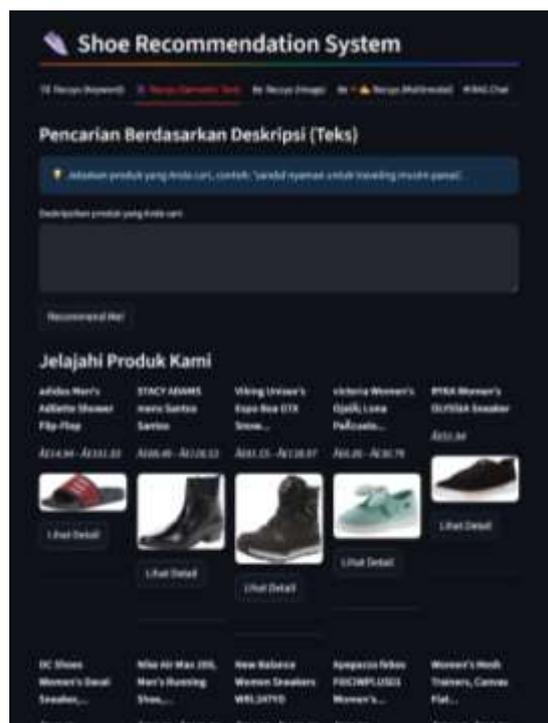


**Gambar 3.** Antarmuka Pencarian Kata Kunci.

Pada Gambar 3. Pengguna memasukkan nama produk yang diinginkan pada kolom input yang tersedia. Sistem kemudian akan menjalankan pencarian leksikal pada database untuk menemukan produk dengan nama yang cocok dan menampilkan hasilnya dalam format grid. Jika tidak ada input, halaman akan menampilkan produk-produk secara acak untuk dijelajahi.

b. Recsys (Semantic Text): Pencarian Berdasarkan Deskripsi

Tab kedua menyediakan pencarian semantik, di mana pengguna dapat memasukkan deskripsi tekstual mengenai sepatu yang mereka cari.

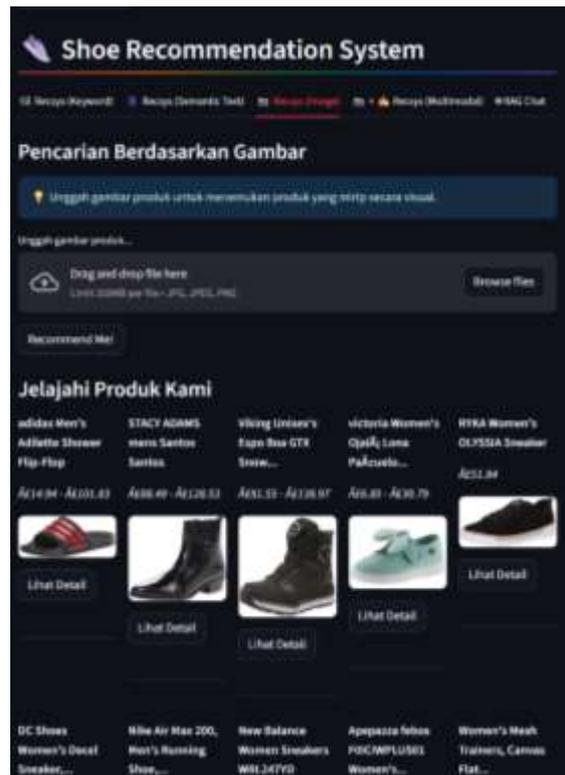


**Gambar 4.** Antarmuka Pencarian Semantik Teks

Pada Gambar 4. Pengguna dapat mengetikkan deskripsi konseptual, seperti "sandal nyaman untuk traveling musim panas". Sistem akan menganalisis makna dari deskripsi tersebut menggunakan model embedding untuk menemukan produk yang cocok secara kontekstual, tidak hanya berdasarkan kata kunci yang sama persis.

c. Reccsys (Image): Pencarian Berdasarkan Gambar

Pada tab ini, pengguna dapat mengunggah gambar produk sepatu untuk menemukan produk yang mirip secara visual.

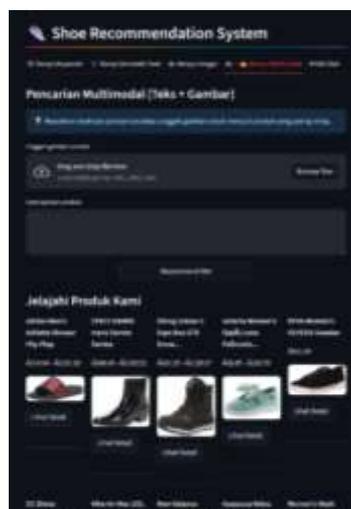


**Gambar 5.** Antarmuka Pencarian Semantik Gambar

Pada Gambar 5. sistem akan menggunakan model pengenalan gambar untuk menganalisis fitur visual dari sepatu dalam gambar yang diunggah. Setelah itu, sistem akan merekomendasikan produk lain dari database yang memiliki kemiripan visual paling tinggi.

d. Recsys (Multimodal): Pencarian Teks dan Gambar

Tab keempat mengimplementasikan pencarian multimodal yang menggabungkan input gambar dan teks untuk hasil yang lebih spesifik.

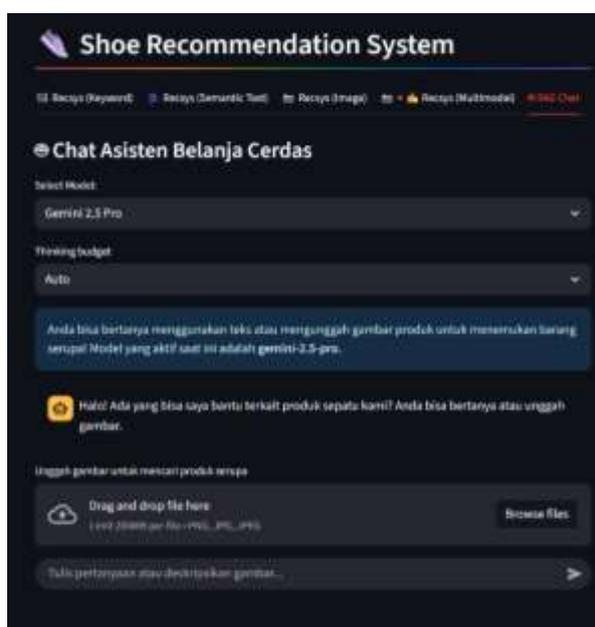


**Gambar 6.** Antarmuka Pencarian Multimodal

Pada Gambar 6. Pengguna dapat mengunggah gambar sepatu sambil memberikan deskripsi tambahan untuk memperjelas pencarian, misalnya mengunggah gambar sepatu lari dan menambahkan teks "untuk lari di medan berat". Sistem akan menggabungkan kedua informasi ini untuk memberikan hasil yang lebih akurat.

e. RAG Chat: Asisten Belanja Cerdas

Tab terakhir adalah antarmuka percakapan (chat) dengan asisten belanja cerdas yang didukung oleh model bahasa canggih (Gemini).



**Gambar 7.** Antarmuka Asisten Chat RAG

Pada Gambar 7. pengguna dapat berinteraksi secara natural, mengajukan pertanyaan, meminta rekomendasi, atau bahkan mengunggah gambar produk dalam percakapan untuk menemukan barang serupa. Sistem RAG akan memberikan jawaban yang konsultatif dan informatif, sehingga menciptakan pengalaman belanja yang lebih interaktif dan personal.

Implementasi sistem menggunakan metodologi *Incremental Model* berhasil menghasilkan sebuah aplikasi web fungsional dengan lima antarmuka pencarian utama yang diakses melalui tab terpisah. Pertama, fitur pencarian berbasis kata kunci (*Recsys Keyword*) yang memungkinkan pengguna menemukan produk melalui pencocokan leksikal. Kedua, fitur pencarian semantik teks (*Recsys Semantic Text*) yang dirancang untuk memahami deskripsi konseptual pengguna. Ketiga, fitur pencarian semantik gambar (*Recsys Image*) yang memungkinkan

pencarian produk berdasarkan kemiripan visual. Keempat, fitur pencarian multimodal (*Recsys Multimodal*) yang menggabungkan input teks dan gambar untuk hasil yang lebih spesifik. Kelima, fitur inti penelitian, yaitu Asisten Belanja Cerdas (*RAG Chat*), yang menyediakan antarmuka percakapan untuk interaksi yang konsultatif dan personal.

**Tabel 1.** Pengujian Fungsional

<b>Skenario Pengujian</b>	<b>Langkah-langkah Pengujian</b>	<b>Hasil yang Diharapkan</b>	<b>Hasil Aktual</b>	<b>Status</b>
Fungsionalitas Pencarian Kata Kunci	1. Buka tab "Recsys (Keyword)". 2. Masukkan kueri "adidas shoes". 3. Tekan tombol "Recommend Me!".	Sistem menampilkan daftar produk yang judulnya mengandung kata "adidas shoes".	Sistem berhasil menampilkan produk yang relevan dengan kata kunci.	Lulus
Fungsionalitas Pencarian Semantik Teks	1. Buka tab "Recsys (Semantic Text)". 2. Masukkan kueri "sepatu hiking tahan air". 3. Tekan tombol "Recommend Me!".	Sistem menampilkan daftar produk sepatu hiking atau sepatu outdoor yang relevan secara semantik.	Sistem berhasil menampilkan produk yang relevan secara kontekstual.	Lulus
Fungsionalitas Pencarian Semantik Gambar	1. Buka tab "Recsys (Image)". 2. Unggah gambar sepatu boots. 3. Tekan tombol "Recommend Me!".	Sistem menampilkan daftar produk sepatu boots lain yang mirip secara visual.	Sistem berhasil menampilkan produk yang relevan secara visual.	Lulus
Fungsionalitas Asisten Chat RAG	1. Buka tab "RAG Chat". 2. Masukkan kueri "sepatu lari untuk maraton". 3. Tekan tombol kirim.	Sistem memberikan respons percakapan yang berisi justifikasi dan rekomendasi produk yang relevan.	Sistem berhasil memberikan respons percakapan yang koheren dan relevan.	Lulus

Setelah implementasi, dilakukan pengujian fungsional dengan metode *black box testing* untuk memvalidasi stabilitas teknis aplikasi. Skenario pengujian mencakup semua alur kerja utama, mulai dari pencarian kata kunci sederhana ("adidas shoes") hingga kueri RAG yang kompleks ("sepatu lari untuk maraton"). Hasil pengujian menunjukkan bahwa semua fungsionalitas sistem berjalan sesuai dengan spesifikasi yang diharapkan dan dinyatakan **lulus**. Dengan demikian, aplikasi divalidasi sebagai alat yang andal untuk tahap evaluasi kualitatif dengan pengguna.

### **Hasil Evaluasi Kualitatif**

Setelah validasi teknis, evaluasi kualitatif dengan 5 partisipan menghasilkan temuan yang kaya mengenai pengalaman pengguna. Analisis tematik terhadap transkrip *Think-Aloud Protocol* dan wawancara mendalam memunculkan tiga tema utama yang secara konsisten membedakan efektivitas masing-masing fitur pencarian.

- 1. Tema 1: Kegagalan Pencarian Leksikal pada Niat Konseptual.** Tema ini muncul secara dominan saat partisipan dihadapkan pada Skenario 2 (mencari 'sepatu hiking yang bagus untuk pemula dan tahan air'). Partisipan yang menggunakan fitur Pencarian Kata Kunci mengalami frustrasi karena sistem gagal memahami maksud di balik kueri mereka. Fitur ini hanya mampu melakukan pencocokan literal, sehingga tidak menghasilkan output yang relevan. Seorang partisipan menyatakan, "*Saya ketik 'sepatu buat lari jauh', malah tidak keluar apa-apa. Sistemnya tidak mengerti maksud saya 'lari jauh' itu butuh bantalan lebih.*" Temuan ini menegaskan bahwa pencarian leksikal hanya efektif untuk Skenario 1, di mana pengguna memiliki niat yang sangat spesifik dan mengetahui nama produk secara pasti.
- 2. Tema 2: Pencarian Semantik sebagai Jembatan Intuitif.** Pada Skenario 2 dan Skenario 3 (pencarian visual), partisipan menunjukkan reaksi positif yang signifikan saat beralih ke fitur Pencarian Semantik. Mereka merasa sistem mampu "mengerti" niat mereka di luar kata kunci literal. Pada Skenario 3, seorang partisipan berkomentar, "*Wow, ini keren. Saya unggah gambar sepatu boots coklat, dan yang muncul benar-benar boots lain dengan model yang mirip, bukan cuma sepatu coklat biasa.*" Namun, batasan yang teridentifikasi adalah meskipun hasilnya relevan secara semantik, partisipan masih dihadapkan pada daftar panjang produk yang perlu

mereka saring dan bandingkan secara manual, yang masih menyisakan beban kognitif.

**3. Tema 3: Asisten RAG sebagai Puncak Pengalaman Konsultatif.** Seluruh partisipan (5 dari 5) secara eksplisit menyatakan bahwa fitur Asisten Chat RAG adalah yang paling mereka sukai dan paling bermanfaat, terutama pada Skenario 4 (multimodal) dan Skenario 5 (ambigu & konsultatif). Fitur ini berhasil mengubah interaksi dari "mencari" menjadi "berkonsultasi". Dua sub-tema utama yang mendukung hal ini adalah:

- a. **Sintesis Informasi Kompleks:** Pada Skenario 4 dan 5, fitur RAG menunjukkan kemampuannya untuk memproses dan mensintesis beberapa batasan (visual, fungsional, gaya, anggaran) menjadi satu set rekomendasi yang koheren dan dipersonalisasi.
- b. **Pembangunan Kepercayaan melalui Justifikasi:** Alasan utama preferensi terhadap RAG adalah adanya justifikasi dalam responsnya. Penjelasan *mengapa* sebuah produk direkomendasikan membuat partisipan merasa lebih yakin dengan pilihan mereka. Seorang partisipan menyimpulkannya pada Skenario 5: *"Yang chat ini paling bagus. Rasanya seperti ngobrol sama teman yang jago sepatu. Dia tidak cuma kasih daftar, tapi juga kasih tahu kenapa sepatu A lebih bagus buat lari maraton daripada sepatu B."*

### **Pembahasan**

Temuan penelitian ini secara konkret menunjukkan bahwa arsitektur RAG Multimodal memberikan nilai tambah yang signifikan pada pengalaman pengguna, melampaui kemampuan sistem pencarian konvensional. Kegagalan pencarian kata kunci pada niat konseptual menegaskan kembali keterbatasan fundamental dari pencocokan leksikal, yang menyebabkan "friksi pencarian" dan meningkatkan beban kognitif ekstrinsik pengguna, sejalan dengan Teori Beban Kognitif (Li et al., 2025). Disisi lain, pencarian semantik, meskipun merupakan kemajuan signifikan, pada dasarnya hanya menyelesaikan tahap *retrieval* menyajikan kandidat yang relevan namun masih menyerahkan tugas berat untuk melakukan sintesis, perbandingan, dan pengambilan keputusan sepenuhnya kepada pengguna.

Keunggulan fitur Asisten Chat RAG dapat dijelaskan melalui alur kerja arsitekturalnya yang lengkap. Kemampuan sistem untuk pertama-tama mengambil

(*retrieve*) sekumpulan produk yang relevan, lalu memperkaya (*augment*) kueri pengguna dengan konteks produk tersebut, dan terakhir menghasilkan (*generate*) respons yang beralasan, secara efektif mengurangi beban kognitif pengguna. Proses inilah yang mengubah interaksi menjadi dialog konsultatif, sejalan dengan teori sistem rekomendasi percakapan (Jannach, 2021).

Pemberian justifikasi oleh model Gemini 2.5 Pro menjadi faktor krusial dalam membangun kepercayaan, sebuah aspek penting dalam interaksi manusia-AI (Bai et al., 2020). Dengan mereplikasi peran seorang asisten penjualan yang berpengalaman, sistem RAG tidak hanya menyediakan daftar produk, tetapi juga "mendidik" pengguna tentang pilihan mereka. Hal ini terbukti sangat berharga dalam skenario eksplorasi di mana pengguna tidak memiliki niat yang jelas atau membutuhkan panduan ahli, sejalan dengan prinsip kewirausahaan yang berfokus pada pemahaman mendalam dan pemenuhan kebutuhan pelanggan secara holistik (Sari, 2023).

## **SIMPULAN**

Berdasarkan implementasi sistem dan analisis hasil evaluasi kualitatif, penelitian ini menghasilkan tiga kesimpulan utama yang secara langsung menjawab rumusan masalah.

1. Pertama, sistem RAG Multimodal yang fungsional berhasil dirancang dan diimplementasikan dengan mengintegrasikan *framework* Streamlit, model *embedding* SigLIP, basis data PostgreSQL dengan pgvector, dan model generatif Gemini 2.5 Pro. Arsitektur ini terbukti valid dan andal dalam mendukung empat kapabilitas pencarian yang berbeda.
2. Kedua, pengalaman pengguna saat berinteraksi dengan sistem RAG *Multimodal* secara signifikan lebih superior dibandingkan dengan metode pencarian tradisional. Temuan kualitatif menunjukkan bahwa fitur RAG secara konsisten dinilai paling bermanfaat dan disukai oleh seluruh partisipan karena kemampuannya memberikan justifikasi dan rekomendasi yang dipersonalisasi, mengubah interaksi menjadi sebuah dialog konsultatif.
3. Ketiga, sistem RAG *Multimodal* memberikan nilai tambah yang paling signifikan dalam skenario pencarian yang ditandai oleh ambiguitas, kompleksitas, dan kebutuhan eksplorasi. Dalam situasi di mana pengguna tidak yakin dengan apa yang mereka cari atau memiliki kebutuhan yang rumit, kemampuan RAG untuk

mensintesis informasi dan memberikan respons yang beralasan menjadi pembeda utama.

4. Untuk penelitian di masa depan, disarankan untuk melakukan ekspansi ke domain produk yang lebih beragam, seperti otomotif atau olahraga, untuk menguji generalisasi sistem. Selain itu, perlu dipertimbangkan penggunaan arsitektur dan teknologi skala produksi (misalnya, FastAPI untuk *backend* dan basis data vektor terdistribusi seperti Pinecone) untuk mengevaluasi kelayakan sistem pada skala yang lebih besar.

### UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Bapak Mukhammad Andri Setiawan, S.T., M.Sc., Ph.D., selaku dosen pembimbing atas arahan dan bimbingannya. Terima kasih juga kepada seluruh partisipan yang telah meluangkan waktu dan memberikan wawasan berharga dalam sesi evaluasi.

### DAFTAR PUSTAKA

- Bach, T. A., Khan, A., Hallock, H., Beltrão, G., & Sousa, S. (2024). A Systematic Literature Review of User Trust in AI-Enabled Systems: An HCI Perspective. *International Journal of Human-Computer Interaction*, 40(5), 1251–1266. <https://doi.org/10.1080/10447318.2022.2138826>
- Bai, J., Chen, M., Xiaosheng, J. L., Daniel, M., Xu, Y., Akolakis, C., Allen, T., Atkin, D., Bandiera, O., Banerjee, A., Bergquist, L., Chor, D., Donaldson, D., Duflo, E., Faber, B., Hanson, G., Hsieh, C.-T., Jia, P., Khan-Delwal, A., ... Xu, D. Y. (2020). *Search and Information Frictions on Global E-Commerce Platforms: Evidence from AliExpress*. <https://doi.org/10.3386/W28100>
- Erlina F. Santika. (2024). *Tingkat Penetrasi Internet Indonesia Capai 79,5% per 2024*. <https://databoks.katadata.co.id/teknologi-telekomunikasi/statistik/e6f9d69e252de32/tingkat-penetrasi-internet-indonesia-capai-795-per-2024>
- Ganis Sanhaji, K., & Irsyaad Hizbullah, A. (2024). ) 2024 | 234 Edusaintek: Jurnal Pendidikan. *Sains Dan Teknologi*, 11(1), 2024–2234. <https://doi.org/10.47668/edusaintek.v11i1.999>
- Google, T. and B. & C. (2025). *e-Conomy SEA 2020*. <https://www.temasek.com.sg/content/dam/temasek-corporate/news-and-views/resources/reports/e-conomy-sea-2020-report.pdf>
- Gupta, S., Ranjan, R., & Singh, S. N. (2024). *A Comprehensive Survey of Retrieval-Augmented Generation (RAG): Evolution, Current Landscape, and Future Directions*. <https://arxiv.org/pdf/2410.12837>
- Jannach, D. (2021). A Survey on Conversational Recommender Systems. *ACM Comput. Surv*, 54.

- Jovanović, M., Voss, P., Ai, A., & Tx, A. (2024). *Towards Incremental Learning in Large Language Models: A Critical Review*. <https://arxiv.org/pdf/2404.18311>
- Li, Z., Wang, B., Prince, O., Le Texier, T., & Addai-Dansoh, S. (2025). *EXPLORING PERCEPTION OF PLATFORM LANGUAGE ON CONSUMER DECISION MAKING: A COGNITIVE LOAD PERSPECTIVE* \*.
- Liang, W., Zhang, Y., Kwon, Y., Yeung, S., & Zou, J. (2022). Mind the Gap: Understanding the Modality Gap in Multi-modal Contrastive Representation Learning. *Advances in Neural Information Processing Systems*, 35. <https://arxiv.org/pdf/2203.02053>
- Mortaheb, M., Amir, M. A., Khojastepour, ), Chakradhar, S. T., & Ulukus, S. (2025). *RAG-Check: Evaluating Multimodal Retrieval Augmented Generation Performance*. <https://arxiv.org/pdf/2501.03995>
- Pratiwi, K. D. (2022). E-COMMERCE AND ECONOMIC GROWTH IN INDONESIA: ANALYSIS OF PANEL DATA REGRESSION. *GEMA PUBLICA*, 7(1), 171–186. <https://doi.org/10.14710/GP.7.1.2022.171-186>
- Rokon, O. F., Simion, A., Du, W., Wen, M., Yao, H., & Lee, K.-C. (2024). *Enhancement of E-commerce Sponsored Search Relevancy with LLM*.
- Tizar, M., & Azizah, N. (2023). RANCANG BANGUN SISTEM INFORMASI PENJUALAN BERBASIS WEB ( E-COMMERCE) PADA TOKO RUMAH POPOK KINAN. *EDUSAINTEK: Jurnal Pendidikan, Sains Dan Teknologi*, 10(1), 154–170. <https://doi.org/10.47668/EDUSAINTEK.V10I1.664>