BAB V.
KESIMPULAN DAN SARAN

5.1. Kesimpulan

Penelitian ini menunjukkan bahwa *Few-shot Prompting* unggul tipis dibandingkan *Zero-shot Prompting* dengan total skor rata-rata 88% berbanding 84%. Ditinjau dari kelima aspek, keunggulan *Few-shot* tidak bersifat menyeluruh: metode ini unggul signifikan pada konsistensi dan *robustness* (masing-masing 20%) karena kemampuannya mereplikasi format dan detail jawaban sesuai contoh percakapan CS, namun justru tertinggal pada penalaran (20%) akibat satu kontradiksi internal terkait kebijakan pengiriman luar kota.

Pada aspek akurasi dan penanganan pertanyaan repetitif, kedua metode menunjukkan performa yang setara (100%), mengindikasikan bahwa kedua kemampuan tersebut lebih dipengaruhi oleh kelengkapan *knowledge base* daripada oleh ada-tidaknya contoh percakapan pada *Response Prompt*. Dengan demikian, pemilihan antara *Zero-shot* dan *Few-shot Prompting* untuk kasus serupa sebaiknya mempertimbangkan konteks penggunaan: *Few-shot* lebih sesuai ketika konsistensi format jawaban dan kepatuhan terhadap detail spesifik perusahaan menjadi prioritas, sedangkan *Zero-shot* tetap menjadi pilihan yang wajar ketika efisiensi biaya token dan kecepatan implementasi tanpa proses kurasi contoh menjadi pertimbangan utama.

Berdasarkan keseluruhan hasil tersebut, *Few-shot Prompting* sangat direkomendasikan untuk implementasi *chatbot customer service* pada perusahaan dengan kebijakan spesifik, vocabulary teknis industri, dan kebutuhan konsistensi komunikasi brand yang tinggi.

5.2. Saran

Berdasarkan temuan dan keterbatasan penelitian, beberapa saran dikemukakan untuk pengembangan ke depan. Knowledge base perlu diperkaya dan diperbarui secara berkala khususnya mengenai informasi promo, estimasi pengiriman per rute, dan kebijakan dinamis perusahaan sambil memantau log percakapan *chatbot* secara rutin sebagai dasar penambahan contoh percakapan *Few-shot* yang lebih relevan. Selain itu, perlu dipertimbangkan penambahan

mekanisme pengalihan otomatis yang lebih tegas agar pelanggan dapat segera diarahkan ke tim CS manusia ketika *chatbot* mendeteksi pertanyaan di luar knowledge base. Untuk penelitian lanjutan, eksplorasi teknik *prompting* yang lebih mendalam seperti Chain-of-Thought (CoT) *Prompting* atau kombinasi *Few-shot* dengan *Retrieval-Augmented Generation (RAG)* berpotensi meningkatkan kemampuan penalaran pada pertanyaan yang lebih kompleks. Terakhir, replikasi penelitian ini pada perusahaan distribusi lain dengan karakteristik produk dan kebijakan yang berbeda akan memperkuat generalisabilitas temuan yang telah diperoleh.

DAFTAR PUSTAKA

[1] CHIDIOGO C. NWOKEDI and CHIDINMA A. NWAFOR, "Enhancing customer service and user experience through the use of machine learning powered intelligent chatbots," *World J. Adv. Res. Rev.*, vol. 23, no. 2, pp. 181–191, 2024, doi: 10.30574/wjarr.2024.23.2.2307.

[2] Abel Uzoka, Emmanuel Cadet, and Pascal Ugochukwu Ojukwu, "Leveraging AI-Powered chatbots to enhance customer service efficiency and future opportunities in automated support," *Comput. Sci. IT Res. J.*, vol. 5, no. 10, pp. 2485–2510, 2024, doi: 10.51594/csitrj.v5i10.1676.

[3] P. B. Brandtz, A. Jarvis, S. Poncho, J. Humani, and G. Assistant, "Chatbots and the New World of HCI," vol. 24, no. 4, pp. 38–42, 2016, doi: 10.1145/3085558.Chatbots.

[4] S. K. Dam, C. S. Hong, Y. Qiao, and C. Zhang, "A Complete Survey on LLM-based AI Chatbots," pp. 1–23, 2024, [Online]. Available: http://arxiv.org/abs/2406.16937

[5] Sudeep Meduri, "Revolutionizing Customer Service : The Impact of Large Language Models on Chatbot Performance," *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, vol. 10, no. 5, pp. 721–730, 2024, doi: 10.32628/cseit241051057.

[6] P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi, and G. Neubig, "Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing," *ACM Comput. Surv.*, vol. 55, no. 9, pp. 1–46, 2023, doi: 10.1145/3560815.

[7] L. Reynolds and K. McDonell, "Prompt Programming for Large Language Models: Beyond the *Few-Shot* Paradigm," *Conf. Hum. Factors Comput. Syst. - Proc.*, 2021, doi: 10.1145/3411763.3451760.

[8] T. Kojima, S. S. Gu, M. Reid, Y. Matsuo, and Y. Iwasawa, "Large Language Models are *Zero-Shot* Reasoners," *Adv. Neural Inf. Process. Syst.*, vol. 35, no. NeurIPS, 2022.

[9] S. Min *et al.*, "Rethinking the Role of Demonstrations: What Makes In-Context Learning Work?," *Proc. 2022 Conf. Empir. Methods Nat. Lang. Process. EMNLP 2022*, no. 2, pp. 11048–11064, 2022, doi: 10.18653/v1/2022.emnlp-main.759.

[10] T. B. Brown *et al.*, "Language Models are *Few-Shot* Learners," vol. 33, pp. 1877–1901, 2020.

[11] J. Wei *et al.*, "Finetuned Language Models Are *Zero-Shot* Learners," *ICLR 2022 - 10th Int. Conf. Learn. Represent.*, pp. 1–46, 2022.

[12] G. Ramesh *et al.*, "A review on NLP *zero-shot* and *few-shot* learning: methods and applications," *Discov. Appl. Sci.*, vol. 7, no. 9, 2025, doi: 10.1007/s42452-025-07225-5.

[13] Y. K. Dwivedi *et al.*, "Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and

policy," *Int. J. Inf. Manage.*, vol. 57, 2021, doi: 10.1016/j.ijinfomgt.2019.08.002.

[14] H. Meshkin, "Harnessing large language models ' *zero-shot* and *few-shot* learning capabilities for regulatory research," vol. 25, no. 5, 2024.

[15] S. Sivarajkumar, M. Kelley, A. Samolyk-Mazzanti, S. Visweswaran, and Y. Wang, "An Empirical Evaluation of Prompting Strategies for Large Language Models in *Zero-Shot* Clinical Natural Language Processing: Algorithm Development and Validation Study," *JMIR Med. Informatics*, vol. 12, pp. 1–14, 2024, doi: 10.2196/55318.

[16] I. Haromain, S. Munir, and A. Rahmah, "Analisa Prompt Engineering pada Large Language Model dengan Retrieval-Augmented Generation untuk Informasi Obat dan Vitamin," *Indones. J. Comput. Sci.*, vol. 4, no. 2, pp. 144–153, 2025, doi: 10.31294/ijcs.v4i2.10005.

[17] S. Schulhoff *et al.*, "The Prompt Report: A Systematic Survey of Prompt Engineering Techniques," 2025, [Online]. Available: http://arxiv.org/abs/2406.06608

[18] Y. Gao *et al.*, "Retrieval-Augmented Generation for Large Language Models : A Survey," pp. 1–21, 2023.

[19] X. Ye and G. Durrett, "The Unreliability of Explanations in *Few-shot* Prompting for Textual Reasoning," *Adv. Neural Inf. Process. Syst.*, vol. 35, no. NeurIPS, pp. 1–15, 2022.

[20] P. Santoso Lukman, "Mengoptimalkan Proses Pembersihan Data dalam Analisis Big Data Menggunakan Pipeline Berbasis AI," *Elkom J. Elektron. dan Komput.*, vol. 17, no. 2, pp. 657–666, 2024, [Online]. Available: https://journal.stekom.ac.id/index.php/elkom/article/view/2311

[21] Q. Xie, X. Ma, Z. Dai, and E. Hovy, "An interpretable knowledge transfer model for knowledge base completion," *ACL 2017 - 55th Annu. Meet. Assoc. Comput. Linguist. Proc. Conf. (Long Pap.*, vol. 1, pp. 950–962, 2017, doi: 10.18653/v1/P17-1088.

[22] L. Pawlik, "How the Choice of LLM and Prompt Engineering Affects Chatbot Effectiveness," *Electron.*, vol. 14, no. 5, pp. 1–16, 2025, doi: 10.3390/electronics14050888.

[23] Q. Zhao, R. Wang, Y. Cen, D. Zha, S. Tan, and J. Tang, "Pref RAG : Preference-Driven Multi-Source Retrieval Augmented Generation," 2020.

[24] Y. Wang *et al.*, "KEPT : Knowledge-Enhanced Prediction of Trajectories from Consecutive Driving Frames with Vision-Language Models".

[25] T. Petter, "How to use Large Language Models for Text Analysis," pp. 1–6, 2023.

[26] I. Azimi, M. Qi, L. Wang, A. M. Rahmani, and Y. Li, "Accuracy and Consistency of LLMs in the Registered Dietitian Exam : The Impact of Prompt Engineering and Knowledge Retrieval," 2024.

[27] G. Caldarini and S. Jaf, "A Literature Survey of Recent Advances in Chatbots,"

2022.

[28] J. Nielsen, “Why You Only Need to Test with 5 Users,” 2000. https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/

[29] L. Faulkner, “Beyond the five-user assumption : Benefits of increased sample sizes in usability testing,” vol. 35, no. 3, pp. 379–383, 2003.

[30] I. Etikan, S. A. Musa, and R. S. Alkassim, “Comparison of Convenience Sampling and Purposive Sampling,” vol. 5, no. 1, pp. 1–4, 2016, doi: 10.11648/j.ajtas.20160501.11.

[31] N. Radziwill and M. Benton, “Evaluating Quality of Chatbots and Intelligent Conversational Agents,” 2016.

[32] W. Ye *et al.*, “Assessing Hidden Risks of LLMs : An Empirical Study on Robustness , Consistency , and Credibility,” pp. 1–29.

[33] R. Zhao, Y. Liu, L. Altinger, H. Schütze, and M. A. Hedderich, “Evaluating Robustness of Large Language Models Against Multilingual Typographical Errors,” 2026.

[34] H. Arsham, “Excel For Statistical Data Analysis,” 2020.

[35] D. Hamdhana, “A low code approach to q&a on care records using flowise ai with llm integration and rag method,” vol. 9, no. 4, pp. 2435–2445, 2024.

[36] S. Edirimannage, C. Elvitigala, K. K. Don, W. Daluwatta, P. Wijesekara, and I. Khalil, “Developers Are Victims Too : A Comprehensive Analysis of The VS Code Extension Ecosystem,” vol. 7, pp. 95–106, 2021.

[37] T. K. Frey and B. S. Bloch, “Using Microsoft Teams to Facilitate Asynchronous Online Focus Groups,” vol. 22, pp. 1–15, 2023, doi: 10.1177/16094069231211251.

[38] V. R. N and N. S. Harinarayana, “Online Survey Tools : A Case Study of Google Forms 1 Online Vs . Other Survey Tools,” no. 1999.

[39] R. E. L. Hamdani, T. Bonald, F. Malliaros, N. Holzenberger, and F. Suchanek, “The Factuality of Large Language Models in the Legal Domain,” 2024.

LAMPIRAN

Transkrip Wawancara Pengumpulan Data (Teams)

Evan: Oke selamat siang Kawan, jadi kali ini aku mau wawancara kawan buat data *chatbot* yang lagi aku buat. Jadi nanti knowledge-nya bisa menjawab pertanyaan-pertanyaan dari customer. Sebelum itu aku mau nanya sedikit, PT. Samudra ini bergerak di bidang apa dan lokasinya di mana? Boleh dijelasin sedikit?

Ridwan: Jadi PT. Samudra itu distributor alat pertukangan, perkakas, sama sanitary. Kalau kantornya kita ada di Komplek Pergudangan Modern Business Park, Jalan KH. Hasyim Asy'ari, Tangerang, Banten. Dan kita itu udah lebih dari 20 tahun berdirinya. Terus kita punya 2 merek unggulan, namanya OWNER dan YO-ZURI.

Evan: Wah udah lama juga ya, 20 tahun. Nah terus untuk produk-produk dari OWNER sama YO-ZURI itu contohnya jual apa aja?

Ridwan: Sebenarnya kita punya 4 merek. Yang tadi itu produk unggulannya, jadi selain OWNER dan YO-ZURI, ada juga HELIOS sama Smart Choice. Nah produknya itu macam-macam, ada peralatan cat, perkakas tangan kayak mata bor, mata gerinda, kunci pintu, sanitary, sampai alat-alat safety atau keselamatan juga.

Evan: Oke, nah yang lebih menonjol kan OWNER sama YO-ZURI. Nah keduanya kan sama-sama jual kuas cat, bedanya apa ya?

Ridwan: Kalau YO-ZURI itu kelas atas, kalau OWNER itu kelas sedang. Bedanya lumayan banyak sebenarnya. OWNER itu merek standar, harganya lebih terjangkau tapi cuma cocok buat cat air atau cat tembok. Kalau YO-ZURI itu produk premium, bulunya lebih tebal dan ada campuran kulit binatang juga, jadi hasilnya lebih rata dan lebih awet. Kuas YO-ZURI juga bisa dipakai buat cat epoxy atau cat minyak. Jadi kalau ada pelanggan yang mau pakai cat epoxy, kita saranin YO-ZURI, karena kalau OWNER itu memang khusus untuk cat air, tidak disarankan buat cat epoxy karena takutnya agak rontok.

Evan: Jadi disesuaikan dengan kebutuhannya ya. Nah terus ada kunci pintu juga, OWNER sama YO-ZURI dua-duanya jualan kunci pintu juga?

Ridwan: Betul, dua-duanya jual kunci pintu. Tapi kemarin OWNER sempat discontinue, udah nggak ada lagi. Sekarang kita fokus ke kunci pintu YO-ZURI aja karena dia kelas premium, variasinya lebih lengkap dan ada garansinya juga.

Evan: Oke noted. Nah terus misalnya ada yang mau beli, cara ordernya gimana?

Ridwan: Kalau misalkan ada yang mau beli bisa langsung telepon ke nomor kita di (021) 55711055. Kita ada WhatsApp-nya juga, terus ada email juga yaitu cssamudra3@gmail.com. Kalau produknya ready, *customer service* kita langsung bikin invoice sesuai pesanan customer. Kalau misalkan nggak ready, kita buatkan PO ke supplier untuk pesanan khusus.

Evan: Oh jadi lengkap ya cara pesannya. Terus ada minimal pembelian nggak?

Ridwan: Minimal pembelian pasti ada. Minimal beli itu per kotak karena nggak bisa satuan, kan kita packagingan. Isi per kotaknya tergantung produk, ada yang isinya 6 pcs ada yang 12 pcs. Kita emang nggak buka kotak buat jual satuan biar stok kita nggak berantakan di gudang dan tetap rapi. Tapi kalau pelanggan pengen beli satuan, kita punya toko cabang namanya Bakti Jaya, di sana bisa jual eceran juga.

Evan: Oh ada toko cabangnya juga ya. Oke terus untuk harga, ada perbedaan nggak antara yang beli banyak atau beli retail biasa?

Ridwan: Perbedaan harga pasti ada. Kita punya beberapa kategori harga, ada 3 kategori. Perorangan itu harga paling mahal, Toko atau Grosir lebih murah, terus Agen atau Partai itu paling murah karena belinya dalam volume yang lebih banyak. Intinya makin banyak yang dibeli, makin murah harga per pcs-nya. Selisihnya bisa 2 sampai 15% antar kategori. Terus khusus yang bayar cash, OWNER dan YO-ZURI dapat diskon 5%, produk lainnya seperti HELIOS dan Smart Choice ada yang 3% tergantung mereknya.

Evan: Oke lumayan. Terus kalau di *chatbot*-nya nanti, harganya dikasih tahu ke pelanggan atau dijadiin privasi dan diarahkan ke WhatsApp CS?

Ridwan: Karena harga kita selalu update terus, jadi sebaiknya diarahkan ke CS aja biar pelanggan dapat harga yang paling akurat dan terkini.

Evan: Oke. Terus kalau kirim ke luar kota bisa nggak?

Ridwan: Bisa, kita kirim ke seluruh Indonesia. Ada 3 pilihan metode. Pertama Reguler pakai armada kantor, estimasinya H+1 paling cepat sampai H+3 atau H+4 tergantung rute, tapi pengirimannya cuma sampai ekspedisi, bukan door-to-door. Kedua Ambil Langsung ke gudang kita di Tangerang. Ketiga kalau urgent bisa pesan via Lalamove tapi ongkirnya ditanggung pembeli dan cuma bisa buat area yang dijangkau Lalamove. Kalau same-day ke luar kota jauh kayak Surabaya itu nggak bisa.

Evan: Oke oke. Nah terus untuk barang, kalau ada kerusakan ada garansi nggak? Dan bagaimana prosedurnya?

Ridwan: Sebenarnya nggak semua produk ada garansinya. Kunci Pintu YO-ZURI ada garansi resmi, ada jaminan penggantian. HELIOS bisa ditukar produk baru asal masih dalam 1 tahun pembelian. Buat OWNER dan YO-ZURI yang lain belum ada garansi tertulis, tapi bisa retur kalau ada kerusakan asal ada bukti foto atau video. Kalau merek lain seperti Smart Choice, kita konfirmasi dulu ke supplier, kemungkinan bisa retur atau diservis.

Evan: Terus untuk komplain prosedurnya gimana?

Ridwan: Komplain itu sebenarnya gampang. Yang pertama siapkan foto atau video bukti kerusakannya, terus langsung hubungi *customer service* kita via WhatsApp. Nanti CS yang koordinasi langsung sama supplier. Solusinya bisa diganti barang baru atau diperbaiki, karena kita cek dulu rusaknya separah apa.

Evan: Nah pertanyaan terakhir, kalau pelanggan mau lihat katalog produk, ada buku atau website-nya nggak?

Ridwan: Ada, kita punya e-katalog PDF yang bisa diunduh gratis di website. Ada katalog OWNER dan YO-ZURI. *Customer service* juga bisa jelasin produk langsung via telepon atau chat, dan bisa kasih rekomendasi produk sesuai kebutuhan pelanggan. Katalognya juga selalu di-update.

Evan: Oke, udah banyak informasi dari kawan. Mungkin cukup sampai di sini aja. Makasih banyak ya sudah meluangkan waktunya!

Ridwan: Iya, mantap. Sama-sama!

Dokumentasi Pengumpulan data + Pengujian responden

https://studentsuajyac-my.sharepoint.com/my?id=%2Fpersonal%2F221712040%5Fstudents%5Fuajy%5Fac%5Fid%2FDocuments%2FWAWANCARA%2BUJICOBA&ga=1

Kuesioner Evaluasi SUS

Petunjuk Pengisian: Berilah tanda (✓) pada kolom yang sesuai dengan penilaian Anda. Skala penilaian: 1 = Sangat Tidak Setuju, 2 = Tidak Setuju, 3 = Netral, 4 = Setuju, 5 = Sangat Setuju.

NO	Pernyataan	1	2	3	4	5
1	Saya pikir saya akan sering menggunakan *chatbot* ini					
2	Saya merasa *chatbot* ini tidak perlu dibuat serumit ini					
3	Saya pikir *chatbot* ini mudah digunakan					
4	Saya membutuhkan bantuan orang teknis untuk menggunakan *chatbot* ini					
5	Saya merasa berbagai fungsi *chatbot* ini terintegrasi dengan baik					
6	Saya rasa terlalu banyak ketidakkonsistenan pada *chatbot* ini					
7	Saya yakin orang lain akan cepat belajar menggunakan *chatbot* ini					
8	Saya merasa *chatbot* ini sangat rumit untuk digunakan					
9	Saya merasa percaya diri menggunakan *chatbot* ini					

10	Saya perlu belajar banyak hal sebelum bisa menggunakan *chatbot* ini					

Kuesioner Evaluasi Kualitas *Chatbot*

Petunjuk Pengisian: Berilah tanda (✓) pada kolom yang sesuai dengan penilaian Anda. Skala penilaian: 1 = Sangat Tidak Setuju, 2 = Tidak Setuju, 3 = Netral, 4 = Setuju, 5 = Sangat Setuju.

Dimensi 1 Akurasi

NO	Pernyataan	1	2	3	4	5
A1	*Chatbot* memberikan informasi yang BENAR tentang produk yang dijual PT. Samudra Teknik Sejahtera					
A2	*Chatbot* memberikan informasi yang BENAR tentang cara pemesanan produk.					
A3	*Chatbot* memberikan informasi yang BENAR tentang estimasi waktu pengiriman					
A4	*Chatbot* memberikan informasi yang BENAR tentang kebijakan garansi produk					
A5	*Chatbot* tidak memberikan informasi yang menyesatkan atau tidak sesuai fakta					
A6	*Chatbot* memberikan informasi harga produk yang akurat dan terkini					
A7	Ketika *chatbot* tidak tahu jawabannya, ia mengakui keterbatasannya secara jujur					

Dimensi 2 Konsistensl

NO	Pernyataan	1	2	3	4	5
C1	*Chatbot* memberikan jawaban yang SAMA ketika pertanyaan serupa diajukan lebih dari sekali					
C2	*Chatbot* tidak memberikan jawaban yang saling bertentangan dalam satu sesi percakapan					
C3	*Chatbot* memberikan informasi yang konsisten meskipun pertanyaan diajukan dengan kata-kata berbeda					
C4	*Chatbot* tidak berubah-ubah memberikan informasi tentang produk yang sama					
C5	Jawaban *chatbot* pada sesi yang berbeda tetap konsisten untuk topik yang sama					

Dimensi 3 Robustness

NO	Pernyataan	1	2	3	4	5
R1	*Chatbot* tetap memberikan jawaban yang relevan meskipun pertanyaan disampaikan dengan bahasa informal atau singkatan					
R2	*Chatbot* mampu memahami pertanyaan meskipun terdapat kesalahan ketik (typo) dalam pesan saya					
R3	*Chatbot* memberikan respon tepat terhadap pertanyaan yang disampaikan dari berbagai sudut pandang					
R4	*Chatbot* mampu memahami pertanyaan dalam bahasa campuran (Indonesia + Inggris)					

R5	*Chatbot* memberikan jawaban yang bermakna bahkan saat pertanyaan tidak lengkap atau ambigu					
R6	*Chatbot* tidak mudah terkecoh oleh pertanyaan yang tidak relevan atau di luar topik					

Pertanyaan Terbuka

NO	Pernyataan	Jawaban
R1	Tuliskan pertanyaan yang tidak bisa dijawab *chatbot* dengan benar	
R2	Apakah ada topik penting yang belum bisa dijawab *chatbot* sama sekali?	
R3	Saran perbaikan untuk *chatbot*	
R4	Menurut Anda, apakah *chatbot* ini sudah layak digunakan oleh customer? (Ya / Belum / Perlu perbaikan dulu)	
R5	Jika belum layak, apa alasan utamanya? Jika Iya cukup (-)	

FORM TANGGAPAN REVISI TUGAS AKHIR

Nama : Evan Austin

NPM : 221712040

Judul TA : Analisis Perbandingan Metode Zero-shot dan Few-shot Prompting dalam Pembuatan Chatbot Customer Service Berbasis LLM untuk Website Profil Perusahaan PT. Samudra Teknik Sejahtera

Penguji

1. Elisabeth Marsella, S.S., M.Li.

2. Hendro Gunawan, S.Si., M.T.

No	Komentar Penguji	Tanggapan
1	Knowledge Base (KB) antara metode Zero-shot dan Few-shot tidak berimbang, sehingga hasil perbandingan berpotensi bias.	KB Zero-shot dan Few-shot telah disusun ulang agar identik pada Bagian 1–6 dan 8. Seluruh 5 contoh percakapan dipindahkan dari knowledge base ke Response Prompt Few-shot, sehingga perbedaan antar metode sepenuhnya berada pada komponen prompt, bukan pada akses informasi BAB IV subbab 4.4 dan 4.5
2	Proses pengujian sebelumnya bersifat mandiri (self-assessment), berpotensi bias penilaian manual, keterbatasan skoring biner, dan tidak ada validasi antar-penilai (inter-rater reliability).	Pengujian direvisi menjadi evaluasi terstruktur bersama expert Customer Service PT. Samudra Teknik Sejahtera melalui sesi Microsoft Teams, menggunakan skala Ya/Tidak pada 5 metrik evaluasi (lihat BAB IV subbab 4.6).
3	Arah penelitian (pengembangan atau analisis perbandingan) perlu ditentukan, dengan konsekuensi terhadap judul, rumusan masalah, tujuan, dan kesimpulan.	Penelitian dipertahankan pada arah analisis perbandingan, sesuai dengan judul, rumusan masalah, dan tujuan yang telah ditulis pada BAB I. Konsekuensi dari pilihan arah ini KB berimbang sejak awal dan pengujian terstruktur dengan expert telah dipenuhi
4	Jika studi literatur telah menunjukkan Few-shot lebih optimal, mengapa perlu dilakukan pengujian kembali?	Ditambahkan argumentasi pada BAB I bahwa keunggulan Few-shot pada literatur belum diuji secara empiris pada domain distribusi alat pertukangan, perkakas, dan sanitary di Indonesia, sehingga pengujian pada domain spesifik ini tetap memberikan kontribusi ilmiah.
5	Perlu analisis lebih mendalam mengenai bagian mana yang gagal dijawab oleh chatbot Zero-shot dan Few-shot.	BAB IV subbab 4.6 memuat tabel detail kegagalan per metrik (Akurasi, Konsistensi, Robustness, Pertanyaan Repetitif, Penalaran) yang menampilkan case spesifik, jawaban

No	Komentar Penguji	Tanggapan
		chatbot, dan penyebab kegagalan masing-masing metode.
6	Perlu dijelaskan dan diteliti trade-off antara performa dan biaya token; kapan Zero-shot tetap lebih sesuai; tantangan skalabilitas Few-shot; serta perspektif berbasis kasus penggunaan (biaya, lead time, use case).	Ditambahkan subbab 4.10.2 Trade-off Biaya Token dan Lead Time Implementasi, yang membandingkan estimasi jumlah token pada Response Prompt kedua metode (Zero-shot ≈696 token, Few-shot ≈1.483 token), membahas perbedaan lead time kurasi contoh, serta memberikan rekomendasi kondisional mengenai kapan masing-masing metode lebih sesuai digunakan.
7	Belum ada bagian yang secara eksplisit mengakui keterbatasan penelitian: ukuran dataset kecil, jumlah responden terbatas, cakupan domain tunggal, tidak adanya alat evaluasi otomatis, dan keterbatasan chatbot terhadap data real-time.	Penambahan pada BAB IV subbab 4.10.4 Keterbatasan dan Rekomendasi yang mengakui secara eksplisit seluruh keterbatasan tersebut.
8	Perlu justifikasi jumlah data/kasus pengujian, jumlah responden, dan parameter teknis (Top-K) yang digunakan.	Justifikasi jumlah 5 responden pengujian usability mengacu pada Nielsen dan Faulkner telah tercantum di BAB III. Justifikasi nilai Top-K = 2 telah dijelaskan di BAB III dengan rujukan studi terkait retrieval. Justifikasi jumlah 5 case pengujian expert sedang disusun.
9	Perbaikan penulisan (penggunaan kata baku) dan gambar harus disebut/dijelaskan pada badan paragraf.	Seluruh gambar pada bagian yang telah direvisi (BAB IV subbab 4.4–4.6) telah dirujuk secara eksplisit pada badan paragraf. Pemeriksaan bahasa baku menyeluruh menggunakan fitur editor Microsoft Word sedang dilakukan pada seluruh dokumen.