JURNAL ILMIAH SAINS TEKNOLOGI DAN INFORMASI
Volume 4, Nomor 3, Juli 2026
e-ISSN: 2964-3104; p-ISSN: 2964-3090, Hal. 399-408
DOI: https://doi.org/10.59024/jiti.v4i3.2353
Tersedia: https://jurnal.alimspublishing.co.id/index.php/JITI

Klasifikasi Keluhan Pengguna Berbasis *Large Language Model* untuk Layanan Digital

Syarifah Akmal[1*], Salahuddin[2], Hartanto Satyo Nugraha[3], Sumarsid[4]
[1] Program Studi Teknik Industri, Fakultas Teknik, Universitas Malikussaleh, Indonesia
[2] Program Studi Teknik Multimedia, Jurusan Manajemen Informatika, Politeknik Negeri Sambas, Indonesia
[3] Program Studi Teknik Listrik, Akademi Teknologi Bogor, Indonesia
[4] Program Studi Manajemen, Fakultas Manajemen, Sekolah Tinggi Manajemen LABORA, Indonesia
Email: syarifah.akmal@unimal.ac.id [1*], chees095@gmail.com [2], tantoakatektel21@gmail.com [3], marsiddpk05@gmail.com [4]

**Penulis Korespondensi:* syarifah.akmal@unimal.ac.id

Abstract*. The rapid growth of digital services such as e-commerce applications, digital banking, and app-based public services has significantly increased the volume of user complaints, making manual complaint classification inefficient. This study aims to develop and evaluate a Large Language Model (LLM)-based approach for classifying user complaints on digital services using a zero-shot prompting scheme. A small sample of 60 user complaints was collected and grouped into five categories: customer service, application or technical issues, payment and transaction, delivery, and data security and privacy. The research method consists of data collection, text preprocessing, labelling, LLM-based classification, and model performance evaluation. Testing on the small sample shows that the LLM was able to classify complaints with an accuracy of 86.7%, a weighted precision of 87.2%, a weighted recall of 86.7%, and a weighted F1-score of 86.8%. Most misclassifications occurred between contextually overlapping categories, namely payment or transaction and application or technical issues. The findings indicate that LLMs have strong potential as an efficient and accurate solution to support automated user complaint handling in digital services.*

Keywords*: Large Language Model; Service; Text Classification; User Complaint; Zero-Shot Prompting.*

Abstrak. Pertumbuhan layanan digital seperti aplikasi *e-commerce*, perbankan digital, dan layanan publik berbasis aplikasi telah meningkatkan volume keluhan pengguna secara signifikan, sehingga proses klasifikasi keluhan secara manual menjadi tidak efisien. Penelitian ini bertujuan mengembangkan dan mengevaluasi pendekatan klasifikasi keluhan pengguna layanan digital berbasis *Large Language Model* (LLM) menggunakan skema *zero-shot prompting*. Sebanyak 60 sampel keluhan pengguna dikumpulkan dan dikelompokkan ke dalam lima kategori, yaitu layanan pelanggan, masalah aplikasi atau teknis, pembayaran dan transaksi, pengiriman, serta keamanan dan privasi data. Metode penelitian meliputi tahap pengumpulan data, pra-pemrosesan teks, pelabelan, klasifikasi menggunakan LLM, dan evaluasi kinerja model. Hasil pengujian pada sampel kecil menunjukkan bahwa model LLM mampu mengklasifikasikan keluhan dengan akurasi sebesar 86,7%, presisi tertimbang 87,2%, *recall* tertimbang 86,7%, dan F1-*score* tertimbang 86,8%. Kesalahan klasifikasi paling banyak terjadi pada kategori yang memiliki kemiripan konteks, yaitu antara kategori pembayaran atau transaksi dan masalah aplikasi atau teknis. Temuan ini menunjukkan bahwa LLM berpotensi menjadi solusi yang efisien dan akurat untuk mendukung otomatisasi penanganan keluhan pengguna pada layanan digital.

Kata kunci: Klasifikasi Teks; Keluhan Pengguna; *Large Language Model*; Layanan Digital; *Zero-Shot Prompting*.

1. LATAR BELAKANG

Perkembangan layanan digital seperti aplikasi *e-commerce*, perbankan digital, dan layanan publik berbasis aplikasi telah mengubah pola interaksi antara penyedia layanan dan pengguna. Seiring dengan meningkatnya jumlah pengguna, jumlah keluhan yang disampaikan melalui berbagai kanal digital seperti aplikasi, media sosial, dan pusat layanan pelanggan turut meningkat secara signifikan. Pengelolaan keluhan yang efektif berperan penting dalam

Naskah Masuk: 21 Mei 2026; Revisi: 12 Juni 2026; Diterima: 27 Juli 2026; Terbit: 29 Juli 2026

meningkatkan kepuasan pelanggan, mengidentifikasi masalah sistemik pada layanan, serta mendukung pengambilan keputusan bisnis (Vairetti, Aránguiz, Maldonado, Karmy, & Leal, 2024). Analitik ulasan pelanggan juga membantu organisasi memantau pola ketidakpuasan secara sistematis dan berkelanjutan (Kim & Lim, 2021).

Penanganan keluhan secara manual memiliki keterbatasan dari sisi waktu, konsistensi, dan skalabilitas ketika volume keluhan sangat besar. Oleh karena itu, otomatisasi klasifikasi keluhan menjadi kebutuhan penting bagi penyedia layanan digital agar keluhan dapat diarahkan ke unit penanganan yang tepat secara cepat dan konsisten (Khadija & Nurharjadmo, 2024). Pendekatan berbasis kecerdasan buatan, khususnya model bahasa berbasis arsitektur *transformer* seperti BERT (Devlin, Chang, Lee, & Toutanova, 2019) yang dibangun di atas mekanisme *attention* (Vaswani et al., 2017), telah banyak digunakan untuk klasifikasi teks keluhan pelanggan pada berbagai domain, mulai dari layanan keuangan dan transportasi hingga isu lingkungan (Zhou et al., 2024).

Dalam beberapa tahun terakhir, *Large Language Model* (LLM) generasi baru menunjukkan kemampuan pemahaman bahasa alami yang lebih luas dibandingkan model klasifikasi konvensional, termasuk dalam skema *zero-shot* tanpa proses pelatihan ulang khusus pada domain tertentu (Zhao et al., 2026). Studi terbaru menunjukkan bahwa LLM mampu mengklasifikasikan keluhan konsumen di sektor keuangan dengan akurasi yang kompetitif dibandingkan model *deep learning* konvensional (Roumeliotis, Tselikas, & Nasiopoulos, 2025). Pendekatan serupa juga memberikan hasil yang menjanjikan pada klasifikasi ulasan pelanggan di sektor pariwisata (Roumeliotis, Tselikas, & Nasiopoulos, 2024). Pada domain perbankan digital, kombinasi LLM dan *fine-tuning* mampu meningkatkan akurasi klasifikasi keluhan pelanggan secara signifikan (Güllü, Atagün, Biroğul, & Barışçı, 2025).

Meskipun demikian, studi yang dirujuk lebih banyak berfokus pada domain keuangan, pariwisata, atau penggunaan model yang telah di-*fine-tuning*. Bukti awal mengenai penggunaan *zero-shot prompting* untuk mengelompokkan keluhan layanan digital berbahasa Indonesia ke dalam kategori operasional yang beragam masih terbatas. Berdasarkan kesenjangan tersebut, penelitian ini bertujuan mengembangkan dan menguji pendekatan klasifikasi keluhan pengguna layanan digital berbasis LLM menggunakan skema *zero-shot prompting* pada sampel data berskala kecil. Kontribusi penelitian ini adalah merancang tahapan penelitian yang sistematis mulai dari pengumpulan data hingga evaluasi model, serta memberikan gambaran awal mengenai performa LLM dalam mengklasifikasikan keluhan pengguna layanan digital ke dalam kategori yang relevan bagi penyedia layanan.

2. KAJIAN TEORITIS

Keluhan Pengguna pada Layanan Digital

Keluhan pengguna merupakan bentuk umpan balik yang berisi ketidakpuasan terhadap suatu produk atau layanan. Pada layanan digital, keluhan umumnya disampaikan melalui ulasan aplikasi, formulir pengaduan, media sosial, maupun layanan pelanggan berbasis *chat*. Pengelolaan keluhan yang tidak optimal dapat menurunkan tingkat kepuasan pengguna dan berdampak pada reputasi penyedia layanan (Song, Rong, & Tang, 2024) Analisis berbasis *deep learning* juga dapat menghubungkan isi umpan balik dengan tingkat kepuasan pelanggan secara lebih terstruktur (Aldunate, Maldonado, Vairetti, & Armelini, 2022).

Keluhan pengguna dapat menjadi indikator penting bagi perbaikan layanan dan kebijakan. Pada layanan publik, laporan warga dapat digunakan untuk memantau kualitas lingkungan dan mendorong perbaikan kualitas udara (Zhou et al., 2024). Pada konteks layanan digital, pengolahan keluhan secara otomatis membantu penyedia layanan mengenali jenis masalah, tingkat urgensi, dan unit penanganan yang sesuai.

Large Language Model (LLM) untuk Klasifikasi Teks

Large Language Model (LLM) merupakan model bahasa berskala besar yang dilatih pada korpus teks dalam jumlah masif menggunakan arsitektur *transformer* (Vaswani et al., 2017). LLM memiliki kemampuan memahami konteks bahasa sehingga dapat digunakan untuk berbagai tugas pemrosesan bahasa alami, termasuk klasifikasi teks, tanpa memerlukan pelatihan ulang khusus domain (Zhao et al., 2026). Perkembangan *generative artificial intelligence* memperluas penggunaan model bahasa dari pembuatan teks menuju analisis, penalaran, dan dukungan keputusan berbasis bahasa (Feuerriegel, Hartmann, Janiesch, & Zschech, 2024).

Pendekatan *zero-shot prompting* memungkinkan LLM mengklasifikasikan teks ke dalam kategori yang telah ditentukan hanya berdasarkan instruksi yang diberikan, tanpa data latih berlabel (Roumeliotis et al., 2025). Liu et al. (2023) menunjukkan bahwa rancangan *prompt* menjadi komponen utama dalam mengarahkan model pralatih untuk menyelesaikan tugas bahasa. Gao, Fisch, dan Chen (2021) juga menunjukkan bahwa formulasi *prompt* dapat meningkatkan kemampuan model pada kondisi data berlabel yang terbatas. Namun, performa LLM dapat berbeda antar tugas dan domain, sehingga evaluasi empiris tetap diperlukan (Kocoń et al., 2023).

Penelitian Terkait

Beberapa penelitian terdahulu telah mengevaluasi penggunaan model bahasa untuk klasifikasi keluhan pelanggan pada berbagai domain, sebagaimana dirangkum pada Tabel 1.

Tabel 1. Ringkasan Penelitian Terkait Klasifikasi Keluhan Pelanggan Berbasis LLM/NLP.

No.	Peneliti (Tahun)	Metode	Domain	Hasil Utama
1	Vairetti et al. (2024)	*Deep learning* + BERT (BETO)	Prioritas keluhan pelanggan	Akurasi 92,1%
2	Roumeliotis et al. (2025)	LLM *zero-shot* dan *reasoning model*	Keluhan konsumen keuangan (CFPB)	Akurasi 70%-78%
3	Roumeliotis et al. (2024)	GPT Omni dan BERT NLP	Ulasan pelanggan pariwisata	Klasifikasi sentimen kompetitif
4	Güllü et al. (2025)	LSTM, CNN, BERT, RoBERTa, LLM *fine-tuned*	Keluhan pelanggan perbankan	LLM *fine-tuned* unggul dibanding model lain
5	Khadija dan Nurharjadmo (2024)	LDA + BERT *embedding*	Keluhan pelanggan berbahasa Indonesia	Peningkatan koherensi topik

Sumber: Disusun dari literatur yang dirujuk.

Berdasarkan Tabel 1, pendekatan berbasis LLM secara konsisten menunjukkan hasil yang kompetitif pada berbagai domain keluhan, sehingga relevan untuk diadaptasi pada konteks layanan digital secara lebih luas. Penelitian lain memperluas analisis keluhan ke aspek keparahan, interpretabilitas, dan transfer pembelajaran. Singh, Bhatia, dan Saha (2024) mengembangkan identifikasi keluhan dan tingkat keparahan dari konten keuangan daring, sedangkan Das, Singh, Saha, dan Maurya (2024) menekankan pentingnya interpretabilitas dalam membedakan ulasan negatif dan keluhan keuangan.

Lee dan Zhao (2024) menunjukkan bahwa pengumpulan data, *data mining*, dan *transfer learning* dapat mendukung sistem penanganan keluhan yang berpusat pada karakter pelanggan. Temuan-temuan tersebut memperkuat kebutuhan akan pendekatan klasifikasi yang efisien, tetapi belum secara langsung menguji skema *zero-shot* pada sampel kecil yang mencakup beberapa jenis layanan digital sebagaimana dilakukan dalam penelitian ini.

3. METODE PENELITIAN

Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif deskriptif dengan metode eksperimen untuk menguji kemampuan LLM dalam mengklasifikasikan keluhan pengguna layanan digital. Klasifikasi dilakukan menggunakan skema *zero-shot prompting*, yaitu model diminta mengklasifikasikan teks keluhan ke dalam kategori yang telah ditentukan

tanpa proses *fine-tuning* tambahan, mengikuti pendekatan penelitian sejenis (Roumeliotis et al., 2025).

Data Penelitian

Data penelitian berupa 60 sampel keluhan pengguna layanan digital, khususnya aplikasi *e-commerce* dan perbankan digital, yang dikumpulkan sebagai studi awal berskala kecil. Setiap keluhan dikelompokkan secara manual oleh peneliti ke dalam lima kategori operasional, yaitu layanan pelanggan, masalah aplikasi atau teknis, pembayaran dan transaksi, pengiriman, serta keamanan dan privasi data. Distribusi data disajikan pada Tabel 2.

Tabel 2. Distribusi Sampel Data Keluhan Pengguna Berdasarkan Kategori.

No.	Kategori Keluhan	Jumlah Sampel	Persentase
1	Layanan Pelanggan	12	20,0%
2	Masalah Aplikasi/Teknis	15	25,0%
3	Pembayaran dan Transaksi	14	23,3%
4	Pengiriman	10	16,7%
5	Keamanan dan Privasi Data	9	15,0%
	Total	60	100%

Sumber: Hasil pengolahan data penelitian.

Tahapan Penelitian

Penelitian ini dilaksanakan melalui delapan tahapan utama, mulai dari identifikasi masalah hingga penarikan kesimpulan, sebagaimana ditunjukkan pada Gambar 1.

Gambar 1. Diagram alur penelitian klasifikasi keluhan pengguna berbasis LLM

Sumber: Hasil pengolahan data penelitian.

Tahap pra-pemrosesan teks meliputi pembersihan teks (*cleaning*), *case folding*, tokenisasi, dan penghapusan *stopword* untuk menstandarkan format keluhan sebelum diklasifikasikan. Tahap pelabelan dilakukan secara manual oleh peneliti sebagai label acuan (*ground truth*) untuk keperluan evaluasi.

Skenario Klasifikasi

Klasifikasi dilakukan dengan menyusun *prompt* yang berisi instruksi tugas, daftar kategori target, dan teks keluhan, mengikuti prinsip rekayasa *prompt* sebagaimana diterapkan pada penelitian klasifikasi keluhan konsumen sebelumnya (Güllü et al., 2025). Model diminta mengembalikan hasil klasifikasi dalam format terstruktur berupa satu label kategori untuk setiap keluhan.

Evaluasi Kinerja Model

Kinerja model dievaluasi menggunakan metrik akurasi, presisi, *recall*, dan F1-*score* dengan pendekatan *weighted average* untuk mengakomodasi perbedaan jumlah data antar kategori. Evaluasi juga didukung dengan analisis matriks konfusi untuk mengidentifikasi pola kesalahan klasifikasi antar kategori (Roumeliotis et al., 2025).

4. HASIL DAN PEMBAHASAN

Contoh Hasil Klasifikasi

Tabel 3 menyajikan contoh hasil klasifikasi pada sebagian sampel keluhan, dibandingkan antara label aktual hasil pelabelan manual dan label prediksi dari LLM.

Tabel 3. Contoh Hasil Klasifikasi Keluhan (Label Aktual dan Label Prediksi).

No.	Cuplikan Keluhan	Label Aktual	Label Prediksi LLM	Ket.
1	"Aplikasi selalu *force close* saat saya coba membuka menu riwayat transaksi."	Masalah Aplikasi/Teknis	Masalah Aplikasi/Teknis	Benar
2	"Saya sudah *transfer* tapi saldo tidak masuk-masuk ke rekening tujuan."	Pembayaran dan Transaksi	Pembayaran dan Transaksi	Benar
3	"Paket saya sudah tiga hari tidak ada *update* pengiriman, kurir tidak merespons."	Pengiriman	Pengiriman	Benar
4	"Kenapa data pribadi saya bisa muncul di aplikasi lain? Ini sangat mengkhawatirkan."	Keamanan dan Privasi Data	Keamanan dan Privasi Data	Benar
5	"CS tidak membalas *chat* saya selama dua hari, padahal masalah saya *urgent*."	Layanan Pelanggan	Layanan Pelanggan	Benar
6	"Fitur pembayaran otomatis gagal terus setiap bulan, harus bayar manual."	Pembayaran dan Transaksi	Masalah Aplikasi/Teknis	Salah
7	"Aplikasi lemot dan sering *error* saat proses *checkout*."	Masalah Aplikasi/Teknis	Pembayaran dan Transaksi	Salah
8	"Akun saya tiba-tiba *logout* sendiri dan minta verifikasi ulang terus-menerus."	Keamanan dan Privasi Data	Masalah Aplikasi/Teknis	Salah

Sumber: Hasil pengolahan data penelitian.

Matriks Konfusi

Hasil klasifikasi terhadap keseluruhan 60 sampel keluhan disajikan dalam bentuk matriks konfusi pada Tabel 4. Baris menunjukkan kategori aktual, sedangkan kolom menunjukkan kategori hasil prediksi model.

Tabel 4. Matriks Konfusi Hasil Klasifikasi Keluhan (n = 60).

Aktual-Prediksi	Layanan Pelanggan	Aplikasi/Teknis	Pembayaran/Transaksi	Pengiriman	Keamanan/Privasi
Layanan Pelanggan	10	1	1	0	0
Aplikasi/Teknis	1	13	1	0	0
Pembayaran/Transaksi	0	1	12	1	0
Pengiriman	0	0	1	9	0
Keamanan/Privasi	0	1	0	0	8

Sumber: Hasil pengolahan data penelitian.

Kinerja Model per Kategori

Berdasarkan matriks konfusi pada Tabel 4, nilai presisi, *recall*, dan F1-*score* dihitung untuk masing-masing kategori sebagaimana disajikan pada Tabel 5.

Tabel 5. Kinerja Klasifikasi Per Kategori Keluhan.

Kategori	Presisi	*Recall*	F1-*score*	Jumlah Sampel
Layanan Pelanggan	0,909	0,833	0,869	12
Aplikasi/Teknis	0,813	0,867	0,839	15
Pembayaran/Transaksi	0,800	0,857	0,828	14
Pengiriman	0,900	0,900	0,900	10
Keamanan/Privasi	1,000	0,889	0,941	9

Sumber: Hasil pengolahan data penelitian.

Kinerja Model Secara Keseluruhan

Evaluasi kinerja model secara keseluruhan dilakukan untuk mengetahui tingkat ketepatan klasifikasi LLM terhadap seluruh sampel keluhan pengguna. Pengukuran dilakukan menggunakan metrik akurasi, presisi tertimbang, *recall* tertimbang, dan F1-*score* tertimbang. Ringkasan hasil evaluasi tersebut disajikan pada Tabel 6.

Tabel 6. Kinerja Klasifikasi Keseluruhan model LLM pada Sampel Uji.

Metrik Evaluasi	Nilai
Akurasi	86,7%
Presisi (*weighted average*)	87,2%
Recall (*weighted average*)	86,7%
F1-*score* (*weighted average*)	86,8%

Sumber: Hasil pengolahan data penelitian.

Hasil pengujian pada sampel kecil menunjukkan bahwa LLM mampu mengklasifikasikan keluhan pengguna layanan digital dengan akurasi 86,7%, presisi tertimbang 87,2%, *recall* tertimbang 86,7%, dan F1-*score* tertimbang 86,8%. Nilai ini sejalan dengan penelitian sebelumnya yang melaporkan bahwa LLM mampu mencapai akurasi klasifikasi keluhan konsumen pada rentang 70%-80% dalam skema *zero-shot* tanpa *fine-tuning* (Roumeliotis et al., 2025). Performa dapat melebihi 90% ketika model dikombinasikan dengan *fine-tuning* pada domain spesifik (Vairetti et al., 2024). Hasil Güllü et al. (2025) juga menunjukkan keunggulan model yang telah disesuaikan pada data keluhan perbankan.

Berdasarkan Tabel 4, kesalahan klasifikasi paling banyak terjadi antara kategori pembayaran atau transaksi dan masalah aplikasi atau teknis. Hal ini konsisten dengan temuan bahwa kategori keluhan yang memiliki kemiripan konteks bahasa, seperti keluhan pembayaran otomatis yang gagal akibat error pada aplikasi, cenderung sulit dibedakan oleh model klasifikasi berbasis teks (Khadija & Nurharjadmo, 2024). Kategori keamanan dan privasi data menunjukkan performa klasifikasi terbaik dengan F1-*score* 0,941, kemungkinan karena kosakata yang digunakan pada kategori ini relatif khas dan tidak banyak tumpang tindih dengan kategori lain.

Secara umum, hasil ini menunjukkan bahwa pendekatan LLM berbasis *zero-shot prompting* berpotensi menjadi solusi awal yang efisien untuk mendukung otomatisasi klasifikasi keluhan pada layanan digital, khususnya bagi penyedia layanan berskala kecil hingga menengah yang belum memiliki data latih berlabel dalam jumlah besar. Dalam penerapannya, organisasi tetap perlu mengelola keseimbangan antara skalabilitas, akurasi, dan pengawasan manusia sebagaimana dibahas oleh Ferraro, Demsar, Sands, Restrepo, dan Campbell (2024). Mengingat ukuran sampel masih terbatas sebanyak 60 keluhan, pengujian lanjutan pada *dataset* yang lebih besar dan representatif diperlukan untuk memastikan generalisasi hasil.

5. KESIMPULAN DAN SARAN

Penelitian ini mengembangkan dan menguji pendekatan klasifikasi keluhan pengguna layanan digital berbasis LLM menggunakan skema *zero-shot prompting* pada sampel kecil berjumlah 60 keluhan yang terbagi ke dalam lima kategori. Hasil pengujian menunjukkan bahwa model mencapai akurasi 86,7%, presisi tertimbang 87,2%, *recall* tertimbang 86,7%, dan F1-*score* tertimbang 86,8%. Kesalahan klasifikasi didominasi oleh tumpang tindih konteks antara kategori pembayaran atau transaksi dan masalah aplikasi atau teknis. Hasil tersebut

menunjukkan bahwa LLM berpotensi mendukung otomatisasi klasifikasi keluhan pengguna secara cepat dan konsisten tanpa memerlukan proses pelatihan ulang model.

Penelitian ini terbatas pada ukuran sampel yang kecil dan belum membandingkan beberapa jenis LLM. Penelitian selanjutnya disarankan menggunakan *dataset* yang lebih besar dan representatif, membandingkan performa beberapa LLM secara langsung, serta mengeksplorasi kombinasi *prompt* engineering dan *fine-tuning* untuk meningkatkan akurasi pada kategori yang tumpang tindih.

DAFTAR REFERENSI

Aldunate, Á., Maldonado, S., Vairetti, C., & Armelini, G. (2022). Understanding customer satisfaction via deep learning and natural language processing. *Expert Systems with Applications*, *209*, 118309. https://doi.org/10.1016/j.eswa.2022.118309

Das, S., Singh, A., Saha, S., & Maurya, A. (2024). Negative review or complaint? Exploring interpretability in financial complaints. *IEEE Transactions on Computational Social Systems*, *11*(3), 3606–3615. https://doi.org/10.1109/TCSS.2023.3338357

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of deep bidirectional transformers for language understanding*. *1*, 4171–4186. Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423

Ferraro, C., Demsar, V., Sands, S., Restrepo, M., & Campbell, C. L. (2024). The paradoxes of generative AI-enabled customer service: A guide for managers. *Business Horizons*, *67*(5), 549–559. https://doi.org/10.1016/j.bushor.2024.04.013

Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2024). Generative AI. *Business & Information Systems Engineering*, *66*(1), 111–126. https://doi.org/10.1007/s12599-023-00834-7

Gao, T., Fisch, A., & Chen, D. (2021). *Making pre-trained language models better few-shot learners*. *1*, 3816–3830. Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.acl-long.295

Güllü, M., Atagün, E., Biroğul, S., & Barışçı, N. (2025). *Customer complaint classification with large language models*. 1–6. IEEE. https://doi.org/10.1109/ACDSA65407.2025.11166338

Khadija, M. A., & Nurharjadmo, W. (2024). Enhancing Indonesian customer complaint analysis: LDA topic modelling with BERT embeddings. *SINERGI*, *28*(1), 153–162. https://doi.org/10.22441/sinergi.2024.1.015

Kim, J., & Lim, C. (2021). Customer complaints monitoring with customer review data analytics: An integrated method of sentiment and statistical process control analyses. *Advanced Engineering Informatics*, *49*, 101304. https://doi.org/10.1016/j.aei.2021.101304

Kocoń, J., Cichecki, I., Kaszyca, O., Kochanek, M., Szydło, D., Baran, J., … Kazienko, P. (2023). ChatGPT: Jack of all trades, master of none. *Information Fusion*, *99*, 101861. https://doi.org/10.1016/j.inffus.2023.101861

Lee, C.-H., & Zhao, X. (2024). Data collection, data mining and transfer of learning based on customer temperament-centered complaint handling system and one-of-a-kind complaint handling dataset. *Advanced Engineering Informatics*, *60*, 102520. https://doi.org/10.1016/j.aei.2024.102520

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, *55*(9), 1–35. https://doi.org/10.1145/3560815

Roumeliotis, K. I., Tselikas, N. D., & Nasiopoulos, D. K. (2024). Leveraging large language models in tourism: A comparative study of the latest GPT Omni models and BERT NLP for customer review classification and sentiment analysis. *Information*, *15*(12), 792. https://doi.org/10.3390/info15120792

Roumeliotis, K. I., Tselikas, N. D., & Nasiopoulos, D. K. (2025). Think before you classify: The rise of reasoning large language models for consumer complaint detection and classification. *Electronics*, *14*(6), 1070. https://doi.org/10.3390/electronics14061070

Singh, A., Bhatia, R., & Saha, S. (2024). Complaint and severity identification from online financial content. *IEEE Transactions on Computational Social Systems*, *11*(1), 660–670. https://doi.org/10.1109/TCSS.2022.3215528

Song, W., Rong, W., & Tang, Y. (2024). Quantifying risk of service failure in customer complaints: A textual analysis-based approach. *Advanced Engineering Informatics*, *60*, 102377. https://doi.org/10.1016/j.aei.2024.102377

Vairetti, C., Aránguiz, I., Maldonado, S., Karmy, J. P., & Leal, A. (2024). Analytics-driven complaint prioritisation via deep learning and multicriteria decision-making. *European Journal of Operational Research*, *312*(3), 1108–1118. https://doi.org/10.1016/j.ejor.2023.08.027

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., … Polosukhin, I. (2017). *Attention is all you need*. *30*. Retrieved from https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

Zhao, W. X., Zhou, K., Li, J., Tang, T., Dong, Z., Hou, Y., … Wen, J.-R. (2026). A survey of large language models. *Frontiers of Computer Science*, *20*(12), 2012627. https://doi.org/10.1007/s11704-026-60308-3

Zhou, X., Cao, G., Peng, B., Xu, X., Yu, F., Xu, Z., … Du, H. (2024). Citizen environmental complaint reporting and air quality improvement: A panel regression analysis in China. *Journal of Cleaner Production*, *434*, 140319. https://doi.org/10.1016/j.jclepro.2023.140319