DIGIVENTURE
Journal of Applied Digital Business Research

Vol. 2, No. 1 (2026), hlm. 210-245
e-ISSN: 3123-5212
Situs web: https://journal.unesa.ac.id/digiventure

Prototipe *AI Agent* Layanan Konsultasi Berbasis WhatsApp untuk Rekomendasi Laptop pada Platform n8n

Archam Zahidin[1] dan Anita Safitri[2]

[1,2] Bisnis Digital, Universitas Negeri Surabaya, Indonesia

*) Corresponding Author (Email: archam.22070@mhs.unesa.ac.id)

ABSTRACT

This study develops and evaluates a prototype WhatsApp-based AI Agent consultation service that delivers laptop recommendations, orchestrated on the n8n platform. Following a research-and-development approach, the system was designed as a conversational recommender system that separates natural-language understanding from factual product determination: a customer-service agent handles need elicitation and dialogue control, while a recommendation agent composes explanations over candidates produced by a deterministic subworkflow. Recommendations are grounded in a laptop catalog stored in a database, combining constraint-based filtering with aspiration-level utility ranking so that outputs remain faithful to available products. Evaluation combined grey-box testing of functional behavior with a qualitative user study involving purposively selected informants interviewed through an adapted conversational-recommender questionnaire. The technical tests confirmed correct dialogue control, filtering, ranking, and catalog fidelity with no fabricated products, while users reported positive perceptions of accuracy, ease of use, transparency, and trust. The study contributes a traceable architecture that separates conversational language from deterministic recommendation, yielding catalog-grounded novelty for pre-purchase consultation in conversational commerce.

Keywords: *conversational recommender system, AI agent, WhatsApp, laptop recommendation, n8n*

I. PENDAHULUAN

Perkembangan kecerdasan buatan (AI) semakin mengubah pola interaksi antara perusahaan dan konsumen, terutama melalui kemunculan AI agent yang mampu menjalankan tugas, berinteraksi dengan pengguna, dan bertindak secara otonom (Trzebiński et al., 2025). Sejumlah studi menunjukkan bahwa integrasi AI dalam layanan meningkatkan efisiensi, ketersediaan informasi, serta pengalaman pelanggan pada berbagai konteks bisnis (Abdelhalim et al., 2024; Mari et al., 2024).

Copyright © 2026 by Authors; this is an open-access article under the CC BY-SA License (https://creativecommons.org/licenses/by-sa/4.0).

Fenomena ini mendorong pengembangan sistem yang tidak sekadar merekomendasikan produk, tetapi juga mampu memberikan konsultasi interaktif secara real-time menyerupai interaksi manusiawi untuk membangun kepercayaan pengguna (Mari et al., 2024). Tren tersebut diperkuat laporan McKinsey State of AI 2025: proporsi organisasi yang secara reguler menggunakan AI pada setidaknya satu fungsi bisnis mencapai 88% (naik dari 78% pada tahun sebelumnya), sementara adopsi generative AI meningkat dari 33% (2023) menjadi 79% (2025).

Layanan berbasis AI berpotensi mengurangi beban tenaga kerja manusia dan mendorong keputusan pembelian melalui rekomendasi terpersonalisasi (Ahn et al., 2022; Kim et al., 2023), sehingga AI agent menarik dikembangkan sebagai layanan konsultasi yang membantu pengguna mempersempit pilihan. Meski demikian, resistensi konsumen terhadap rekomendasi AI masih terjadi akibat bias, keterbatasan empati, maupun persepsi bahwa AI belum sepenuhnya menggantikan sentuhan manusia (Frank et al., 2024; Shaalan et al., 2024), sehingga pengembangan agen yang lebih kontekstual dan adaptif tetap relevan.

Di sisi lain, AI generatif yang mengandalkan kemampuan generatif cenderung menghasilkan jawaban yang terdengar meyakinkan tetapi berisiko tidak faktual atau tidak dapat ditelusuri. Risiko *hallucination* pada large language models telah banyak dibahas sebagai tantangan reliabilitas pada implementasi nyata (L. Huang et al., 2023). Risiko ini menjadi krusial pada konteks rekomendasi produk, karena rekomendasi yang tidak valid atau salah menyebutkan spesifikasi dapat menyesatkan calon pembeli. Beberapa studi menunjukkan bahwa penguatan model dengan basis pengetahuan eksternal (katalog) melalui pendekatan retrieval-augmented atau pemanfaatan tools dapat meningkatkan keterlacakan dan menekan kesalahan informasi dibanding sekadar mengandalkan pengetahuan parametris model (Gao et al., 2023; Lewis et al., 2020).

Sebagai penguat dasar empiris, dilakukan survei pendahuluan berupa wawancara semi-terstruktur terhadap lima calon pengguna yang dipilih secara *purposive* pada tahap pra-pembelian (tahap interest dalam Model AIDA); jumlah ini lazim pada studi eksploratif karena tema utama umumnya telah menjenuh setelah sekitar lima wawancara pada populasi yang relatif homogen (Guest et al., 2006). Hasilnya menunjukkan proses pencarian laptop belum efisien, yakni partisipan berpindah antar-marketplace, menelusuri ulasan di YouTube maupun TikTok, dan membandingkan spesifikasi secara manual, sehingga kerap memicu *choice overload*. Seluruh partisipan menilai konsultasi berbasis percakapan relevan dengan kebutuhan mereka, dengan ekspektasi sistem menanyakan kebutuhan dan anggaran di awal, menyesuaikan preferensi selama interaksi, serta menyertakan alasan pada tiap rekomendasi. Kekhawatiran terbesar justru berkaitan dengan kemungkinan produk yang direkomendasikan tidak benar-benar tersedia atau harganya tidak sesuai, temuan yang memperkuat urgensi *grounding* rekomendasi terhadap katalog dan menegaskan WhatsApp sebagai kanal berhambatan rendah. Temuan-temuan ini secara langsung memandu keputusan rancangan, yakni menyediakan konsultasi percakapan terpusat yang mempersempit pilihan dan menyertakan alasan pada tiap

rekomendasi, menerapkan alur dialog adaptif yang menggali kebutuhan dan anggaran lebih dulu, menambatkan (*grounding*) rekomendasi pada katalog produk, serta memakai WhatsApp sebagai antarmuka berhambatan rendah.

Penelaahan terhadap penelitian terdahulu memperlihatkan beberapa celah yang berkaitan langsung dengan konteks ini. Pada lingkup AI agent untuk *customer service*, sebagian besar penelitian masih bersifat eksperimental atau berbasis skenario pada lingkungan terbatas dan lebih menekankan aspek psikologis pengguna, seperti anthropomorphism, kepercayaan, dan adoption *intent* (Adam et al., 2021; Araujo, 2018; Sheehan et al., 2020), sehingga pengujian implementasi sistem konsultasi secara utuh (*end-to-end*) pada kondisi nyata masih jarang. Pada lingkup *conversational recommender system* dan pemanfaatan LLM, validitas keluaran terhadap katalog belum sepenuhnya terselesaikan; model masih berpotensi merekomendasikan produk yang tidak tersedia dalam inventaris (Granada et al., 2023), sementara evaluasi tidak dapat hanya bertumpu pada metrik otomatis karena korelasinya rendah terhadap persepsi pengguna (Jannach, 2023; Manzoor et al., 2024). Adapun kajian *grey-box testing* sejauh ini cenderung konseptual dan deskriptif serta minim pembuktian empiris pada studi kasus nyata (Coulter, 1999; Ehmer & Khan, 2012).

Merujuk pemetaan tersebut, belum terdapat penelitian yang memadukan secara terintegrasi antara agen konsultasi pra-pembelian berbasis percakapan, rekomendasi produk laptop yang tervalidasi terhadap katalog, pemanfaatan WhatsApp sebagai antarmuka, dan orkestrasi low-code melalui n8n, yang pada saat bersamaan dievaluasi menggunakan *grey-box testing* dan uji pengguna. Dengan demikian, kebaruan penelitian ini tidak terletak pada satu komponen tunggal, melainkan pada pemaduannya menjadi satu prototipe utuh yang diuji secara empiris pada kondisi penggunaan nyata.

Pemilihan komponen berpijak pada pertimbangan berikut. Konsultasi pra-pembelian melalui aplikasi pesan makin relevan dalam praktik social commerce, di mana relasi sosial dan *trust* memengaruhi keputusan belanja (Cao et al., 2021; Y. Huang et al., 2022; Nyazabe et al., 2025). Kanal dipilih berdasarkan kebiasaan pengguna: di Indonesia, WhatsApp digunakan oleh 91,7% pengguna internet, jauh melampaui platform lain seperti Instagram (84,6%), Facebook (83,0%), dan TikTok (77,4%), sehingga menjadi antarmuka berhambatan rendah (Maddalena, 2025).

Laptop dipilih sebagai domain karena tergolong complex buying behaviour, yaitu keterlibatan konsumen tinggi pada produk yang relatif mahal, berisiko, dan kaya atribut (Xhema, 2019), dengan banyak kriteria keputusan seperti CPU, RAM, penyimpanan, kartu grafis, layar, hingga harga (Mochammad Rafi Diaz Ardhana & Aries Dwi Indriyanti, 2025). Untuk menghubungkan kanal percakapan, logika konsultasi, akses katalog, dan pemanggilan layanan AI dalam satu alur kerja, penelitian ini memakai n8n karena bersifat open-source, low-code, *node*-based, mendukung integrasi API, dan dapat self-hosted (McFeetors & Pant, 2022). Evaluasi menggunakan *grey-box testing* karena prototipe bersifat integratif dan melibatkan komponen pihak ketiga, sehingga inspeksi internal menyeluruh (*white-box*) tidak

memungkinkan sementara *black-box* berisiko melewatkan kesalahan alur data (Ehmer & Khan, 2012; Myers et al., 2012). Kerangka AIDA dipakai semata untuk menetapkan kriteria partisipan pada tahap interest, tempat *choice overload* lazim muncul (Chernev et al., 2015; Iwamoto, 2023).

Penelitian ini bertujuan merancang, mengimplementasikan, dan mengevaluasi prototipe AI Agent layanan konsultasi berbasis WhatsApp untuk rekomendasi laptop yang diorkestrasi n8n, menilai kinerjanya dalam menghasilkan rekomendasi yang valid, relevan, dan sesuai kebutuhan pengguna, serta mengidentifikasi kelebihan, keterbatasan, dan peluang pengembangannya melalui pengujian fungsionalitas dan uji pengguna dengan pendekatan *grey-box testing*.

II. TINJAUAN PUSTAKA

Landasan Konseptual

Model AIDA dan Tahap Interest

Model AIDA (Attention, Interest, Desire, Action) adalah model klasik komunikasi pemasaran bertipe hierarchy-of-effects yang memandang respons konsumen sebagai rangkaian tahap kognitif, afektif, hingga perilaku sebelum tindakan, dari attention, interest, desire, sampai action (Barry, 1987; Barry & Howard, 1990). Meski kerap dikritik karena tidak selalu linear, AIDA berguna sebagai kerangka praktis pemetaan posisi konsumen pra-pembelian, dan penelitian ini memakainya sebagai lensa perancangan dengan fokus pada tahap interest, yaitu fase ketika seseorang telah tertarik namun preferensinya belum matang. Karakteristik ini melandasi keputusan rancangan berupa penggalian kebutuhan bertahap alih-alih menuntut spesifikasi teknis di awal, penyempitan pilihan melalui penyaringan dan pemeringkatan untuk menekan *choice overload*, dan penyertaan penjelasan pada tiap rekomendasi, mengingat banyaknya pilihan pada produk kompleks terbukti memperberat proses memilih dan menurunkan kualitas keputusan (Chernev et al., 2015).

AI Agent dan *Large Language Model*

AI agent adalah entitas komputasi yang mengindra lingkungannya, mengambil keputusan, dan bertindak secara otonom untuk mencapai tujuan tertentu, dan kemajuan *large language model* (LLM) mendorong kemunculan LLM-based agent yang memakai LLM sebagai pusat penalaran untuk memahami instruksi, menyusun rencana, serta memanggil perkakas dan sumber pengetahuan eksternal, dengan arsitektur lazim berupa otak sebagai pusat penalaran, persepsi untuk menerima masukan, dan aksi untuk menghasilkan keluaran (L. Wang et al., 2024; Xi et al., 2025). LLM sendiri adalah model bahasa berskala besar yang dilatih pada korpus teks masif, umumnya di atas arsitektur Transformer dengan mekanisme attention (Vaswani et al., 2023); ketika skalanya diperbesar, LLM memperoleh kemampuan few-shot learning yang menyelesaikan beragam tugas hanya dari instruksi dan sedikit contoh (Brown et al., 2020), dan kemampuan generalisasi inilah yang

menjadikannya fondasi sistem percakapan dan agen cerdas yang menggali kebutuhan pengguna serta menyusun rekomendasi (Zhao et al., 2023).

Grounding dan *Hallucination*

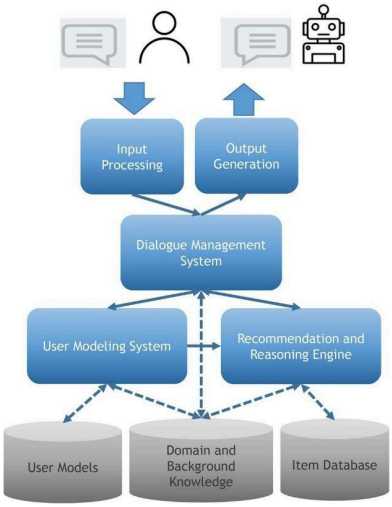
Kefasihan LLM menyimpan paradoks: model yang mampu merangkai kalimat meyakinkan juga dapat menghasilkan *hallucination*, yaitu keluaran yang terdengar sahih tetapi tidak berpijak pada fakta atau sumber yang dapat ditelusuri, baik bersifat intrinsic yang bertentangan dengan masukan maupun extrinsic yang tidak dapat diverifikasi (Ji et al., 2023); pada rekomendasi produk risikonya berwujud produk fiktif, spesifikasi keliru, atau klaim harga tanpa dasar yang meruntuhkan kepercayaan. *Grounding* menjawab persoalan ini dengan menautkan keluaran model pada sumber pengetahuan eksternal tepercaya, sebagaimana *Retrieval-Augmented Generation* (RAG) yang memadukan kemampuan generatif dengan pengambilan informasi non-parametrik (Lewis et al., 2020) dan terbukti menaikkan akurasi sekaligus menekan halusinasi (Gao et al., 2023). Penelitian ini menempatkan *grounding* sebagai prinsip inti dalam dua lapis, yaitu *grounding* faktual yang mewajibkan tiap produk berasal dari katalog, dan *grounding* keputusan yang mengalihkan penyaringan dan pemeringkatan ke lapis komputasi deterministik sehingga LLM hanya memverbalkan hasil, dan ruang mengarang produk atau mengubah peringkat ditutup pada tingkat arsitektur, bukan ditambal lewat penyempurnaan *prompt*.

Conversational Recommender System (CRS)

Conversational Recommender System (CRS) adalah sistem rekomendasi yang menyampaikan saran item melalui interaksi percakapan multi-giliran (multi-turn) dengan pengguna (Jannach, 2023). Berbeda dengan sistem rekomendasi konvensional yang umumnya bersifat one-shot, CRS memungkinkan interaksi dinamis: sistem menggali preferensi (preference *elicitation*) melalui pertanyaan, sedangkan pengguna dapat memberi umpan balik, meminta alternatif atau penjelasan, dan memperbaiki preferensinya sepanjang percakapan. Cara kerja ini relevan ketika preferensi pengguna belum jelas dan perlu dibentuk bertahap melalui dialog.

Sebagaimana ditunjukkan pada Gambar 1, CRS umumnya tersusun atas *input* processing untuk memahami bahasa dan *intent* pengguna, dialogue management system yang menentukan langkah percakapan, user modeling system yang memperbarui preferensi, recommendation and reasoning engine yang memeringkat item, serta item database sebagai katalog sumber fakta (Jannach, 2023). Survei atas 121 studi menempatkan dialogue management system sebagai pengendali utama alur percakapan dengan strategi interaksi beragam, dari berbasis formulir, berbasis kritik, hingga dialog berbahasa alami (Jannach et al., 2022), sedangkan pemaduan preference *elicitation* dengan pembelajaran berkelanjutan efektif mengatasi cold-start (Christakopoulou et al., 2016). Kehadiran LLM menggeser arah pengembangan menuju LLM-based CRS (Peng et al., 2025), dan model seperti ChatGPT dapat berperan sebagai CRS yang dinilai lebih bermakna (Manzoor et al.,

2024), meski keluarannya perlu dibatasi pada kandidat tepercaya (Granada et al., 2023) melalui *grounding* katalog, sehingga evaluasinya perlu holistik mencakup akurasi, kualitas dialog, dan *usability* (Jannach, 2023).



Sumber: Jannach, 2023

Gambar 1. Arsitektur *Conversational Recommender System*

***Multi-Criteria Decision Making* (MCDM)**

Sistem rekomendasi dikembangkan melalui beberapa paradigma, yaitu collaborative *filtering* yang bertumpu pada riwayat interaksi, content-based *filtering* yang bertumpu pada kemiripan atribut, dan pendekatan berbasis pengetahuan (knowledge-based) yang memanfaatkan pengetahuan domain (Burke, 2002); pendekatan berbasis pengetahuan unggul pada item kompleks berbanyak atribut, tidak terdampak cold-start, dan lebih transparan (Uta et al., 2024). Pendekatan ini bekerja melalui dua mekanisme yang saling melengkapi, yaitu penyaringan berbasis batasan (*constraint-based filtering*) yang memetakan kebutuhan menjadi batasan atas properti item lalu menyaring item yang memenuhi seluruh batasan wajib, dan pemeringkatan berbasis utilitas (utility-based *ranking*) yang menilai serta mengurutkan item lolos menurut kesesuaian preferensi (Felfernig & Burke, 2008). Pemeringkatan kerap dilandasi *Multi-Attribute Utility Theory* (MAUT) yang memodelkan utilitas sebagai gabungan terbobot beberapa kriteria (Winterfeldt & Edwards, 1993), dengan bentuk aditif *Simple Additive Weighting* (SAW) yang menjumlahkan nilai ternormalisasi (Taherdoost & Madanchian, 2023); karena rentang antar-atribut berbeda, normalisasi wajib dilakukan (Vafaei et al., 2022), dan penelitian ini memakai normalisasi berbasis sasaran terhadap *aspiration level* dari sebaran data (Jahan et al., 2012; Wierzbicki, 1982) sesuai prinsip satisficing dan *bounded rationality* (Simon, 1955). Karena skor tiap alternatif dihitung independen, pemeringkatan kebal terhadap rank reversal (Y.-M. Wang & Luo, 2009) dan kestabilannya diuji melalui analisis sensitivitas (Triantaphyllou & Sanchez, 1997).

n8n

n8n adalah platform otomatisasi alur kerja (*workflow automation*) berbasis low-code yang bersifat sumber terbuka (fair-code) dan dapat diinang secara mandiri (self-hosted). Platform ini memungkinkan perancangan alur secara visual berbasis *node* untuk menghubungkan beragam layanan serta mengorkestrasi langkah logika, termasuk pemanggilan LLM dan integrasi API eksternal (McFeetors & Pant, 2022). Pada penelitian ini, n8n dimanfaatkan sebagai lapisan orkestrasi yang menghubungkan kanal percakapan, logika pemrosesan, dan akses data katalog.

Metode Pengujian Perangkat Lunak

Pengujian perangkat lunak umumnya menempuh tiga pendekatan. *Black-box* testing menilai sistem dari hubungan *input-output* tanpa memperhitungkan struktur internal sehingga cocok pada level fitur tetapi berpotensi melewatkan jalur logika (Ehmer & Khan, 2012; Myers et al., 2012), sedangkan *White-box* testing berfokus pada struktur internal sehingga mengungkap kesalahan implementasi tetapi menuntut akses penuh atas kode (Myers et al., 2012). *Grey-box testing* berada di antara keduanya dengan memanfaatkan pengetahuan internal parsial seperti arsitektur, aliran data, dan state untuk merancang skenario uji lebih terarah sambil mengeksekusi melalui antarmuka yang tersedia (Beizer & Wiley, 1996), sehingga menilai bukan hanya keluaran akhir melainkan juga alur proses internal yang dapat diamati dan memudahkan penelusuran kesalahan integrasi (Pressman, 2001). *Grey-box* dipilih karena artefak berupa prototipe integratif dari orkestrasi n8n, kanal percakapan, dan layanan eksternal yang sebagian mekanismenya berada pada komponen pihak ketiga sehingga tidak seluruh struktur dapat diinspeksi seperti *white-box*, sementara *black-box* murni berisiko melewatkan persoalan pada pemetaan preferensi, penyaringan katalog, dan penyusunan rekomendasi (Sommerville, 2016).

Penelitian Terkait

Bagian ini mensintesis penelitian terdahulu ke dalam tiga rumpun yang relevan dengan penelitian ini, yaitu AI agent untuk layanan pelanggan, AI agent untuk sistem rekomendasi, dan *grey-box testing*.

AI dan AI Agent untuk Layanan Pelanggan.

Penelitian pada rumpun ini didominasi eksperimen dan studi berbasis skenario yang menyoroti respons psikologis pengguna dan pengaruh desain interaksi. Isyarat antropomorfik dan strategi percakapan memengaruhi perilaku pengguna, mulai dari peningkatan compliance melalui anthropomorphic design cues dan foot-in-the-door (Adam et al., 2021), perubahan persepsi akibat framing dan isyarat manusiawi (Araujo, 2018), strategi klarifikasi yang menjaga niat adopsi saat terjadi galat (Sheehan et al., 2020), hingga dampak pengungkapan identitas AI pada pembelian (Luo et al., 2019). Dari sisi kualitas layanan, kepuasan ditopang akurasi dan kredibilitas (Chung et al., 2020), keautentikan dan manfaat yang dirasakan (Rese et al., 2020), serta kualitas percakapan dan pemulihan kegagalan (Hsu & Lin,

2023). Celah yang konsisten adalah dominasi lingkungan eksperimental dan fokus pada persepsi, bukan kinerja sistem *end-to-end* pada penggunaan nyata.

AI dan AI Agent untuk Sistem Rekomendasi.

Pada rumpun rekomendasi percakapan, sejumlah studi menyediakan fondasi arsitektural dan taksonomi, dari arsitektur konseptual CRS yang menempatkan dialogue management sebagai pengendali (Jannach et al., 2022) hingga pemetaan agen rekomendasi berbasis LLM ke dalam tiga paradigma dan empat modul inti (Peng et al., 2025). Pemanfaatan LLM menaikkan kualitas CRS yang dinilai lebih bermakna, tetapi rapuh pada konteks dialog panjang dan lemah berkorelasi dengan metrik otomatis (Manzoor et al., 2024), sementara desain preference *elicitation* memengaruhi kinerja dan efisiensi dialog (Ziegfeld et al., 2025) serta efektif mengatasi cold-start (Christakopoulou et al., 2016). Persoalan inti yang belum tuntas adalah validitas keluaran terhadap katalog, karena LLM masih dapat merekomendasikan item tak tersedia saat pengaman dilepas (Granada et al., 2023), sehingga evaluasi CRS ditegaskan harus holistik dan berpusat pada pengguna (Jannach, 2023).

Grey-box Testing

Kajian pada rumpun ini menegaskan pengujian yang kuat perlu memadukan perspektif fungsional dan struktural. *Grey-box* diposisikan sebagai pendekatan tengah yang memakai pengetahuan internal parsial seperti arsitektur, aliran data, dan state untuk merancang skenario, tetapi tetap dieksekusi melalui antarmuka (Ehmer & Khan, 2012); gabungan *black-box* dan *white-box* dipandang kompromi wajar karena pengujian menyeluruh tidak realistis (Myers et al., 2012; Pressman, 2001), dan metodologi *grey-box* yang lebih formal telah dirumuskan untuk siklus hidup perangkat lunak (Coulter, 1999). Celahnya, kajian ini dominan konseptual dan minim pembuktian empiris pada sistem nyata.

Posisi penelitian ini dari ketiga rumpun tersebut, belum ada penelitian yang memadukan secara terintegrasi agen konsultasi pra-pembelian berbasis percakapan, rekomendasi laptop yang tervalidasi terhadap katalog, antarmuka WhatsApp, dan orkestrasi low-code n8n, yang sekaligus dievaluasi secara empiris melalui *grey-box testing* dan uji pengguna. Penelitian ini mengisi celah tersebut dan menyumbang penerapan empiris *grey-box* pada sistem terintegrasi nyata serta desain CRS yang menekankan validitas melalui *grounding* katalog.

III. METODE

Jenis Penelitian

Penelitian ini menggunakan pendekatan penelitian dan pengembangan (*Research and Development*/R&D) model Sugiyono (2018) yang mencakup empat tahapan utama, yaitu penelitian, perancangan, produksi, dan pengujian (4P). Pendekatan ini dipilih karena tujuan penelitian adalah menghasilkan produk dan

menguji kelayakannya, bukan menguji hubungan antarvariabel. Produk yang dimaksud adalah prototipe AI Agent layanan konsultasi rekomendasi laptop berbasis *workflow automation* dengan antarmuka WhatsApp. Agar selaras dengan karakteristik produk yang berupa prototipe perangkat lunak, keempat tahapan 4P diadaptasi menjadi tiga tahap operasional, yaitu studi pendahuluan, pengembangan sistem, dan evaluasi akhir.

Partisipan dipilih dengan teknik *purposive sampling*, yaitu penentuan sumber data berdasarkan pertimbangan tertentu agar sesuai tujuan penelitian (Sugiyono, 2018), dengan kriteria utama individu pada tahap pra-pembelian (interest), yaitu mahasiswa atau pekerja yang berencana membeli laptop dalam kurun kurang lebih 12 bulan dan memiliki kebutuhan serta kisaran anggaran yang jelas, sementara calon yang tidak dapat menyebutkan keduanya tidak dilibatkan. Partisipan ditetapkan lima orang (IF1 sampai IF5) yang mengikuti dua tahap sekaligus, yaitu wawancara pendahuluan sebelum perancangan dan *usability testing* setelah pengembangan, sehingga kebutuhan awal dapat ditinjau ulang melalui pengalaman langsung; jumlah ini mengacu pada prinsip formative *usability testing* bahwa sedikit partisipan sudah mampu mengungkap sebagian besar persoalan *usability* dominan (Myers et al., 2012) serta pada karakteristik penelitian kualitatif yang mengutamakan kedalaman informasi ketimbang keterwakilan statistik (Sugiyono, 2018).

Data dikumpulkan melalui tiga kegiatan berurutan sesuai tahapan R&D, yaitu wawancara pendahuluan (preliminary survey) untuk menghimpun kebutuhan sebelum perancangan, *grey-box testing* untuk menguji sistem, dan *usability testing* untuk menguji dari sisi pengguna, dengan kedua wawancara memakai teknik semi-terstruktur yang terarah oleh pertanyaan pokok namun terbuka untuk penggalian mendalam (Sugiyono, 2018). Wawancara pendahuluan dilaksanakan tatap muka sebelum perancangan dan direkam atas izin partisipan untuk menghimpun kebutuhan, harapan, dan kekhawatiran calon pengguna, sedangkan uji pengguna dilaksanakan setelah prototipe stabil dengan partisipan menjalankan tugas berinteraksi dengan *chatbot* hingga memperoleh rekomendasi, meminta perbandingan antarlaptop, serta mengubah kebutuhan, anggaran, atau preferensi di tengah percakapan; wawancara kemudian difokuskan pada persepsi ketepatan rekomendasi, kemudahan interaksi, kejelasan penjelasan, dan kepuasan penggunaan.

Uji keabsahan data mencakup kredibilitas, keteralihan, reliabilitas, dan objektivitas (Sugiyono, 2018), dengan penekanan pada kredibilitas melalui triangulasi sumber dan member check. Triangulasi sumber membandingkan jawaban kelima informan (IF1 sampai IF5) secara silang untuk menemukan pola yang konsisten maupun perbedaan menonjol, sehingga kesimpulan tidak bertumpu pada satu sumber. Member check mengembalikan ringkasan interpretasi wawancara kepada tiap informan untuk dikonfirmasi, dikoreksi, atau ditambahkan, dan hasilnya didokumentasikan sebagai bukti keabsahan.

Instrumen Penelitian

Dalam penelitian kualitatif, instrumen utama adalah peneliti sendiri sebagai human instrument yang menetapkan fokus, memilih informan, mengumpulkan, menilai, dan menafsirkan data (Sugiyono, 2018), dibantu pedoman wawancara dan alat perekam. Dua instrumen pendukung diadaptasi dari kerangka CRS-Que (Jin et al., 2024). Instrumen preliminary survey disusun dari kisi-kisi (blueprint) demi validitas isi dan mencakup enam aspek, yaitu perceived usefulness, *trust* & confidence, intention to use, accuracy, interaction adequacy, dan *explainability*, masing-masing tiga butir sehingga berjumlah 18 butir yang bersifat ekspektatif karena dilaksanakan sebelum prototipe dibangun. Instrumen *usability testing* mencakup 18 konstruk dalam empat dimensi, yaitu *perceived qualities* (8 konstruk), *user beliefs* (6 konstruk), *user attitudes* (2 konstruk), dan *behavioral intentions* (2 konstruk), dipakai setelah informan mencoba prototipe sehingga diarahkan pada pengalaman nyata, dengan tiap konstruk diwakili satu pernyataan terbuka beserta kolom justifikasi demi keterlacakan.

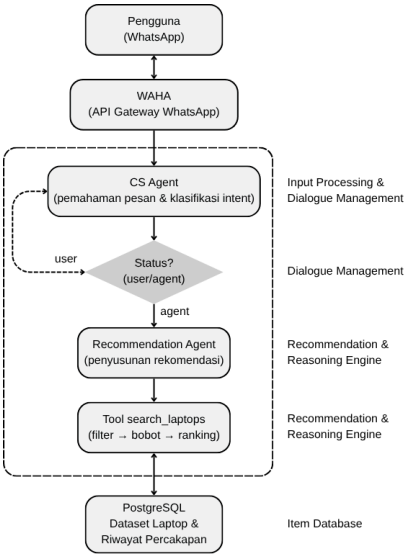
Data dianalisis secara kualitatif dengan model interaktif Miles dan Huberman dalam Sugiyono (2018), yaitu reduksi data, penyajian data, serta penarikan kesimpulan dan verifikasi. Transkrip dirangkum dan dipilah menurut aspek pada pedoman, lalu disajikan sebagai uraian naratif per aspek sehingga pola kebutuhan pada tahap pendahuluan serta pola pengalaman dan persepsi pada tahap uji pengguna terbaca jelas. Penarikan kesimpulan diverifikasi melalui triangulasi sumber, dengan kesimpulan wawancara pendahuluan menjadi dasar perumusan kebutuhan sistem dan kesimpulan uji pengguna menjadi dasar penilaian serta saran perbaikan prototipe.

Desain Prototipe

Prototipe dikembangkan memakai kerangka *Conversational Recommender System* (CRS) yang menempatkan dialog sebagai inti rekomendasi sehingga cocok dengan antarmuka WhatsApp, dengan mesin rekomendasi berbasis pengetahuan (knowledge-based) bermekanisme penyaringan berbasis batasan (*constraint-based filtering*) dan pemeringkatan berbasis utilitas (utility-based *ranking*) (Felfernig & Burke, 2008; Manouselis & Costopoulou, 2007). Pendekatan ini dipilih karena laptop berjenis produk banyak atribut teknis dan, berbeda dari penyaringan kolaboratif yang rentan cold-start, menilai item langsung dari preferensi terkini sehingga dapat merekomendasikan sejak interaksi pertama dan lebih mudah dijelaskan (Burke, 2002; Uta et al., 2024). Arsitektur menyeluruh ditampilkan pada Gambar 2: alur data bermula dari pengguna melalui WhatsApp dan WAHA, masuk ke alur kerja n8n tempat CS Agent (*Customer Service Agent*) dan *Recommendation Agent* bekerja, lalu berakhir pada PostgreSQL yang menyimpan *dataset* laptop dan riwayat percakapan, sehingga percakapan dan rekomendasi berjalan sebagai satu alur khas CRS (Jannach et al., 2022).

Pemetaan arsitektur referensi CRS ke implementasi bersifat fungsional: CS Agent menangani pemrosesan masukan dan keluaran serta manajemen dialog,

sedangkan *Recommendation Agent* menyusun rekomendasi dan penjelasannya; pemodelan pengguna diwujudkan lewat pengelompokan *use-case* dan objek profil, mesin rekomendasi lewat penyaringan (Tabel 2) serta pembobotan dan penilaian (Tabel 1), dan basis data item lewat struktur atribut *dataset*. Antarmuka dirancang pada WhatsApp melalui WAHA mengikuti tren *conversational commerce* dengan indikator seen dan typing serta format keluaran ringkas agar nyaman dibaca di ponsel.



Sumber: Diolah oleh Peneliti, 2026

Gambar 2. Arsitektur Sistem Prototipe

Perancangan *Item Model* (*Dataset*)

Item model sistem berupa atribut *dataset* laptop yang sejajar dengan komponen item database pada arsitektur CRS (Jannach et al., 2022). Atribut dikelompokkan menjadi identitas, informasi umum, CPU, memori, penyimpanan, display, GPU, sistem operasi, dan garansi, dengan atribut numerik seperti harga, kapasitas, dan skor *benchmark* dipakai untuk penyaringan dan penilaian skor, sedangkan atribut kategorikal dan nama *benchmark* bersifat deskriptif atau cadangan. Secara lengkap, ke-27 atribut tersebut mencakup identitas dan informasi umum (id, brand, model, device_category, dan price); spesifikasi prosesor (cpu_brand, cpu_series, cpu_core_count, dan cpu_thread_count) beserta skor *benchmark*-nya (passmark_cpu_name, passmark_cpu_mark, dan passmark_single_thread); kapasitas memori dan penyimpanan (ram_gb dan storage_gb); spesifikasi layar (display_width_px, display_height_px, dan display_size_inch); spesifikasi grafis (gpu_brand, gpu_model, gpu_vram_gb, dan gpu_type) beserta skor *benchmark*-nya (passmark_gpu_name, passmark_g3d_mark, dan passmark_g2d_mark) serta penanda ketersediaannya

(gpu_benchmark_available); dan sistem operasi (os_name) beserta masa garansi (warranty_years).

Perancangan Pemodelan Pengguna dan Strategi Elisitasi

Kebutuhan pengguna dimodelkan melalui pengelompokan *use-case* dan objek profil berisi kebutuhan, anggaran, dan filter yang terus diperbarui selama percakapan, sebagai wujud user modeling pada arsitektur CRS (Jannach et al., 2022) sekaligus selaras dengan sistem berbasis pengetahuan yang memetakan kebutuhan tingkat tinggi ke properti teknis produk (Burke, 2002; Uta et al., 2024). CS Agent mengenali *use-case* dari percakapan lalu memetakannya ke satu nilai kebutuhan, dan meminta klarifikasi bila ambigu agar tiap kueri berpijak pada satu *use-case* yang jelas sehingga matriks bobot (Tabel 1) dapat dipakai pasti. Profil dibangun dengan strategi elisitasi bergaya System Ask, User Respond (SAUR), yaitu sistem aktif bertanya sebelum merekomendasikan (Zhang et al., 2018), dengan urutan tujuan penggunaan, anggaran, preferensi khusus, lalu konfirmasi, serta pemisahan preferensi lunak (soft preference) yang menentukan bobot dari batasan keras (*hard constraint*) berupa anggaran dan filter. Tersedia tujuh kategori *use-case*, yaitu gaming, programming, office, education, graphic design, video editing, dan general use, dengan prioritas spesifikasi berbeda tiap kategori.

Perancangan Mesin Rekomendasi: Penyaringan dan Penilaian

Pembobotan Multi-Atribut

Matriks pembobotan pada Tabel 1 menerjemahkan tiap *use-case* menjadi sehimpunan bobot komponen penilaian. Matriks ini mewujudkan pemeringkatan berbasis utilitas yang berlandaskan *Multi-Attribute Utility Theory* (MAUT), yaitu pemodelan utilitas pengguna atas suatu item sebagai gabungan terbobot beberapa kriteria (Adomavicius & Kwon, 2015; Manouselis & Costopoulou, 2007). Jumlah bobot tiap baris dibuat sama dengan 1,00 agar skor antarlaptop sebanding. Nilai bobot ditetapkan secara heuristik menurut karakteristik tiap *use-case* dan acuan kategori *benchmark* PassMark, sebagaimana lazim pada kerangka MAUT (Winterfeldt & Edwards, 1993). Setiap *use-case* menekankan komponen paling menentukan performanya, misalnya gaming pada GPU 3D, programming dan video editing pada CPU multi-core dan RAM, sedangkan office, education, dan general use pada CPU single-thread dan RAM. Yang menentukan adalah urutan prioritas antarkomponen, bukan ketepatan angkanya, sehingga pemeringkatan cukup tahan terhadap perubahan kecil bobot, dan ketahanan ini diuji secara empiris melalui analisis sensitivitas (Triantaphyllou & Sanchez, 1997).

Tabel 1. Matriks Pembobotan

Use-case	CPU multi	CPU single	RAM	Storage	GPU 3D	GPU 2D	VRAM	Total
Gaming	0,12	0,15	0,11	0,10	0,38	0,00	0,14	1,00
Programming	0,30	0,24	0,28	0,18	0,00	0,00	0,00	1,00
Office	0,17	0,37	0,30	0,16	0,00	0,00	0,00	1,00
Education	0,21	0,33	0,28	0,18	0,00	0,00	0,00	1,00

Graphic design	0,13	0,22	0,20	0,08	0,15	0,12	0,10	1,00
Video editing	0,26	0,11	0,20	0,10	0,20	0,00	0,13	1,00
General use	0,22	0,33	0,28	0,17	0,00	0,00	0,00	1,00

Sumber: Diolah oleh Peneliti, 2026

Normalisasi Berbasis Acuan

Tiap komponen penilaian diambil dari atribut *dataset*, misalnya CPU multi-core dari passmark_cpu_mark, RAM dari ram_gb, dan GPU 3D dari passmark_g3d_mark. Karena rentang nilai antaratribut jauh berbeda, nilai mentah tidak dapat langsung dijumlahkan dan harus dinormalisasi lebih dulu (Vafaei et al., 2022). Karena semua komponen bersifat menguntungkan (benefit), normalisasi dilakukan dengan membagi nilai mentah terhadap acuan tetap lalu membatasinya paling tinggi bernilai satu, seperti pada Persamaan 1, sehingga kelebihan kapasitas di atas acuan tidak menambah skor.

$$r_{ij} = \left(\frac{x_{ij}}{a_j}, 1\right) \quad (1)$$

dengan r_{ij} nilai ternormalisasi komponen ke-j pada laptop ke-i, x_{ij} nilai mentah atribut, dan a_j nilai acuan, sedangkan fungsi min menjamin hasil tidak melampaui satu. Normalisasi berbasis acuan tetap ini dipilih alih-alih normalisasi terhadap nilai maksimum himpunan yang lazim pada *Simple Additive Weighting* (Taherdoost & Madanchian, 2023; Vafaei et al., 2022); acuan diambil dari data, yaitu persentil ke-95 sebaran skor PassMark, sebagai penerapan metode tingkat aspirasi (*aspiration level*) (Wierzbicki, 1982) yang mencerminkan prinsip satisficing (Simon, 1955) dan target-based normalization dalam MCDM (Jahan et al., 2012). Karena skor tiap laptop dihitung mandiri, pemeringkatan kebal terhadap rank reversal (Y.-M. Wang & Luo, 2009); acuan RAM, penyimpanan, dan VRAM dipatok pada 32 GB, 1.024 GB, dan 8 GB.

Agregasi Skor

Skor akhir dihitung sebagai penjumlahan utilitas terbobot atas semua nilai ternormalisasi, yakni bentuk aditif MAUT yang dalam praktik MCDM dikenal sebagai *Simple Additive Weighting*, seperti pada Persamaan 2 (Manouselis & Costopoulou, 2007; Taherdoost & Madanchian, 2023).

$$V_i = \sum_j w_j \cdot r_{ij} \quad (2)$$

dengan V_i menyatakan skor akhir laptop ke-i, w_j bobot komponen sesuai *use-case* terpilih (Tabel 1), dan r_{ij} nilai ternormalisasi. Karena jumlah bobot sama dengan satu dan tiap nilai ternormalisasi berada pada rentang nol sampai satu, skor V_i juga berada pada rentang nol sampai satu sehingga langsung dapat diperingkat. Pendekatan ini dipilih karena sederhana, transparan, dan mudah ditelusuri, dengan semua nilai dibulatkan ke tiga angka di belakang koma memakai pembulatan setengah ke atas.

Penyaringan Berbasis Batasan

Tabel 2 memuat seluruh filter atau batasan keras (*hard constraint*) sebagai mekanisme penyaringan berbasis batasan (*constraint-based filtering*), yaitu pemetaan kebutuhan pengguna menjadi sehimpunan kondisi atas properti produk (Felfernig & Burke, 2008). Sebagian besar filter opsional melalui pengabaian nilai bawaan, sedangkan anggaran (budget) diperlakukan sebagai filter wajib berupa batas atas harga; karena anggaran bernilai nol menghasilkan himpunan kosong, alur elisitasi memastikan anggaran selalu terisi dan terkonfirmasi sebelum pencarian. Penyaringan dijalankan lebih dulu untuk mempersempit kandidat sebelum pembobotan dan penilaian.

Tabel 2. Filter atau Constraint

Kelompok Filter	Key Filter	Atribut Target	Operator
Anggaran (wajib)	budget	price	price ≤ budget
Batas minimum numerik	ram_min_gb, storage_min_gb, gpu_vram_min_gb, warranty_min_years, display_size_min_inch	ram_gb, storage_gb, gpu_vram_gb, warranty_years, display_size_inch	≥
Batas maksimum layar	display_size_max_inch	display_size_inch	≤
Tipe GPU	gpu_type	gpu_type	=
Merek laptop	brand / brand_exclude	brand	= ANY / NOT ANY
Merek CPU	cpu_brand / cpu_brand_exclude	cpu_brand	= ANY / NOT ANY
Seri CPU	cpu_series_contains / cpu_series_exclude_contains	cpu_series	substring (OR) / NOT substring
Merek GPU	gpu_brand / gpu_brand_exclude	gpu_brand	= ANY / NOT ANY
Model GPU	gpu_model_contains / gpu_model_exclude_contains	gpu_model	substring (OR) / NOT substring
Sistem operasi	os_name / os_name_exclude	os_name	= ANY / NOT ANY
Kategori perangkat	device_category / device_category_exclude	device_category	= ANY / NOT ANY

Sumber: Diolah oleh Peneliti, 2026

Seluruh filter selain anggaran bersifat opsional dan diabaikan saat bernilai nol, *string* kosong, atau *array* kosong, dengan filter batas numerik bertipe *numeric* dan filter kesetaraan atau kategori bertipe *string* atau *array*. Bila hasil penyaringan kosong, sistem tidak mengarang produk pengganti melainkan memberi tahu pengguna secara jujur dan menyarankan pelonggaran anggaran atau preferensi.

Rangkaiannya membentuk satu alur utuh, yaitu penyaringan berbasis batasan, lalu normalisasi, pembobotan, perhitungan skor, dan pemeringkatan, sesuai praktik lazim pada sistem rekomendasi berbasis pengetahuan (Adomavicius & Kwon, 2015; Felfernig & Burke, 2008).

Perancangan Penyajian dan Penjelasan Rekomendasi

Recommendation Agent menyajikan hasil pemeringkatan berdasarkan dua prinsip. Pertama, penyajian bersifat berbasis pengambilan (retrieval-based): agen tidak mengarang kandidat atau spesifikasi, melainkan menampilkan hasil dari basis data tanpa mengubah urutannya, sehingga risiko halusinasi ditekan dan pengalaman dipersepsikan lebih baik dibanding pembangkitan murni (He et al., 2023; Manzoor & Jannach, 2022). Kedua, penjelasan berorientasi manfaat, yaitu menerjemahkan spesifikasi teknis menjadi manfaat nyata sambil mengaitkan tiap alasan dengan kebutuhan dan anggaran pengguna, berpijak pada explainable recommendation (Zhang & Chen, 2020) dengan sasaran efektivitas dan kepercayaan (Tintarev & Masthoff, 2015). Agar tetap dipercaya, agen jujur menyatakan keterbatasan data, misalnya berat, daya tahan baterai, atau kualitas layar yang tidak ada dalam *dataset*, dan secara bawaan menampilkan paling banyak lima kandidat teratas beserta tawaran tindak lanjut berupa rincian, perbandingan, atau alternatif.

Perancangan Model Dialog

Model dialog mengatur jalannya percakapan dan menjadi pembeda utama CRS dari sistem rekomendasi biasa, sebagai wujud dialogue management pada arsitektur CRS (Jannach et al., 2022). Variabel kontrol eksplisit yang menggerakkan alur adalah status keluaran CS Agent bernilai user atau agent: user menandakan sistem masih menggali kebutuhan sehingga CS Agent bertanya, mengklarifikasi, atau menjawab langsung, sedangkan agent menandakan sistem siap merekomendasikan sehingga proses diteruskan ke *Recommendation Agent*, dan keputusan biner inilah inti kebijakan aksi CRS (Christakopoulou et al., 2016). Ragam maksud pengguna ditangani sebagai kebijakan perilaku, bukan field tersendiri: pencarian baru dan ulang memicu pemanggilan alat search_laptops, minta alternatif dijawab dengan shortlist yang belum muncul, minta rincian dan perbandingan dijawab dari memori, sedangkan kritik atau revisi profil memperbarui profil lalu konfirmasi ulang mengikuti prinsip critiquing-based recommendation (Chen & Pu, 2012). Untuk menopang interaksi banyak giliran, memori percakapan dan shortlist terakhir disimpan di PostgreSQL sehingga tindak lanjut dijawab dari konteks, sebagaimana pentingnya pengelolaan state pada CRS banyak giliran (Gao et al., 2023).

Pemilihan *Large Language Model* (LLM)

LLM berperan sebagai mesin penalaran yang menafsirkan maksud pengguna, menyusun respons, dan memicu pemanggilan *tool* pada alur kerja n8n, dan pemilihannya diposisikan sebagai design decision, bukan variabel yang diuji eksperimental, sehingga sasarannya bukan menemukan LLM terbaik mutlak melainkan model yang memenuhi kebutuhan fungsional secara memadai dan efisien. Kriteria utamanya adalah dukungan *function calling* sebagai syarat wajib

karena Recommendation Agent memanggil *tool* search_laptops, ditambah latensi dan *throughput* untuk percakapan mendekati *real-time*, biaya, ketersediaan *free tier*, serta kompetensi berbahasa Indonesia. Atas dasar itu, CS Agent memakai Gemini 2.5 Flash-Lite dan Recommendation Agent memakai Gemini 2.5 Flash karena memenuhi seluruh kriteria wajib melalui *free tier* Google AI Studio, dengan Gemma sebagai *fallback* melalui OpenRouter; pemilihan sengaja dibatasi pada satu *fast* atau *lightweight tier* sehingga tidak mencakup model kelas *frontier*, dan keterbatasan ini dibahas pada bagian keterbatasan.

Desain dan Pendekatan Evaluasi

Evaluasi dirancang menilai kualitas prototipe dari dua sudut pandang yang saling melengkapi, yaitu berpusat pada sistem (system-centric) yang menilai kebenaran fungsi secara objektif melalui pengujian perangkat lunak, dan berpusat pada pengguna (user-centric) yang menilai pengalaman melalui uji pengguna, sejalan dengan penekanan CRS pada pengukuran yang tidak semata berorientasi akurasi (Jannach et al., 2022). Pengujian fungsional memakai pendekatan *grey-box*, kombinasi *black-box* dan *white-box*: perilaku diuji dari sisi masukan dan keluaran, namun kasus uji dirancang dengan memanfaatkan pengetahuan atas struktur internal seperti alur kerja n8n, logika mesin rekomendasi, dan skema basis data (Beizer & Wiley, 1996; Ehmer & Khan, 2012; Pressman, 2001; Sommerville, 2016). Pendekatan ini sesuai untuk sistem berbasis LLM yang keluarannya bervariasi karena pengujian dapat difokuskan pada komponen deterministik, dan kebenaran keluaran ditentukan memakai *oracle* pengujian sebagai acuan independen (Barr et al., 2015), yaitu hasil hitung ulang terpisah untuk komponen deterministik dan rubrik kesesuaian untuk keluaran bahasa.

Kasus Uji, Metrik, dan Prosedur Pengujian

Kasus uji dirancang memakai teknik baku perancangan pengujian agar cakupannya sistematis (ISTQB, 2024), meliputi partisi ekuivalensi, analisis nilai batas untuk tepi rentang seperti ambang anggaran, tabel keputusan untuk kombinasi filter, transisi keadaan untuk perpindahan status dialog, pengujian berbasis *use-case* untuk skenario percakapan menyeluruh, serta pengujian negatif untuk masukan tidak valid atau di luar topik, sehingga mencakup jalur normal maupun kondisi tepi dan kesalahan. Metrik dipilih sesuai sistem rekomendasi berbasis pengetahuan yang keluarannya berupa daftar terperingkat, bukan teks bebas, sehingga metrik pembangkitan teks seperti BLEU tidak relevan; komponen rekomendasi memakai precision, recall, dan F1-score yang lazim pada information retrieval (Manning et al., 2008), sedangkan komponen fungsional memakai metrik berbasis kelulusan kasus uji, dengan definisi operasional tiap metrik dirangkum pada Tabel 3.

Tabel 3. Metrik dan Definisi Operasional

Metrik	Definisi Operasional	Satuan
Precision	Proporsi rekomendasi yang relevan terhadap seluruh rekomendasi yang ditampilkan.	Persentase

Recall	Proporsi item relevan yang berhasil ditampilkan terhadap seluruh item relevan yang seharusnya muncul.	Persentase
F1-score	Rata-rata harmonis precision dan recall sebagai ukuran keseimbangan keduanya.	Persentase
Akurasi klasifikasi maksud	Ketepatan sistem menentukan status user atau agent sesuai konteks percakapan.	Persentase
Keberhasilan pemanggilan tool	Proporsi kasus saat tool search_laptops dipanggil dengan parameter yang benar ketika seharusnya dipanggil.	Persentase
Ketepatan penyaringan	Kesesuaian hasil penyaringan dengan batasan keras yang diminta pengguna.	Persentase
Ketepatan perhitungan skor	Kesesuaian skor dan urutan sistem dengan hasil perhitungan ulang secara independen.	Persentase
Ketepatan penanganan hasil kosong	Ketepatan sistem memberi tahu dan menyarankan pelonggaran saat kandidat kosong.	Persentase
Konsistensi profil	Kesesuaian rekomendasi dengan profil kebutuhan dan anggaran yang telah dikonfirmasi.	Skala rubrik
Kesetiaan penjelasan	Tingkat penjelasan yang berpijak pada data tanpa klaim di luar dataset.	Skala rubrik
Keberhasilan penanganan kritik	Ketepatan sistem memperbarui profil dan hasil setelah kritik atau revisi pengguna.	Persentase
Penanganan pesan di luar topik	Ketepatan sistem mengarahkan kembali percakapan ke topik rekomendasi laptop.	Persentase

Sumber: Diolah oleh Peneliti, 2026

Dataset pengujian disusun secara sengaja (*purposive*) agar mewakili keragaman *use-case*, rentang anggaran, kombinasi filter, serta skenario tepi dan negatif, dengan tiap kasus dilengkapi masukan, kondisi awal, dan keluaran yang diharapkan sebagai *oracle*. Prosedur berjalan tiga langkah, yaitu penyiapan kasus uji beserta *oracle* pada Google Sheets, eksekusi otomatis melalui n8n Evaluations pada alur produksi, lalu pencatatan keluaran dan metrik untuk dianalisis. Komponen deterministik dibandingkan dengan perhitungan ulang independen, sedangkan keluaran bahasa dinilai memakai rubrik agar objektif meski keluaran LLM bervariasi.

Validitas dan Indikator Keberhasilan

Beberapa ancaman validitas dipertimbangkan sesuai praktik penelitian rekayasa perangkat lunak (Runeson & Höst, 2009; Wohlin et al., 2012): validitas konstruk dijaga dengan menyandarkan metrik pada rujukan baku dan definisi operasional, validitas internal diperkuat oleh *oracle* independen dan pengujian pada alur produksi, validitas eksternal dibatasi cakupan *dataset* laptop sehingga generalisasi dilakukan hati-hati, dan reliabilitas dijaga melalui otomasi n8n Evaluations yang dapat diulang. Evaluasi ini menilai kebenaran fungsional dan kualitas rekomendasi terhadap *oracle*, bukan membandingkan akurasi antaralgoritma pada tolok ukur publik, sehingga terhindar dari sebagian persoalan reproduktifitas pada penelitian sistem rekomendasi (Ferrari Dacrema et al., 2021).

Indikator keberhasilan disusun memakai pendekatan Goal-Question-Metric (GQM) yang menautkan tujuan penelitian dengan pertanyaan evaluatif lalu ke metrik terukur (Van Solingen et al., 2002), mencakup keberhasilan implementasi fungsional, kualitas hasil rekomendasi, kualitas interaksi percakapan, dan penerimaan pengguna, dengan kriteria tiap indikator dirangkum pada Tabel 4.

Tabel 4. Indikator dan Kriteria Keberhasilan

Indikator	Deskripsi	Kriteria Keberhasilan
Fungsionalitas prototipe	Seluruh komponen inti berjalan sesuai rancangan dari elisitasi hingga penyajian rekomendasi.	Mayoritas kasus uji fungsional lulus tanpa kegagalan kritis.
Ketepatan pemrosesan dialog	Sistem menentukan status user atau agent dan memanggil tool dengan benar.	Akurasi klasifikasi maksud dan keberhasilan pemanggilan tool tinggi.
Ketepatan penyaringan dan skor	Hasil penyaringan dan pemeringkatan sesuai perhitungan ulang independen.	Ketepatan penyaringan dan perhitungan skor mendekati sempurna.
Kualitas rekomendasi	Rekomendasi relevan dengan profil kebutuhan dan anggaran pengguna.	Precision, recall, dan F1-score berada pada tingkat memadai.
Kualitas penjelasan	Penjelasan berpijak pada data tanpa klaim di luar dataset.	Skor rubrik kesetiaan penjelasan dan konsistensi profil tinggi.
Ketangguhan sistem	Sistem menangani hasil kosong, kritik, dan pesan di luar topik dengan tepat.	Kasus tepi dan negatif ditangani sesuai perilaku yang diharapkan.
Penerimaan pengguna	Pengguna menilai layanan bermanfaat dan nyaman digunakan.	Umpan balik uji pengguna dominan positif pada aspek utama.

Sumber: Diolah oleh Peneliti, 2026

Analisis Sensitivitas Pembobotan

Karena bobot pada Tabel 1 ditetapkan secara heuristik, dilakukan analisis sensitivitas untuk menguji ketahanan pemeringkatan terhadap perubahan bobot, sesuai anjuran pengambilan keputusan multikriteria bahwa yang menentukan adalah urutan prioritas, bukan nilai bobot yang presisi (Triantaphyllou & Sanchez, 1997). Simulasi Monte Carlo 2.000 iterasi dengan memperturbasi bobot sekitar sepuluh sampai dua puluh persen menghasilkan korelasi peringkat Spearman tidak kurang dari 0,999 pada katalog penuh dan tetap di atas 0,997 pada berbagai subset anggaran, meski konsistensi sepuluh besar sedikit menurun pada subset ketat karena skor kandidat berdekatan. Hasil ini menegaskan keputusan pemeringkatan tidak bergantung pada nilai bobot yang presisi, melainkan pada urutan prioritas antarkomponen penilaian, sehingga skema pembobotan heuristik cukup absah, dan penurunan konsistensi tersebut menjadi alasan penerapan kunci pengurutan tambahan berupa harga dan id agar urutan tetap pasti dan dapat diulang.

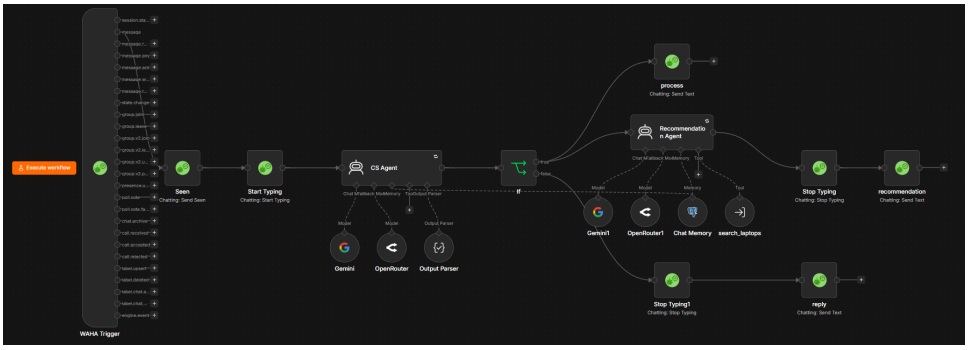
IV. HASIL DAN PEMBAHASAN

Pengembangan *Workflow* n8n

Subbab ini membahas implementasi rancangan ke dalam *workflow* n8n yang menjadi penggerak *chatbot*. Sistem dibangun dari dua bagian, yaitu *workflow* utama sebagai lapisan percakapan WhatsApp melalui WAHA, dan *subworkflow* yang mengambil data laptop dari basis data. Pada *workflow* utama terdapat dua agen, yaitu CS Agent sebagai pengelola dialog dan *Recommendation Agent* sebagai penghasil rekomendasi. Penjelasan diuraikan per *node* beserta konfigurasi dan fungsinya.

Workflow Utama

Workflow utama disusun sebagai satu alur berantai yang memadukan lapisan percakapan WhatsApp dan orkestrasi dua agen. Pesan yang tiba melalui WAHA *Trigger* ditandai terbaca dan memunculkan indikator mengetik, lalu diteruskan ke CS Agent, yaitu *node* AI Agent pengelola dialog yang memelihara konteks melalui Chat Memory dan merumuskan keluaran terstruktur tervalidasi *Output* Parser ke skema JSON berisi status, pesan, dan profil, dengan Gemini sebagai model utama dan OpenRouter sebagai cadangan.



Sumber: Diolah oleh Peneliti, 2026

Gambar 3. *Workflow* Utama

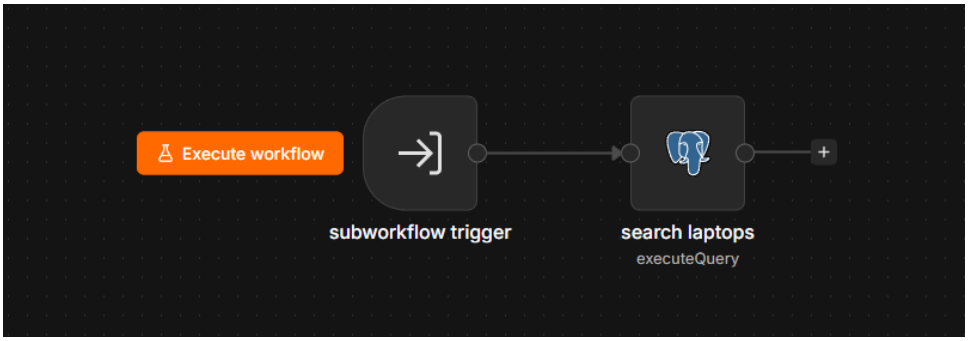
Field status menentukan arah pada *node* IF: bila user, sistem membalas langsung; bila agent, *Recommendation Agent* menyusun rekomendasi dengan memanggil alat search_laptops yang menjembatani *subworkflow* pencarian basis data sebelum jawaban akhir dikirim. Prinsip anti-halusinasi ditanamkan pada system message kedua agen, yaitu CS Agent dilarang menyaring atau memeringkat produk dan *Recommendation Agent* wajib menyusun jawaban semata dari shortlist search_laptops tanpa menghitung ulang skor atau mengubah urutan, sehingga setiap klaim berpijak pada katalog aktual dan *grounding* menjadi mekanisme utama penekan halusinasi. Rangkuman peran seluruh *node* ada pada Tabel 5.

Tabel 5. Ringkasan Node Workflow Utama

Node	Fungsi	Peran dalam alur
WAHA Trigger	Node pemicu berbasis webhook WAHA	Titik masuk seluruh alur; aktif setiap pesan WhatsApp masuk
Seen, Start Typing, Stop Typing, Stop Typing1	WAHA (Send Seen, Start Typing, Stop Typing)	Menandai pesan telah dibaca serta memunculkan dan menghentikan indikator mengetik
CS Agent	AI Agent (LangChain) layanan pelanggan dan pengelola dialog	Menggali kebutuhan, mengelola dialog, menetapkan status (user/agent)
Output Parser	Structured Output Parser (JSON)	Menjamin keluaran CS Agent sesuai skema JSON (status, pesan, profil)
Chat Memory	Postgres Chat Memory	Menyimpan dan memanggil riwayat percakapan (window 8 pesan)
Gemini, OpenRouter	Model bahasa utama dan cadangan CS Agent	Menghasilkan respons CS Agent; OpenRouter menggantikan model utama saat gagal (fallback)
Status	Node IF	Mempercabang alur berdasarkan field status
Reply, Proses	WAHA (Send Text)	Mengirim balasan langsung pada jalur user (Reply) atau pesan tunggu pada jalur agent (Proses)
Recommendation Agent	AI Agent (LangChain) penghasil rekomendasi	Menyusun rekomendasi dari profil dan hasil alat search_laptops
search_laptops	Tool Workflow	Menghubungkan Recommendation Agent ke subworkflow pencarian
Gemini1, OpenRouter1	Model bahasa utama dan cadangan Recommendation Agent	Menghasilkan narasi rekomendasi; OpenRouter1 menggantikan model utama saat gagal (fallback)
Recommendation	WAHA (Send Text)	Mengirim hasil akhir rekomendasi ke pengguna

Sumber: Diolah oleh Peneliti, 2026

Subworkflow



Sumber: Diolah oleh Peneliti, 2026

Gambar 4. *Subworkflow*

Subworkflow menjadi titik masuk ketika alat search_laptops dipanggil. *Node* Execute *Workflow Trigger* menetapkan kontrak masukan berupa kebutuhan, budget, dan filters yang memuat 21 kunci preferensi, sedangkan inti komputasionalnya adalah *Node* search laptop bertipe Postgres yang menjalankan satu kueri parameterized berbasis Common Table Expression sehingga penyaringan dan pemeringkatan dapat ditelusuri utuh. Kueri berjalan bertahap, yaitu membungkus parameter, memuat tabel bobot tujuh profil sesuai Tabel 1 dan memilih baris sesuai ranking_profile (kembali ke profil general bila tak dikenal), menyaring keras berdasarkan harga di bawah atau sama dengan budget ditambah kriteria opsional yang diabaikan saat netral, lalu menghitung recommendation_score sebagai penjumlahan terbobot nilai ternormalisasi yang dibatasi maksimum satu. Tahap akhir mengurutkan kandidat menurut skor menurun, harga menaik, lalu id sebagai pemecah seri deterministik, membatasi sepuluh kandidat teratas, dan mengembalikan jumlah kandidat lolos filter beserta array kandidat teratas.

Hasil Evaluasi Prototipe

***Workflow* Utama**

Pada setiap kasus, nilai aktual hasil ekstraksi identik dengan nilai harapan sehingga seluruh metrik kesesuaian bernilai 100 persen.

Tabel 6. Skor Metrik per Kasus Workflow Utama

No	Case ID	Status	Kebutuhan	Budget	Filters Sim.	E2E Format	E2E Tool	E2E Faithful
1	M-EP-01	100%	100%	100%	100%	-	-	-
2	M-EP-02	100%	100%	100%	100%	-	-	-
3	M-NEG-01	100%	100%	100%	100%	-	-	-
4	M-NEG-02	100%	100%	100%	100%	-	-	-

5	M-UNS-01	100%	100%	100%	100%	-	-	-
6	M-ST-01	100%	100%	100%	100%	100%	100 %	100%
7	M-UC-01	100%	100%	100%	100%	100%	100 %	100%
8	M-EMPTY-01	100%	100%	100%	100%	100%	100 %	100%
9	M-CMP-01	100%	100%	100%	100%	100%	100 %	100%
10	M-ALT-01	100%	100%	100%	100%	100%	100 %	100%

Sumber: Diolah oleh Peneliti, 2026

Skor kuantitatif hasil pengujian *workflow* utama disajikan pada Tabel 6, yang menampilkan skor tiap metrik untuk setiap kasus. Rata-rata seluruh metrik pada sepuluh kasus tersebut bernilai 100 persen.

Seluruh metrik tahap layanan pelanggan, yaitu ketepatan status, kebutuhan, budget, dan kemiripan filter, mencapai 100 persen pada sepuluh kasus. Ini menunjukkan CS Agent konsisten menempatkan percakapan pada status yang benar, mengekstraksi kebutuhan dan anggaran secara akurat, serta menyusun filter sesuai acuan, termasuk pada kasus negatif dan kontradiktif yang menuntut penolakan halus atau klarifikasi tanpa memaksakan atribut yang tidak didukung.

Pada lima kasus berjalur menyeluruh, ketiga metrik *end-to-end*, yaitu kepatuhan format maksimal lima kandidat, ketepatan pemanggilan tool, dan *faithfulness* yang ter-*grounding*, juga bernilai 100 persen. *Recommendation Agent* terbukti hanya menyusun jawaban dari shortlist yang dikembalikan *subworkflow*, menangani hasil kosong secara ramah dengan saran melonggarkan kriteria, dan menjalankan follow-up perbandingan maupun alternatif dari memori tanpa memanggil ulang tool.

Secara keseluruhan, seluruh kasus lulus dengan skor gabungan 100 persen dan memenuhi seluruh kriteria keberhasilan pada Tabel 4, yang menetapkan ambang tinggi untuk ketepatan status dan kebutuhan, kemiripan filter, kepatuhan format, ketepatan tool, *faithfulness* penuh, dan tingkat kelulusan skenario. Perlu dicatat, evaluasi ini deterministik terhadap *oracle* yang ditetapkan pada sepuluh skenario terkurasi, sehingga capaian sempurna mencerminkan kesesuaian perilaku pada ruang kasus yang diuji, bukan generalisasi tanpa batas.

Subworkflow

Kelima kasus *subworkflow* memperoleh kesesuaian penuh, baik pada jumlah kandidat lolos filter maupun urutan id teratas, sehingga metrik matched_total_match dan kesesuaian top-K bernilai 100 persen. Karena nilai *oracle* dikomputasi ulang independen dari logika penyaringan dan rumus skor, kesesuaian ini menegaskan implementasi kueri Common Table Expression berjalan persis seperti rancangan.

Pasangan SR-BVA-01 dan SR-BVA-02 membuktikan ketepatan penanganan nilai batas harga. Ketika budget disetel sama dengan harga sebuah laptop, produk itu

ikut lolos sehingga jumlah kandidat 898; ketika budget diturunkan satu rupiah, produk itu tereliminasi sehingga jumlah kandidat menjadi 897 dan komposisi id teratas bergeser. Perilaku ini konsisten dengan syarat harga di bawah atau sama dengan budget.

Kasus SR-DT-01 menunjukkan kombinasi filter GPU dedicated dan merek tertentu menyusutkan kandidat menjadi 78 dengan urutan teratas tepat, sedangkan SR-DT-02 dengan filter kontradiktif (merek sekaligus di-include dan di-exclude) menghasilkan nol kandidat sesuai harapan. Kasus SR-EP-01 memperlihatkan profil programming menghasilkan bobot berbeda sehingga urutan id teratas mencerminkan preferensi profil. Secara keseluruhan, *subworkflow* terbukti deterministik dan setia pada spesifikasi penyaringan dan pemeringkatan.

Hasil Wawancara *Usability Testing*

Uji pengguna dilaksanakan terhadap lima informan yang sebelumnya terlibat dalam survei pendahuluan, diberi kode IF01 sampai IF05. Setelah menggunakan prototipe, setiap informan menyelesaikan tiga kelompok tugas, yaitu memperoleh rekomendasi laptop, meminta perbandingan antarrekomendasi, serta mengubah kebutuhan, anggaran, atau preferensi selama percakapan. Wawancara dipetakan ke 18 item CRS-Que dengan kode AKU sampai ITP. Rekapitulasi jawaban disajikan pada Tabel 7 dan diuraikan menurut empat dimensi, yaitu *perceived qualities*, *user beliefs*, *user attitudes*, dan *behavioral intentions*, dengan pemeriksaan silang antarinforman untuk mengenali pola konsisten sekaligus perbedaan pengalaman.

Tabel 7. Rekapitulasi Temuan Wawancara CRS-Que (per dimensi)

Dimensi	Konstruk tercakup	Temuan kunci (IF01-IF05)
Perceived Qualities	*Accuracy*, *Novelty*, *Interaction Adequacy*, *Explainability*, *CUI Adaptability*, *CUI Understanding*, *CUI Response Quality*, *CUI Attentiveness*	Seluruh informan menilai rekomendasi tepat terhadap budget, spesifikasi, dan kebutuhan, serta hemat waktu dibanding mencari manual di *marketplace* atau YouTube. Aspek *novelty* menonjol karena beberapa menemukan laptop yang jarang diulas atau merek yang belum dikenal, dan sistem sempat memetakan batas atas pilihan seperti ROG Zephyrus. Interaksi dinilai mudah: kebutuhan spesifik tertangkap dan sistem proaktif bertanya, sedangkan setiap rekomendasi disertai alasan yang jelas dan *reasonable* serta meluas saat dilakukan komparasi. Sistem juga dinilai memahami maksud, menyimak, dan memberi respons yang sesuai; perpindahan konteks lancar dan menyesuaikan setelah umpan balik, meski umpan balik ganda sekaligus sempat membuat respons menumpuk.
User Beliefs	*Perceived Ease of Use*, *Perceived Usefulness*, *User Control*, *Transparency*, *CUI Humanness*, *CUI Rapport*	Sistem dinilai mudah digunakan dan membantu; pengguna leluasa mengganti atau menambah preferensi seperti budget, merek, dan jenis pemakaian, lalu sistem menyesuaikan. Transparansi dinilai baik karena kelebihan serta

		keterbatasan dijelaskan saat diminta dan disertai kesimpulan serta spesifikasi detail. Kesan manusiawi bervariasi: sebagian merasa mengalir seperti mengobrol dengan orang dan terbantu indikator mengetik ala WhatsApp, sementara satu informan menilai masih kaku dan menyarankan gaya bahasa lebih luwes. Rapport dirasakan positif melalui pertanyaan lanjutan yang membantu pengguna awam.
User Attitudes	*Trust & Confidence*, *Satisfaction*	Seluruh informan menilai sistem dapat diandalkan sebagai referensi awal sebelum membeli, bahkan layak direkomendasikan ke teman. Tingkat kepuasan berkisar cukup hingga sangat puas, dengan satu catatan bahwa gaya bahasa masih terasa seperti *chatbot*.
Behavioral Intentions	*Intention to Use*, *Intention to Purchase*	Semua informan menyatakan niat memakai kembali sistem saat mencari laptop, dan sebagian akan merekomendasikannya ke teman. Niat beli dipertimbangkan, terutama bila argumen sistem dinilai sangat cocok dengan kebutuhan.

Sumber: Diolah oleh Peneliti, 2026

Perceived Qualities

Pada aspek accuracy (AKU), kelima informan menilai rekomendasi sesuai kebutuhan, anggaran, dan spesifikasi, serta menghemat waktu dibanding pencarian mandiri, dan bagi informan yang kurang memahami spesifikasi, penjelasan performa dan penyimpanan membantu memahami alasan pemilihan. Pada aspek novelty (NOV), seluruh informan memperoleh pilihan yang sebelumnya jarang diketahui, termasuk produk di luar ulasan rujukan dan merek baru, sedangkan pada interaction adequacy (INT) preferensi dinilai mudah disampaikan karena sistem menangkap kebutuhan spesifik seperti resolusi layar dan batasan merek. Pada *explainability* (EXP), setiap rekomendasi disertai alasan yang dapat dipahami, dan penilaian CUI adaptability, understanding, response quality, serta attentiveness cenderung positif dengan dua catatan, yaitu respons menumpuk saat beberapa umpan balik diberikan berdekatan dan rekomendasi sempat di luar jangkauan ketika batas anggaran belum spesifik lalu disesuaikan pada percakapan yang sama. Secara keseluruhan, dimensi *perceived qualities* menunjukkan fungsi inti rekomendasi memenuhi kebutuhan, memberi nilai penemuan pilihan baru dengan alasan yang membantu, meski pengelolaan respons belum selalu optimal saat umpan balik berdekatan atau anggaran belum eksplisit.

User Beliefs

Pada aspek perceived ease of use (PEU) dan perceived usefulness (PUS), seluruh informan menilai prototipe mudah digunakan dan memberi saran yang sesuai, misalnya menemukan laptop dari merek yang disukai atau membandingkan dua laptop gaming berdasarkan spesifikasi dan *benchmark*, sedangkan pada user

control (CTL) seluruh informan berhasil mengubah preferensi di tengah percakapan meliputi anggaran, merek, kualitas layar, dan peralihan kebutuhan, yang diikuti perubahan rekomendasi. Pada *transparency* (TRP), informan dapat menghubungkan rekomendasi dengan preferensi karena sistem menyertakan spesifikasi, kelebihan, keterbatasan, dan kesimpulan, sementara perbedaan paling jelas muncul pada CUI humanness (HUM), yaitu sebagian menilai interaksi terasa alami, satu informan menilai gaya bahasa masih seperti *chatbot*, dan satu informan bersikap netral karena mengutamakan kejelasan. Secara ringkas, dimensi *user beliefs* menunjukkan perceived ease of use, perceived usefulness, user control, dan *transparency* dinilai positif oleh seluruh informan, dengan CUI humanness sebagai satu-satunya aspek yang tidak seragam.

User Attitudes

Pada aspek *trust* and confidence (TRU), seluruh informan bersedia menjadikan rekomendasi prototipe sebagai acuan dalam memilih laptop, sebagian sebagai langkah awal menyusun preferensi dan sebagian menilai sistem dapat diandalkan, sedangkan pada satisfaction (SAT) kelima informan puas karena informasi dan kesimpulan dinilai cukup lengkap dan jelas, dengan catatan gaya bahasa yang masih terasa seperti *chatbot* bagi satu informan. Dengan demikian, dimensi *user attitudes* memperlihatkan *trust* and confidence serta satisfaction yang konsisten pada kelima informan, namun kepercayaan itu berbentuk acuan awal dalam memilih, bukan penyerahan keputusan pembelian sepenuhnya, dan catatan gaya bahasa menunjukkan ruang perbaikan pengalaman percakapan tanpa menghilangkan kepuasan keseluruhan.

Behavioral Intentions

Pada aspek intention to use (ITU), seluruh informan bersedia menggunakan prototipe pada kesempatan berikutnya, terutama saat membutuhkan laptop atau dimintai rekomendasi, dan satu informan bersedia merekomendasikannya kepada teman, sedangkan pada intention to purchase (ITP) seluruh informan akan mempertimbangkan produk yang direkomendasikan dengan keputusan yang tetap bergantung pada ketersediaan dana, kecocokan kebutuhan kuliah atau pekerjaan, dan kekuatan argumen sistem. Dimensi *behavioral intentions* dengan demikian menunjukkan intention to use dan intention to purchase yang positif, tetapi niat pembelian bersifat kondisional sehingga penilaian positif belum dapat dimaknai sebagai keputusan pembelian yang pasti.

Pembahasan

Pengembangan Prototipe Sistem Rekomendasi Percakapan

Dari sisi perancangan dan implementasi, prototipe dibangun sebagai *conversational recommender system* yang memadukan antarmuka WhatsApp, orkestrasi *workflow* n8n, CS Agent, *Recommendation Agent*, memori percakapan, tool search_laptops, dan katalog laptop pada PostgreSQL, dengan pembagian peran antara CS Agent yang menangani elisitasi, klarifikasi, dan kendali status serta

Recommendation Agent yang menyusun rekomendasi dan penjelasan sehingga merealisasikan arsitektur CRS yang menghubungkan pemrosesan masukan, pemodelan pengguna, pengelolaan dialog, dan mesin rekomendasi (Jannach et al., 2022). Keputusan desain utamanya adalah memisahkan kemampuan bahasa dari penentuan fakta produk: LLM memahami kebutuhan dan menyusun respons, sedangkan penyaringan berbasis batasan (Felfernig & Burke, 2008) dan pemeringkatan berbasis utilitas (Adomavicius & Kwon, 2015; Manouselis & Costopoulou, 2007) ditempatkan pada *subworkflow* deterministik agar fakta bersumber dari data dan aturan yang dapat diperiksa ulang, sementara WhatsApp dan n8n dipilih karena kanal familier dan orkestrasi low-code yang dapat dioperasikan self-hosted (McFeetors & Pant, 2022). Dibandingkan penelitian terdahulu yang menilai isyarat antropomorfik (Araujo, 2018) atau menekankan validitas katalog pada CRS berbasis LLM (Granada et al., 2023; Manzoor et al., 2024), kontribusinya terletak pada integrasi keduanya melalui lapisan rekomendasi deterministik dan evaluasi *grey-box*, meski kesiapan produksi berskala besar belum diuji.

Kinerja dan Keandalan Prototipe dalam Menghasilkan Rekomendasi

Dari sisi kinerja, prototipe dinilai dalam menghasilkan rekomendasi yang valid, relevan, dan sesuai kebutuhan, sejalan dengan pandangan bahwa evaluasi CRS menuntut penilaian objektif dan pengalaman pengguna secara berdampingan (Jannach, 2023). Berdasarkan Tabel 6 seluruh kasus memenuhi nilai harapan dan seluruh indikator yang berlaku mencapai 100 persen, melampaui kriteria pada Tabel 4, meski capaian ini bermakna pada sepuluh kasus *workflow* utama dan lima kasus *subworkflow* yang diuji dan bukan bukti seluruh variasi masukan tercakup: pada *workflow* utama kesesuaian status, kebutuhan, anggaran, filter, dan keputusan pemanggilan tool menunjukkan kendali dialog yang konsisten, sedangkan pengujian *subworkflow* memperlihatkan filter dan pemeringkatan bekerja sesuai *oracle* dengan tiga kandidat teratas identik dan berurutan sama, mengonfirmasi penyaringan pada Tabel 2 serta normalisasi dan agregasi pada Persamaan 1 dan 2, dan analisis sensitivitas menunjukkan urutan stabil terhadap gangguan bobot. Validitas katalog ditunjukkan oleh kesetiaan keluaran pada kandidat pencarian tanpa produk fiktif sehingga menekan risiko *hallucination* (Ji et al., 2023) dan meningkatkan keterlacakan melalui penautan pada sumber deterministik (Gao et al., 2023), yang mendukung efektivitas *grounding* pada kasus yang diuji dan bukan klaim bahwa halusinasi mustahil. Dari sisi metode, *grey-box testing* memungkinkan setiap keluaran ditelusuri ke status, kebutuhan, filter, dan urutan kandidat, melampaui *black-box* dan memberi bukti empiris pada konteks yang dalam literatur lebih sering konseptual (Coulter, 1999; Ehmer & Khan, 2012), sementara *oracle* terpisah pada *subworkflow* mengurangi pemeriksaan melingkar dan menjaga validitas pengujian (Wohlin et al., 2012), sehingga kinerja prototipe terbukti dengan keberlakuan dibatasi *dataset*, konfigurasi bobot, model bahasa, dan cakupan kasus.

Pengalaman dan Penerimaan Pengguna terhadap Prototipe

Dari sisi pengalaman pengguna, evaluasi dilakukan melalui wawancara terhadap lima informan setelah tugas konsultasi, perbandingan, dan perubahan preferensi, memakai instrumen yang mengadaptasi CRS-Que menjadi 18 pertanyaan terbuka dalam dimensi *perceived qualities*, *user beliefs*, *user attitudes*, dan *behavioral intentions* (Jin et al., 2024), sehingga hasilnya memahami pola pengalaman dan bukan mengestimasi populasi. Pada *perceived qualities*, penilaian positif atas accuracy, interaction adequacy, *explainability*, serta pemahaman CUI menunjukkan keluaran teknis yang benar umumnya menjadi pengalaman yang dapat dipahami, sementara novelty memperluas ruang pilihan dan relevan terhadap *choice overload* karena sistem menyaring serta menjelaskan alternatif agar ruang keputusan lebih terkelola (Chernev et al., 2015). Pada *user beliefs*, perceived ease of use, perceived usefulness, user control, dan *transparency* menunjukkan pengguna dapat menyampaikan dan mengubah preferensi tanpa kehilangan kendali sesuai posisi pra-pembelian, sedangkan pada *user attitudes* dan *behavioral intentions* seluruh informan menyatakan *trust* and confidence, satisfaction, intention to use, dan intention to purchase yang positif namun berupa kesediaan memakai rekomendasi sebagai acuan sehingga belum membuktikan pembelian aktual atau loyalitas. Perbedaan paling jelas muncul pada CUI humanness, dengan sebagian menilai interaksi alami dan satu masih merasakan gaya bahasa seperti *chatbot*, selaras dengan temuan bahwa pengaruh isyarat antropomorfik bergantung konteks (Adam et al., 2021; Araujo, 2018). Ketika hasil teknis dan pengguna dipertemukan, kesesuaian filter dan pemeringkatan sejajar dengan accuracy, kendali dengan user control, alasan dengan *explainability* dan *transparency*, memori dengan pemahaman CUI, dan kesetiaan kandidat dengan *trust* and confidence, meski hubungan ini interpretatif dan bukan sebab akibat, sehingga sisi pengguna dijawab melalui penerimaan positif disertai catatan pada pengelolaan respons dan CUI humanness pada lingkungan uji lima informan.

Keterbatasan dan Peluang Pengembangan Sistem

Keterbatasan dibedakan menjadi keterbatasan sistem dan keterbatasan bukti penelitian agar hasil positif tidak digeneralisasi berlebihan. Pada sisi katalog, *dataset* publik diperlakukan sebagai simulasi dan tidak merepresentasikan inventaris satu toko sedangkan harga dan stok tidak real-time; *grounding* memastikan jawaban setia pada data yang diberikan tetapi tidak menjamin data lengkap atau mutakhir, sehingga pengembangan perlu mengintegrasikan katalog nyata, pembaruan harga dan stok, penanda waktu, serta validasi data. Pada filter dan pemeringkatan, filter terlalu ketat dapat menghasilkan himpunan kosong sehingga *constraint* relaxation yang meminta persetujuan pengguna diperlukan, dan bobot heuristik yang stabil pada analisis sensitivitas tetap memerlukan validasi pakar, personalisasi, dan pembandingan dengan metode lain; pada percakapan dan model bahasa, respons dapat menumpuk, konteks lama berpotensi tidak sesuai setelah preferensi berubah, gaya bahasa belum dinilai alami oleh semua informan, dan sistem bergantung pada penyedia layanan tertentu serta integrasi multikomponen melalui n8n dan WAHA

yang menambah titik kegagalan, latensi, dan biaya, sehingga pengembangan diarahkan pada ringkasan konteks, versi preferensi, konfirmasi konflik, pengelolaan giliran, tingkat kerincian respons, dan eskalasi kepada petugas manusia. Dari sisi bukti, cakupan sepuluh kasus *workflow* utama dan lima kasus *subworkflow* mewakili partisi yang dirancang tetapi belum mencakup seluruh kombinasi filter, kegagalan layanan, keamanan, beban, dan konkurensi sehingga nilai 100 persen tidak dapat diekstrapolasi ke produksi, sedangkan lima informan sesuai untuk evaluasi formatif tetapi tidak mewakili populasi dan belum mengukur penggunaan jangka panjang atau pembelian aktual sehingga penelitian lanjutan perlu partisipan lebih beragam dan periode lebih panjang. Peluang lintas domain terbuka karena arsitektur memisahkan katalog, aturan, dan percakapan, meski perpindahan kategori memerlukan *dataset*, definisi atribut, filter, bobot, serta evaluasi baru, sehingga secara keseluruhan prototipe layak sebagai bukti konsep yang masih memerlukan penguatan pada katalog real-time, pengelolaan konteks, validasi pemeringkatan, perluasan pengujian, dan evaluasi bisnis nyata sebelum menjadi layanan komersial berskala besar.

V. KESIMPULAN DAN REKOMENDASI

Penelitian ini bertujuan merancang, mengimplementasikan, dan mengevaluasi prototipe AI agent layanan konsultasi berbasis WhatsApp untuk rekomendasi laptop pada platform n8n. Terkait proses perancangan dan implementasi, kebutuhan konsultasi diturunkan dari studi pendahuluan dan diwujudkan menjadi sistem rekomendasi percakapan yang memisahkan pengelolaan dialog pada *Customer Service Agent*, penyusunan rekomendasi pada *Recommendation Agent*, serta penyaringan dan pemeringkatan produk pada *subworkflow* deterministik, diorkestrasi melalui *workflow* low-code di n8n yang menghubungkan model bahasa, memori percakapan, alat pencarian, dan katalog, lalu dipadukan dengan WhatsApp melalui WAHA. Terkait kinerja, validitas terjaga karena setiap keluaran ditautkan pada kandidat hasil pencarian sehingga tidak keluar dari katalog, relevansi diperoleh dari penyaringan anggaran dan filter serta pemeringkatan multiatribut yang konsisten dengan perhitungan acuan, dan kesesuaian kebutuhan dihasilkan dari ekstraksi status, kebutuhan, anggaran, serta preferensi; cara kerja ini berjalan sesuai rancangan pada seluruh kasus yang diuji, meski hasil itu perlu dimaknai sebatas ruang skenario terkurasi. Terkait evaluasi, pengujian objektif dengan pendekatan *grey-box testing* terhadap *oracle* memperlihatkan perilaku konsisten dengan rancangan, sedangkan uji pengguna terhadap lima informan dengan instrumen CRS-Que menunjukkan penerimaan positif pada kualitas rekomendasi, kemudahan, kendali, dan transparansi, dengan catatan pada respons yang menumpuk dan kesan kemanusiaan antarmuka yang belum seragam. Studi ini juga menghasilkan dua temuan tambahan, yaitu kebaruan yang tetap tertambat pada katalog sehingga penemuan produk berjalan tanpa mengorbankan validitas, dan posisi sistem sebagai pendamping keputusan pra-pembelian, bukan pengambil keputusan yang menggantikan pertimbangan pengguna. Secara praktis, prototipe

berperan sebagai sistem pendukung keputusan bagi calon pembeli sekaligus antarmuka konsultasi pra-penjualan bagi peritel, dengan manfaat bisnis yang masih bersifat potensial karena latensi, biaya token, konversi, dan imbal hasil investasi belum diukur. Kesimpulan ini perlu dibaca dalam batasan penelitian, yaitu penggunaan *dataset* publik sebagai simulasi katalog dengan harga dan kurs tetap, domain terbatas pada laptop, pembobotan heuristik, cakupan kasus yang terbatas, dan evaluasi formatif dengan lima informan.

Berdasarkan temuan dan keterbatasan tersebut, beberapa rekomendasi diajukan. Bagi praktisi seperti peritel dan pengelola marketplace, prototipe disarankan diintegrasikan dengan katalog produk sebenarnya beserta pembaruan harga dan stok berkala, lalu diterapkan sebagai antarmuka konsultasi pra-penjualan dengan mekanisme eskalasi kepada petugas manusia untuk kasus kompleks, sementara pemanfaatan riwayat percakapan untuk memahami kebutuhan pasar perlu disertai persetujuan pengguna, anonimisasi, dan tata kelola akses yang memadai. Bagi regulator, disarankan aturan transparansi yang mewajibkan pengungkapan status promosi atau sponsor bila kelak dimasukkan ke rekomendasi, serta penguatan perlindungan data dan privasi percakapan agar rekomendasi berbantuan kecerdasan buatan tetap dipercaya. Bagi penelitian lanjutan, diperlukan integrasi katalog *real-time* dan kurs dinamis, validasi dan personalisasi bobot pemeringkatan, penerapan *constraint* relaxation yang meminta persetujuan pengguna, perbaikan pengelolaan konteks dan penyajian respons, serta pengujian beban, keamanan, perluasan cakupan kasus, dan evaluasi jangka panjang dengan partisipan lebih beragam, termasuk skema *human-in-the-loop* pada operasi layanan sesungguhnya.

REFERENSI

Abdelhalim, E., Anazodo, K. S., Gali, N., & Robson, K. (2024). A framework of diversity, equity, and inclusion safeguards for chatbots. Business Horizons, 67(5), 487–498. https://doi.org/10.1016/j.bushor.2024.03.003

Adam, M., Wessel, M., & Benlian, A. (2021). AI-based chatbots in customer service and their effects on user compliance. Electronic Markets, 31(2), 427–445. https://doi.org/10.1007/s12525-020-00414-7

Adomavicius, G., & Kwon, Y. (2015). Multi-Criteria Recommender Systems. In F. Ricci, L. Rokach, & B. Shapira (Eds.), Recommender Systems Handbook (pp. 847–880). Springer US. https://doi.org/10.1007/978-1-4899-7637-6_25

Ahn, J., Kim, J., & Sung, Y. (2022). The effect of gender stereotypes on artificial intelligence recommendations. Journal of Business Research, 141, 50–59. https://doi.org/10.1016/j.jbusres.2021.12.007

Araujo, T. (2018). Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent

and company perceptions. Computers in Human Behavior, 85, 183–189. https://doi.org/10.1016/j.chb.2018.03.051

Barr, E. T., Harman, M., McMinn, P., Shahbaz, M., & Yoo, S. (2015). The Oracle Problem in Software Testing: A Survey. IEEE Transactions on Software Engineering, 41(5), 507–525. https://doi.org/10.1109/TSE.2014.2372785

Barry, T. E. (1987). The Development of the Hierarchy of Effects: An Historical Perspective. Current Issues and Research in Advertising, 10(1–2), 251–295

Barry, T. E., & Howard, D. J. (1990). A Review and Critique of the Hierarchy of Effects in Advertising. International Journal of Advertising, 9(2), 121–135. https://doi.org/10.1080/02650487.1990.11107138

Beizer, B., & Wiley, J. (1996). Black Box Testing: Techniques for Functional Testing of Software and Systems. IEEE Software, 13(5), 98. https://doi.org/10.1109/MS.1996.536464

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). Language Models are Few-Shot Learners (Version 4). arXiv. https://doi.org/10.48550/ARXIV.2005.14165

Burke, R. (2002). Hybrid Recommender Systems: Survey and Experiments. User Modeling and User-Adapted Interaction, 12(4), 331–370. https://doi.org/10.1023/A:1021240730564

Cao, H., Chen, Z., Cheng, M., Zhao, S., Wang, T., & Li, Y. (2021). You Recommend, I Buy: How and Why People Engage in Instant Messaging Based Social Commerce. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW1), 1–25. https://doi.org/10.1145/3449141

Chen, L., & Pu, P. (2012). Critiquing-based recommenders: Survey and emerging trends. User Modeling and User-Adapted Interaction, 22(1–2), 125–150. https://doi.org/10.1007/s11257-011-9108-6

Chernev, A., Böckenholt, U., & Goodman, J. (2015). Choice overload: A conceptual review and meta-analysis. Journal of Consumer Psychology, 25(2), 333–358. https://doi.org/10.1016/j.jcps.2014.08.002

Christakopoulou, K., Radlinski, F., & Hofmann, K. (2016). Towards Conversational Recommender Systems. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 815–824. https://doi.org/10.1145/2939672.2939746

Chung, M., Ko, E., Joung, H., & Kim, S. J. (2020). Chatbot e-service and customer satisfaction regarding luxury brands. Journal of Business Research, 117, 587–595. https://doi.org/10.1016/j.jbusres.2018.10.004

Coulter, A. C. (1999). Graybox software testing methodology: Embedded software testing technique. Gateway to the New Millennium. 18th Digital Avionics Systems Conference. Proceedings (Cat. No.99CH37033), B.6-6 2, 10.A.5-1-10.A.5-8. https://doi.org/10.1109/DASC.1999.822089

Ehmer, M., & Khan, F. (2012). A Comparative Study of White Box, Black Box and Grey Box Testing Techniques. International Journal of Advanced Computer Science and Applications, 3(6). https://doi.org/10.14569/IJACSA.2012.030603

Felfernig, A., & Burke, R. (2008). Constraint-based recommender systems: Technologies and research issues. Proceedings of the 10th International Conference on Electronic Commerce, 1–10. https://doi.org/10.1145/1409540.1409544

Ferrari Dacrema, M., Boglio, S., Cremonesi, P., & Jannach, D. (2021). A Troubling Analysis of Reproducibility and Progress in Recommender Systems Research. ACM Transactions on Information Systems, 39(2), 1–49. https://doi.org/10.1145/3434185

Frank, D.-A., Chrysochou, P., Mitkidis, P., Otterbring, T., & Ariely, D. (2024). Navigating uncertainty: Exploring consumer acceptance of artificial intelligence under self-threats and high-stakes decisions. Technology in Society, 79, 102732. https://doi.org/10.1016/j.techsoc.2024.102732

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, M., & Wang, H. (2023). Retrieval-Augmented Generation for Large Language Models: A Survey (Version 5). arXiv. https://doi.org/10.48550/ARXIV.2312.10997

Granada, M. G., Zilbershtein, D., Odijk, D., & Barile, F. (2023). VideolandGPT: A User Study on a Conversational Recommender System (arXiv:2309.03645). arXiv. https://doi.org/10.48550/arXiv.2309.03645

Guest, G., Bunce, A., & Johnson, L. (2006). How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability. Field Methods, 18(1), 59–82. https://doi.org/10.1177/1525822X05279903

He, Z., Xie, Z., Jha, R., Steck, H., Liang, D., Feng, Y., Majumder, B. P., Kallus, N., & Mcauley, J. (2023). Large Language Models as Zero-Shot Conversational Recommenders. Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, 720–730. https://doi.org/10.1145/3583780.3614949

Hsu, C.-L., & Lin, J. C.-C. (2023). Understanding the user satisfaction and loyalty of customer service chatbots. Journal of Retailing and Consumer Services, 71, 103211. https://doi.org/10.1016/j.jretconser.2022.103211

Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B., & Liu, T. (2023). A Survey on Hallucination in Large Language Models:

Principles, Taxonomy, Challenges, and Open Questions. https://doi.org/10.48550/ARXIV.2311.05232

Huang, Y., Sun, Z., Pilot, A., Zhao, H., Wang, Z., Huang, Y., Sun, Z., Pilot, A., Zhao, H., & Wang, Z. (2022). The Effect of Instant Messaging Social Media Platform Characteristics on Consumers' Purchase Intention: An Empirical Study of WeChat. 系统科学与信息学报(英文), 10(1), 65–83. https://doi.org/10.21078/JSSI-2022-065-19

International Software Testing Qualifications Board (ISTQB). (2024, October 19). Diakses 22 Juli 2026, dari https://istqb.org/

Iwamoto, A. (2023). The Origin of AIDA: Who Invented and Formulated the AIDA model?

Jahan, A., Bahraminasab, M., & Edwards, K. L. (2012). A target-based normalization technique for materials selection. Materials & Design, 35, 647–654. https://doi.org/10.1016/j.matdes.2011.09.005

Jannach, D. (2023). Evaluating conversational recommender systems: A landscape of research. Artificial Intelligence Review, 56(3), 2365–2400. https://doi.org/10.1007/s10462-022-10229-x

Jannach, D., Manzoor, A., Cai, W., & Chen, L. (2022). A Survey on Conversational Recommender Systems. ACM Computing Surveys, 54(5), 1–36. https://doi.org/10.1145/3453154

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of Hallucination in Natural Language Generation. ACM Computing Surveys, 55(12), 1–38. https://doi.org/10.1145/3571730

Jin, Y., Chen, L., Cai, W., & Zhao, X. (2024). CRS-Que: A User-centric Evaluation Framework for Conversational Recommender Systems. ACM Transactions on Recommender Systems, 2(1), 1–34. https://doi.org/10.1145/3631534

Kim, J., Kim, J. H., Kim, C., & Park, J. (2023). Decisions with ChatGPT: Reexamining choice overload in ChatGPT recommendations. Journal of Retailing and Consumer Services, 75, 103494. https://doi.org/10.1016/j.jretconser.2023.103494

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks (Version 4). arXiv. https://doi.org/10.48550/ARXIV.2005.11401

Luo, X., Tong, S., Fang, Z., & Qu, Z. (2019). Frontiers: Machines vs. Humans: The Impact of Artificial Intelligence Chatbot Disclosure on Customer Purchases. Marketing Science, mksc.2019.1192. https://doi.org/10.1287/mksc.2019.1192

Maddalena, S. (2025, February 5). Digital 2025. We Are Social Indonesia. Diakses 22 Juli 2026, dari https://wearesocial.com/id/blog/2025/02/digital-2025/

Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to Information Retrieval (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511809071

Manouselis, N., & Costopoulou, C. (2007). Analysis and Classification of Multi-Criteria Recommender Systems. World Wide Web, 10(4), 415–441. https://doi.org/10.1007/s11280-007-0019-8

Manzoor, A., & Jannach, D. (2022). Towards retrieval-based conversational recommendation. Information Systems, 109, 102083. https://doi.org/10.1016/j.is.2022.102083

Manzoor, A., Ziegler, S. C., Garcia, K. Maria. P., & Jannach, D. (2024). ChatGPT as a Conversational Recommender System: A User-Centric Analysis. Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization, 267–272. https://doi.org/10.1145/3627043.3659574

Mari, A., Mandelli, A., & Algesheimer, R. (2024). Empathic voice assistants: Enhancing consumer responses in voice commerce. Journal of Business Research, 175, 114566. https://doi.org/10.1016/j.jbusres.2024.114566

McFeetors, J., & Pant, T. (2022). Rapid product development with n8n: Practical guide to creating digital products on the web using workflow automation and n8n. Packt Publishing.

Mochammad Rafi Diaz Ardhana & Aries Dwi Indriyanti. (2025). Web-Based Decision Support System for Best Laptop Selection Using MABAC Method. Journal of Emerging Information Systems and Business Intelligence (JEISBI), 5(4). https://doi.org/10.26740/jeisbi.v5i4.64267

Models. (n.d.). OpenRouter. Diakses 22 Juli 2026, dari https://openrouter.ai/models

Models | Gemini API | Google AI for Developers. (n.d.). Diakses 22 Juli 2026, dari https://ai.google.dev/gemini-api/docs/models

Myers, G. J., Sandler, C., & Badgett, T. (2012). The art of software testing (3rd ed). John Wiley & Sons. https://doi.org/10.1002/9781119202486

Nyazabe, S., Lee, C., & Vololoniaina, L. (2025). Sns-Based Exposure Influence on Consumers' Purchasing Behavior: The Evidence from WhatsApp. SSRN. https://doi.org/10.2139/ssrn.5082139

PassMark Software—PC Benchmark and Test Software. (n.d.). Diakses 10 Juli 2026, dari https://www.passmark.com

Peng, Q., Liu, H., Huang, H., Yang, J., Yang, Q., & Shao, M. (2025). A Survey on LLM-powered Agents for Recommender Systems.

Pressman, R. S. (2001). Software engineering: A practitioner's approach (5. ed., 20. anniversary ed). McGraw Hill.

Rese, A., Ganster, L., & Baier, D. (2020). Chatbots in retailers' customer communication: How to measure their acceptance? Journal of Retailing and Consumer Services, 56, 102176. https://doi.org/10.1016/j.jretconser.2020.102176

Runeson, P., & Höst, M. (2009). Guidelines for conducting and reporting case study research in software engineering. Empirical Software Engineering, 14(2), 131–164. https://doi.org/10.1007/s10664-008-9102-8

Shaalan, A., Tourky, M., & Ibrahim, K. (2024). AI Caramba!: The Negative Effects of AI Agents in Customer Relationship Management. In J. D. Santos, P. B. Pires, & N. Grigoriou (Eds.), Advances in Marketing, Customer Relationship Management, and E-Services (pp. 309–352). IGI Global. https://doi.org/10.4018/979-8-3693-5340-0.ch011

Sheehan, B., Jin, H. S., & Gottlieb, U. (2020). Customer service chatbots: Anthropomorphism and adoption. Journal of Business Research, 115, 14–24. https://doi.org/10.1016/j.jbusres.2020.04.030

Simon, H. A. (1955). A Behavioral Model of Rational Choice. The Quarterly Journal of Economics, 69(1), 99. https://doi.org/10.2307/1884852

Sommerville, I. (2016). Software engineering (Tenth edition). Pearson.

Sugiyono. (2018). Metode Penelitian Kuantitatif, Kualitatif, dan R&D (2nd ed). ALFABETA, cv.

Taherdoost, H., & Madanchian, M. (2023). Multi-Criteria Decision Making (MCDM) Methods and Concepts. Encyclopedia, 3(1), 77–87. https://doi.org/10.3390/encyclopedia3010006

The State of AI: Global Survey 2025 | McKinsey. (n.d.). Diakses 22 Juli 2026, dari https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai

Tintarev, N., & Masthoff, J. (2015). Explaining Recommendations: Design and Evaluation. In F. Ricci, L. Rokach, & B. Shapira (Eds.), Recommender Systems Handbook (pp. 353–382). Springer US. https://doi.org/10.1007/978-1-4899-7637-6_10

Triantaphyllou, E., & Sánchez, A. (1997). A Sensitivity Analysis Approach for Some Deterministic Multi-Criteria Decision-Making Methods*. Decision Sciences, 28(1), 151–194. https://doi.org/10.1111/j.1540-5915.1997.tb01306.x

Trzebiński, W., Konuk, F. A., & Marciniak, B. (2025). Is AI suitable to speak about morality? Consumer response to AI-generated moral appeals in the case of fair trade products. Journal of Business Research, 199, 115487. https://doi.org/10.1016/j.jbusres.2025.115487

Uta, M., Felfernig, A., Le, V.-M., Tran, T. N. T., Garber, D., Lubos, S., & Burgstaller, T. (2024). Knowledge-based recommender systems: Overview and research directions. Frontiers in Big Data, 7, 1304439. https://doi.org/10.3389/fdata.2024.1304439

Vafaei, N., Ribeiro, R. A., & Camarinha-Matos, L. M. (2022). Assessing Normalization Techniques for Simple Additive Weighting Method. Procedia Computer Science, 199, 1229–1236. https://doi.org/10.1016/j.procs.2022.01.156

Van Solingen, R., Basili, V., Caldiera, G., & Rombach, H. D. (2002). Goal Question Metric (GQM) Approach. In J. J. Marciniak (Ed.), Encyclopedia of Software Engineering (1st ed.). Wiley. https://doi.org/10.1002/0471028959.sof142

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2023). Attention Is All You Need (arXiv:1706.03762). arXiv. https://doi.org/10.48550/arXiv.1706.03762

Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., Chen, Z., Tang, J., Chen, X., Lin, Y., Zhao, W. X., Wei, Z., & Wen, J. (2024). A survey on large language model based autonomous agents. Frontiers of Computer Science, 18(6), 186345. https://doi.org/10.1007/s11704-024-40231-1

Wang, Y.-M., & Luo, Y. (2009). On rank reversal in decision analysis. Mathematical and Computer Modelling, 49(5–6), 1221–1229. https://doi.org/10.1016/j.mcm.2008.06.019

Wierzbicki, A. P. (1982). A mathematical basis for satisficing decision making. Mathematical Modelling, 3(5), 391–405. https://doi.org/10.1016/0270-0255(82)90038-0

Winterfeldt, D. von, & Edwards, W. (1993). Decision analysis and behavioral research (Repr). Cambridge Univ. Press.

Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., & Wesslén, A. (2012). Experimentation in Software Engineering. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-29044-2

Xhema, J. (2019). Effect of Social Networks on Consumer Behaviour: Complex Buying. IFAC-PapersOnLine, 52(25), 504–508. https://doi.org/10.1016/j.ifacol.2019.12.594

Xi, Z., Chen, W., Guo, X., He, W., Ding, Y., Hong, B., Zhang, M., Wang, J., Jin, S., Zhou, E., Zheng, R., Fan, X., Wang, X., Xiong, L., Zhou, Y., Wang, W., Jiang, C., Zou, Y., Liu, X., … Gui, T. (2025). The rise and potential of large language model based agents: A survey. Science China Information Sciences, 68(2), 121101. https://doi.org/10.1007/s11432-024-4222-0

Zhang, Y., & Chen, X. (2020). Explainable Recommendation: A Survey and New Perspectives. Foundations and Trends® in Information Retrieval, 14(1), 1–101. https://doi.org/10.1561/1500000066

Zhang, Y., Chen, X., Ai, Q., Yang, L., & Croft, W. B. (2018). Towards Conversational Search and Recommendation: System Ask, User Respond. Proceedings of the 27th ACM International Conference on Information and Knowledge Management, 177–186. https://doi.org/10.1145/3269206.3271776

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., … Wen, J.-R. (2023). A Survey of Large Language Models (Version 19). arXiv. https://doi.org/10.48550/ARXIV.2303.18223

Ziegfeld, L., Scala, D. D., & Cremers, A. H. M. (2025). The effect of preference elicitation methods on the user experience in conversational recommender systems. Computer Speech & Language, 89, 101696. https://doi.org/10.1016/j.csl.2024.101696