

Analisis Sentimen Publik Terhadap ChatGPT Menggunakan Arsitektur Transformer: A Systematic Literature Review

Amalia Nur Afifah^{a*}, Endah Sudarmilah^b, Yusuf Sulistyo Nugroho^c
1208250005@student.ums.ac.id

Universitas Muhammadiyah Surakarta^{abc}

Intisari

Kata kunci:
Analisis sentimen,
ChatGPT, BERT,
transformer

Analisis sentimen terhadap opini publik mengenai ChatGPT telah menjadi topik yang semakin banyak diteliti dalam bidang *Natural Language Processing* (NLP) dan deep learning sejak peluncurannya pada November 2022. Tinjauan literatur sistematis ini bertujuan menyajikan referensi menyeluruh mengenai pendekatan, metode, dan teknik yang digunakan dalam analisis sentimen publik terhadap ChatGPT dengan model berbasis transformer. Kajian dilakukan terhadap artikel yang diterbitkan pada periode 2022–2026 melalui basis data Scopus. Dari proses seleksi yang dilakukan, diperoleh 30 artikel yang memenuhi kriteria inklusi untuk dianalisis. Hasil penelitian menunjukkan bahwa tidak terdapat publikasi yang memenuhi kriteria pada tahun 2022, sementara terjadi peningkatan signifikan pada tahun 2024 dengan kontribusi sebesar 53,3% dari total artikel. Model BERT dan variannya, termasuk ALBERT serta kombinasi BERT+RoBERTa, menjadi pendekatan yang paling banyak digunakan (40,0%), sedangkan BERT murni dan fine-tuned BERT mendominasi sebesar 30,0%. Rata-rata nilai F1-score yang dilaporkan mencapai 0,904 dengan rentang 0,87–0,93. Twitter/X merupakan sumber data yang paling sering digunakan (53,3%), sedangkan bahasa Inggris mendominasi 86,7% penelitian. Meskipun performa model tergolong tinggi, tantangan seperti sarkasme, bahasa informal, dan konteks multibahasa masih menjadi perhatian. Tinjauan ini memberikan gambaran terkini sekaligus peluang penelitian untuk meningkatkan kualitas sistem analisis sentimen terhadap teknologi kecerdasan buatan generatif.

Tentang naskah:
-diterima, 17 Juni 2026
-diterbitkan, 1 Juli 2026

Abstract

Sentiment analysis of public opinion regarding ChatGPT has become an increasingly popular research topic in the fields of Natural Language Processing (NLP) and deep learning since its public release in November 2022. This systematic literature review aims to provide a comprehensive reference on the approaches, methods, and techniques employed in analyzing public sentiment toward ChatGPT using transformer-based models. The review examined studies published between 2022 and 2026 and retrieved from the Scopus database. After a rigorous selection process, 30 articles met the inclusion criteria and were analyzed. The findings reveal that no eligible publications were identified in 2022, while a significant increase occurred in 2024, accounting for 53.3% of the total studies. BERT and its variants, including ALBERT and the BERT+RoBERTa ensemble, were the most widely adopted approaches, accounting for 40.0% of the reviewed studies, while standard and fine-tuned BERT models alone represented 30.0% of the total articles. The reported average F1-score was 0.904, ranging from 0.87 to 0.93. Twitter/X was the most frequently used data source (53.3%), and English was the dominant language in 86.7% of the studies. Despite the strong performance achieved by transformer-based models, challenges remain in handling sarcasm, informal language, and multilingual contexts. This review provides an up-to-date overview of current research trends and highlights future opportunities to enhance sentiment analysis systems for generative artificial intelligence technologies.

Keywords:
Sentiment analysis,
ChatGPT, BERT,
transformer

1. Pendahuluan

Perkembangan teknologi kecerdasan buatan (*Artificial Intelligence*), khususnya *Large Language Models* (LLMs), mengalami kemajuan yang sangat pesat dalam beberapa tahun terakhir. Model bahasa

berskala besar mampu memahami serta menghasilkan teks yang menyerupai bahasa manusia dengan tingkat akurasi yang semakin tinggi. Salah satu implementasi LLM yang paling populer adalah ChatGPT yang dikembangkan oleh OpenAI. Model ini dirancang untuk

menghasilkan respons berbasis teks secara kontekstual dan telah dimanfaatkan dalam berbagai bidang, seperti pendidikan, pengembangan perangkat lunak, bisnis, serta layanan informasi digital (Bin-Nashwan, Sadallah, & Bouteraa, 2023; Su & Kabala, 2023). Sejak diperkenalkan kepada publik, ChatGPT dengan cepat memperoleh perhatian luas dari masyarakat global dan menjadi salah satu aplikasi kecerdasan buatan yang paling banyak digunakan.

Popularitas ChatGPT mendorong munculnya berbagai diskusi di platform media sosial seperti Twitter, Reddit, serta berbagai platform ulasan aplikasi. Platform-platform tersebut menjadi ruang bagi pengguna untuk menyampaikan pengalaman, opini, serta persepsi mereka terhadap teknologi ini (Naing & Udomwong, 2024a; Xu, Fang, Huang, & Xie, 2024a). Interaksi pengguna pada media sosial menghasilkan volume data teks yang sangat besar yang mencerminkan beragam pandangan publik terhadap penggunaan ChatGPT. Informasi yang terkandung dalam data tersebut sangat berharga untuk memahami persepsi masyarakat terhadap perkembangan teknologi kecerdasan buatan. Maka dari itu, analisis sentimen menjadi salah satu pendekatan yang banyak digunakan untuk mengidentifikasi dan menganalisis opini publik yang terkandung dalam data teks tersebut (Koonchanok, Pan, & Jang, 2024). ChatGPT secara umum dipersepsikan sebagai teknologi dengan kualitas tinggi yang didominasi oleh sentimen positif dan emosi kegembiraan. Dalam literatur ilmiah, ChatGPT banyak dipandang sebagai peluang besar di berbagai bidang, meskipun juga menimbulkan kekhawatiran dari perspektif etika. Selain itu, persepsi di media sosial menunjukkan tren penurunan seiring waktu, ditandai dengan berkurangnya emosi kegembiraan dan meningkatnya rasa keterkejutan (Leiter et al., 2024).

Analisis sentimen merupakan salah satu cabang penting dalam *Natural Language Processing* (NLP) yang bertujuan untuk mengidentifikasi serta mengklasifikasikan opini dalam teks ke dalam kategori sentimen tertentu, seperti positif, negatif, atau netral (Issa & Abdu Ibrahim, 2024). Sebelumnya, pendekatan analisis sentimen banyak menggunakan metode berbasis leksikon maupun algoritma machine learning tradisional.

Namun, perkembangan teknologi *deep learning* telah membawa perubahan signifikan dalam metode analisis sentimen, khususnya melalui penggunaan model berbasis transformer seperti *Bidirectional Encoder Representations from Transformers* (BERT) dan berbagai variannya (Mujahid, Kanwal, Rustam, Aljedaani, & Ashraf, 2023; Sudheesh et al., 2023). Model berbasis *transformer* mampu memahami konteks linguistik secara lebih komprehensif melalui mekanisme *self-attention*, sehingga memberikan performa yang lebih baik dibandingkan pendekatan konvensional dalam berbagai tugas pemrosesan bahasa alami.

Sejumlah penelitian telah memanfaatkan model transformer seperti BERT, RoBERTa, DistilBERT untuk melakukan analisis sentimen terhadap opini publik mengenai ChatGPT pada berbagai platform media sosial. Misalnya, penelitian oleh Agus Surya Dharma et al. (2023) menunjukkan bahwa model BERT yang di-fine-tune mampu memberikan kinerja yang baik dalam menganalisis sentimen masyarakat Indonesia terhadap ChatGPT pada platform media sosial. Penelitian lain oleh Mujahid et al. (2023) menunjukkan bahwa kombinasi model RoBERTa dan BERT dapat meningkatkan akurasi klasifikasi sentimen pada dataset tweet berbahasa Arab yang berkaitan dengan ChatGPT. Selain itu, Lyu et al. (2025) mengembangkan kerangka berbasis BERT untuk menganalisis kepercayaan publik terhadap ChatGPT dan menemukan bahwa model berbasis *transformer* mampu menangkap konteks semantik secara lebih efektif dibandingkan metode konvensional.

Selain penggunaan model transformer, beberapa penelitian juga mengkaji persepsi publik terhadap ChatGPT dari berbagai perspektif dan domain aplikasi. Penelitian oleh Danish et al. (2025) melakukan analisis sentimen dan analisis tematik terhadap diskusi mengenai penggunaan ChatGPT dalam pendidikan menggunakan data *Twitter*. Hasil penelitian tersebut menunjukkan bahwa sebagian besar opini publik terkait penggunaan ChatGPT dalam pendidikan bersifat positif, meskipun masih terdapat kekhawatiran terkait etika dan integritas akademik (Danish et al., 2025). Studi lain oleh Padarha & Vijaylakshmi (2024) mengeksplorasi opini pengguna Reddit terhadap ChatGPT dan menemukan bahwa persepsi pengguna sangat dipengaruhi oleh konteks penggunaan teknologi tersebut.

Penelitian terkait analisis sentimen terhadap ChatGPT juga telah dilakukan pada berbagai bahasa dan konteks budaya. Alobedalla et al. (2025) meneliti opini publik berbahasa Arab terhadap ChatGPT pada platform Twitter dan menemukan adanya variasi emosi publik yang dipengaruhi oleh faktor sosial dan budaya. Selain itu, Hannani et al. (2024) menunjukkan bahwa penggunaan model BERT yang di-*fine-tune* dapat meningkatkan performa analisis sentimen pada bahasa dengan sumber data terbatas. Safira et al. (2025) juga memanfaatkan model ALBERT untuk melakukan klasifikasi sentimen terhadap ulasan panjang mengenai ChatGPT, yang menunjukkan bahwa model *transformer* ringan dapat memberikan kinerja yang kompetitif.

Dalam konteks yang lebih luas, analisis sentimen terhadap teknologi kecerdasan buatan juga digunakan untuk memahami dinamika opini publik terhadap perkembangan teknologi generatif. Xing et al. (2024) mengkaji fenomena polarisasi opini publik terhadap ChatGPT pada media sosial dan menemukan adanya perbedaan pandangan yang signifikan antar kelompok pengguna. Zhang et al. (2024) menunjukkan bahwa diskusi publik mengenai kecerdasan buatan dalam pendidikan meningkat secara signifikan setelah kemunculan ChatGPT.

Beberapa penelitian juga mengeksplorasi pendekatan analisis sentimen dengan memanfaatkan kombinasi berbagai metode machine learning dan deep learning. Elmadya & Oktavia (2025) melakukan analisis menyeluruh terhadap ulasan aplikasi ChatGPT dengan membandingkan pendekatan berbasis leksikon, machine learning, serta deep learning, dan menunjukkan bahwa model berbasis pembelajaran mendalam memberikan performa yang lebih baik. Sinaga et al. (2025) juga menemukan adanya perubahan tren sentimen masyarakat terhadap *generative AI* di Indonesia seiring meningkatnya adopsi teknologi tersebut.

Berdasarkan telaah terhadap penelitian-penelitian terdahulu, dapat disimpulkan bahwa model berbasis *transformer*, khususnya BERT dan variannya, telah banyak digunakan untuk menganalisis sentimen publik terhadap ChatGPT pada berbagai platform dan bahasa. Sebagian besar penelitian masih berfokus pada eksperimen model tertentu,

dataset yang spesifik, atau konteks aplikasi yang terbatas, sehingga belum tersedia kajian yang secara sistematis merangkum perkembangan penelitian, tren penggunaan model *transformer*, sumber data yang digunakan, metrik evaluasi yang diterapkan, serta tantangan yang masih dihadapi dalam analisis sentimen terhadap ChatGPT. Keterbatasan tersebut menunjukkan adanya kesenjangan penelitian (*research gap*) berupa kurangnya sintesis menyeluruh yang mampu memberikan gambaran menyeluruh mengenai perkembangan bidang ini. Maka, penelitian ini bertujuan untuk melakukan *Systematic Literature Review* (SLR) terhadap studi-studi yang membahas analisis sentimen publik terhadap ChatGPT menggunakan arsitektur *transformer*, untuk mengidentifikasi tren penelitian, mengevaluasi pendekatan yang digunakan, menganalisis sumber data dan performa model, serta mengungkap tantangan dan peluang penelitian yang dapat menjadi dasar pengembangan penelitian selanjutnya.

Secara khusus, penelitian ini memberikan beberapa kontribusi utama sebagai berikut:

- Mengidentifikasi tren publikasi penelitian terkait analisis sentimen terhadap ChatGPT menggunakan pendekatan deep learning dan model *transformer*.
- Mengkaji berbagai varian model BERT yang digunakan dalam penelitian analisis sentimen, termasuk RoBERTa, DistilBERT, dan model *transformer* lainnya.
- Mengidentifikasi sumber data opini publik yang paling banyak digunakan, seperti *Twitter*, *Reddit*, maupun ulasan aplikasi.
- Mengidentifikasi keterbatasan serta tantangan dalam penggunaan model *transformer* untuk analisis sentimen pada teks media sosial.

Artikel ini disusun sebagai berikut. Bagian II mengulas penelitian terkait (kerangka teori) yang berkaitan dengan analisis sentimen serta penerapan model *transformer* dalam pemrosesan bahasa alami. Bagian III menjelaskan metodologi penelitian yang digunakan dalam studi literatur sistematis ini, yang meliputi strategi pencarian literatur, kriteria inklusi dan eksklusi, serta proses seleksi artikel menggunakan pendekatan PRISMA. Bagian IV menyajikan hasil analisis berdasarkan pertanyaan penelitian yang

telah dirumuskan. Selanjutnya, Bagian V membahas temuan penelitian serta implikasi dan arah penelitian di masa mendatang. Terakhir, Bagian VI menyajikan kesimpulan dari penelitian ini.

2. Kerangka Teori

2.1. Analisis Sentimen

Analisis sentimen (sentiment analysis) merupakan salah satu bidang dalam Natural Language Processing (NLP) yang bertujuan untuk mengidentifikasi, mengekstraksi, dan mengklasifikasikan opini atau sikap yang terkandung dalam suatu teks. Dalam konteks media sosial dan platform digital, analisis sentimen digunakan untuk memahami persepsi pengguna terhadap suatu produk, layanan, maupun teknologi tertentu. Seiring meningkatnya penggunaan ChatGPT dan teknologi Generative Artificial Intelligence, analisis sentimen menjadi pendekatan yang banyak digunakan untuk mengevaluasi respons publik terhadap teknologi tersebut (ElAzzouzy, Chanyour, & Andaloussi, 2025).

Secara umum, analisis sentimen mengelompokkan opini ke dalam tiga kategori utama, yaitu sentimen positif, negatif, dan netral. Klasifikasi ini memungkinkan peneliti untuk mengidentifikasi kecenderungan persepsi pengguna terhadap suatu topik tertentu. Dalam beberapa tahun terakhir, pendekatan berbasis *deep learning* menunjukkan performa yang lebih baik dibandingkan metode tradisional karena kemampuannya memahami konteks linguistik yang kompleks, termasuk hubungan antar kata dan makna kalimat secara keseluruhan (Amrullah, 2025; Triawan, Tahyudin, & Purwadi, 2025).

2.2. Transformer dan BERT

2.2.1 Transformer Architecture

Arsitektur Transformer merupakan salah satu perkembangan penting dalam bidang NLP yang menjadi fondasi bagi berbagai *Large Language Models modern*. Transformer dirancang untuk memproses urutan kata secara paralel dan mampu memahami hubungan kontekstual antar kata secara lebih efektif dibandingkan arsitektur berbasis RNN maupun LSTM. Kemampuan ini menjadikan Transformer sangat efektif untuk berbagai tugas NLP, termasuk klasifikasi teks dan analisis sentiment (Devlin, Chang, Lee, & Toutanova, 2019).

Menurut IBM, Transformer merupakan arsitektur jaringan saraf yang mampu mempelajari hubungan antar elemen dalam suatu urutan data melalui mekanisme perhatian (attention mechanism), sehingga dapat menghasilkan representasi bahasa yang lebih kontekstual dan akurat (Stryker & Bergmann, n.d.).

2.2.2 Self-Attention Mechanism

Komponen utama dalam Transformer adalah self-attention mechanism. Mekanisme ini memungkinkan model untuk menentukan tingkat kepentingan setiap kata terhadap kata lainnya dalam sebuah kalimat. Dengan demikian, model dapat memahami konteks secara lebih baik, termasuk hubungan semantik yang berada pada posisi berjauhan dalam teks. Kemampuan ini menjadi salah satu alasan utama mengapa model Transformer mampu mencapai performa tinggi pada berbagai tugas NLP, termasuk analisis sentiment (Amrullah, 2025).

2.2.3 BERT

BERT (*Bidirectional Encoder Representations from Transformers*) merupakan model bahasa berbasis Transformer yang diperkenalkan oleh Devlin et al. (2019). BERT menggunakan pendekatan *bidirectional encoding* yang memungkinkan model memahami konteks kata dari arah kiri dan kanan secara simultan. Pendekatan ini memberikan peningkatan performa yang signifikan pada berbagai tugas NLP, termasuk analisis sentiment.

Dalam penelitian terkait sentimen terhadap ChatGPT, BERT menjadi model yang paling dominan digunakan karena mampu menghasilkan performa yang konsisten dan akurat pada berbagai jenis dataset.

2.2.4 RoBERTa

RoBERTa (*Robustly Optimized BERT Pretraining Approach*) merupakan pengembangan dari BERT yang menggunakan strategi pelatihan yang lebih optimal dan dataset yang lebih besar. Berbagai penelitian menunjukkan bahwa RoBERTa mampu menghasilkan performa yang lebih baik dibandingkan BERT pada beberapa tugas klasifikasi teks dan analisis sentiment (Zhang et al., 2024). Penelitian terbaru juga menunjukkan bahwa RoBERTa memiliki keunggulan dalam memahami konteks teks yang kompleks dan menghasilkan akurasi yang lebih tinggi pada berbagai dataset.

sentimen dibandingkan model BERT standar (Fahmi & Nababan, 2025).

2.2.5 ALBERT

ALBERT (*A Lite BERT*) merupakan varian BERT yang dirancang untuk mengurangi jumlah parameter model tanpa mengurangi performa secara signifikan. Model ini menggunakan teknik parameter sharing sehingga lebih efisien dalam penggunaan memori dan komputasi. Penelitian Darma Pamuja et al. (2025) menunjukkan bahwa ALBERT tetap mampu menghasilkan performa yang kompetitif dalam tugas analisis sentimen meskipun memiliki ukuran model yang lebih ringan dibandingkan BERT.

2.2.6 DistilBERT

DistilBERT merupakan versi ringan dari BERT yang dibangun menggunakan teknik *knowledge distillation*. Model ini dirancang untuk mempertahankan sebagian besar kemampuan BERT dengan kebutuhan komputasi yang lebih rendah. Fajri et al. (2022) serta Dewangga & Prasetyo (2025) menunjukkan bahwa DistilBERT mampu memberikan performa yang kompetitif pada berbagai tugas analisis sentimen dengan waktu inferensi yang lebih cepat dibandingkan BERT standar.

Selain itu, studi komparatif terhadap BERT, RoBERTa, dan DistilBERT menunjukkan bahwa ketiga model tersebut merupakan pendekatan yang paling banyak digunakan dalam analisis sentimen berbasis Transformer karena keseimbangan antara akurasi dan efisiensi komputasi.

2.3 Penelitian Terdahulu

Perkembangan ChatGPT telah mendorong munculnya berbagai penelitian yang berfokus pada analisis sentimen publik menggunakan model berbasis Transformer. Sebagian besar penelitian memanfaatkan data dari media sosial seperti Twitter/X, Reddit, dan platform ulasan pengguna untuk mengukur persepsi masyarakat terhadap teknologi *Generative AI*. Berbagai model seperti BERT, RoBERTa, DistilBERT, ALBERT, hingga model hibrida telah digunakan untuk meningkatkan performa klasifikasi sentimen.

Beberapa penelitian menunjukkan bahwa BERT dan variannya masih mendominasi penelitian analisis sentimen karena mampu menghasilkan nilai akurasi dan F1-score yang tinggi. Tetapi, berbagai tantangan masih ditemukan, terutama dalam menangani sarkasme, bahasa informal, ketidakseimbangan data, serta

konteks multibahasa (ElAzzouzy et al., 2025; Muhammad & Rospocher, 2025).

Meskipun jumlah penelitian terus meningkat, sebagian besar studi masih berfokus pada evaluasi model tertentu atau dataset yang spesifik. Oleh karena itu, diperlukan kajian *Systematic Literature Review* (SLR) yang mampu merangkum tren penelitian, model yang digunakan, sumber data, metrik evaluasi, serta tantangan yang masih dihadapi dalam analisis sentimen terhadap ChatGPT. Temuan inilah yang menjadi dasar dilaksanakannya penelitian ini.

3. Metodologi

Penelitian ini menggunakan metode *Systematic Literature Review* (SLR) untuk mengidentifikasi, mengevaluasi, dan mensintesis penelitian terkait analisis sentimen publik terhadap ChatGPT menggunakan arsitektur Transformer. Metode SLR dipilih karena mampu memberikan kajian yang sistematis, transparan, dan dapat direplikasi. Proses penelitian mengacu pada pedoman PRISMA 2020 (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*) yang terdiri atas tahapan identifikasi, penyaringan, penilaian kelayakan, dan inklusi artikel. Tahapan penelitian yang dilakukan meliputi perumusan *research question*, strategi pencarian literatur, penentuan kriteria inklusi dan eksklusi, *quality assessment*, serta proses seleksi artikel menggunakan diagram PRISMA. Tools yang digunakan dalam penelitian ini meliputi Scopus sebagai sumber data literatur dan Mendeley Reference Manager sebagai perangkat lunak manajemen referensi untuk membantu proses penyimpanan, pengorganisasian, deduplikasi, serta pengelolaan sitasi dan daftar pustaka.

3.1 Perumusan Research Question

Research Question (RQ) dirumuskan untuk memastikan bahwa proses pencarian, seleksi, dan analisis literatur dilakukan secara terarah sesuai tujuan penelitian. Penyusunan RQ mengacu pada pendekatan PICOC (*Population, Intervention, Comparison, Outcomes, Context*) yang umum digunakan dalam penelitian SLR. Berdasarkan tujuan penelitian, dirumuskan empat pertanyaan penelitian sebagai berikut:

RQ1: Bagaimana tren publikasi penelitian terkait analisis sentimen publik terhadap ChatGPT menggunakan arsitektur Transformer?

RQ2: Model Transformer dan varian BERT apa yang paling banyak digunakan dalam penelitian analisis sentimen terhadap ChatGPT?

RQ3: Sumber data, *platform*, dan bahasa apa yang paling sering digunakan dalam penelitian analisis sentimen terhadap ChatGPT?

RQ4: Apa saja keterbatasan, tantangan, dan peluang penelitian yang diidentifikasi dalam studi-studi terkait?

Keempat pertanyaan penelitian tersebut menjadi dasar dalam proses ekstraksi data dan sintesis hasil penelitian.

3.2 Strategi Pencarian Literatur

Pencarian literatur dilakukan menggunakan basis data Scopus karena memiliki cakupan publikasi ilmiah yang luas serta menyediakan artikel berkualitas tinggi yang telah melalui proses peer review. Proses pencarian dilakukan pada periode Januari 2022 hingga Februari 2026.

Strategi pencarian menggunakan kombinasi kata kunci yang berkaitan dengan analisis sentimen, ChatGPT, dan model Transformer. Kata kunci dikombinasikan menggunakan operator *Boolean AND* dan *OR* untuk memperoleh artikel yang relevan dengan topik penelitian.

String pencarian yang digunakan adalah sebagai berikut:

("Sentiment Analysis" OR "Opinion Mining") AND

("ChatGPT" OR "Large Language Models") AND

("BERT" OR "RoBERTa" OR "DistilBERT" OR "ALBERT" OR "Transformer")

Strategi ini dirancang untuk memperoleh cakupan literatur yang luas namun tetap relevan dengan fokus penelitian.

3.3 Kriteria Inklusi dan Eksklusi

Setelah proses pencarian dilakukan, artikel diseleksi berdasarkan kriteria inklusi dan eksklusi yang telah ditentukan sebelumnya.

Tabel 1 Kriteria Inklusi dan Eksklusi

Kriteria Inklusi	Kriteria Eksklusi
Artikel membahas analisis sentimen terhadap ChatGPT atau Large Language Models	Artikel tidak membahas ChatGPT atau analisis sentimen
Menggunakan model berbasis Transformer (BERT, RoBERTa, DistilBERT, ALBERT, dan sejenisnya)	Tidak menggunakan pendekatan Transformer
Artikel penelitian asli (<i>primary study</i>)	Editorial, review, book chapter, atau artikel non-penelitian

Dipublikasikan pada jurnal atau prosiding terindeks Scopus Artikel yang tidak tersedia secara lengkap (*full text unavailable*)

Dipublikasikan pada periode 2022–2026 Publikasi di luar rentang waktu penelitian

Kriteria tersebut digunakan untuk memastikan bahwa artikel yang dianalisis memiliki relevansi dan kualitas yang sesuai dengan tujuan penelitian.

3.4 Quality Assessment

Untuk menjamin kualitas studi yang dianalisis, setiap artikel yang lolos tahap seleksi dievaluasi menggunakan proses *Quality Assessment (QA)*. Evaluasi dilakukan menggunakan empat pertanyaan penilaian yang disesuaikan dengan tujuan penelitian.

Tabel 2 Quality Assessment Criteria

Kode	Pertanyaan
QA1	Apakah tujuan penelitian dijelaskan dengan jelas?
QA2	Apakah metodologi yang digunakan dijelaskan secara memadai?
QA3	Apakah hasil penelitian dilaporkan dengan jelas dan dapat dievaluasi?
QA4	Apakah penelitian memberikan kontribusi yang relevan terhadap analisis sentimen berbasis Transformer?

Setiap pertanyaan diberikan skor menggunakan kategori:

- Yes (Y) = 1
- Partially (P) = 0,5
- No (N) = 0

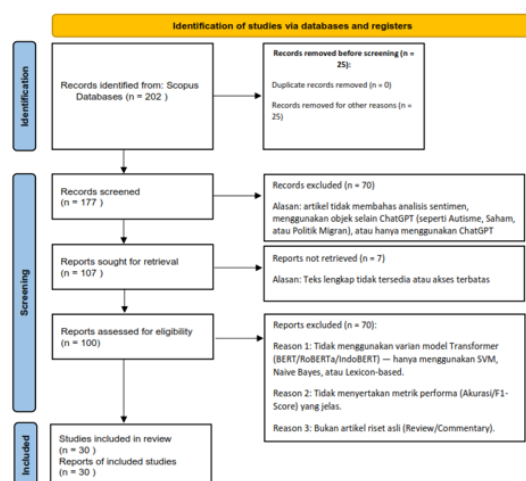
Hasil penilaian digunakan untuk memastikan bahwa artikel yang digunakan dalam proses sintesis memiliki kualitas metodologis yang memadai.

3.5 Proses Seleksi PRISMA

Proses seleksi artikel dilakukan mengikuti tahapan PRISMA 2020 yang terdiri atas empat fase utama, yaitu *Identification*, *Screening*, *Eligibility*, dan *Inclusion*.

Pada tahap *Identification*, diperoleh 202 artikel dari hasil pencarian pada basis data Scopus menggunakan string pencarian yang telah ditentukan. Selanjutnya dilakukan penghapusan artikel duplikat dan artikel yang tidak relevan berdasarkan judul serta abstrak pada tahap *Screening*.

Tahap *Eligibility* dilakukan dengan membaca teks lengkap artikel yang tersisa untuk memastikan kesesuaiannya dengan kriteria inklusi dan eksklusi. Artikel yang tidak memenuhi persyaratan, tidak menggunakan model Transformer, atau tidak membahas analisis sentimen terhadap ChatGPT dikeluarkan dari proses seleksi.



Gambar 1. Diagram PRISMA 2020

Hasil akhir dari proses tersebut adalah 30 artikel yang memenuhi seluruh kriteria dan digunakan sebagai objek analisis dalam penelitian ini. Alur lengkap proses seleksi artikel ditunjukkan pada diagram PRISMA yang disajikan pada Gambar 1.

Dengan tahapan tersebut, proses seleksi literatur dapat dilakukan secara sistematis, transparan, dan dapat direproduksi oleh peneliti lain.

4. Hasil dan Pembahasan

Hasil dari proses *Systematic Literature Review* (SLR) yang telah dilakukan, diperoleh sebanyak 30 artikel dipilih untuk dianalisis lebih lanjut. Analisis ini bertujuan untuk menjawab pertanyaan penelitian (RQ1–RQ4) yang telah dirumuskan, mencakup tren publikasi, metode yang digunakan, sumber data, serta keterbatasan dalam analisis sentimen terhadap ChatGPT berbasis arsitektur transformer.

4.1 Tren Publikasi Penelitian (RQ1)

Berdasarkan hasil analisis terhadap 30 artikel terpilih, penelitian mengenai analisis sentimen terhadap ChatGPT menunjukkan peningkatan signifikan sejak tahun 2023. Hal ini sejalan dengan meningkatnya popularitas ChatGPT sebagai teknologi *generative AI* yang digunakan secara luas oleh masyarakat global. Berdasarkan tahun publikasi seluruh 30 artikel, diperoleh distribusi sebagai berikut: tidak ada artikel yang berhasil lolos kriteria inklusi pada tahun 2022 (0,0%), meskipun ChatGPT resmi diluncurkan pada November 2022. Pada tahun 2023 terdapat 4 artikel (13,3%), kemudian melonjak signifikan pada tahun 2024 dengan 16 artikel (53,3%), diikuti tahun 2025 sebanyak 9 artikel (30,0%), dan 1 artikel (3,3%) dari tahun 2026.

Grafik tren publikasi artikel per tahun ditunjukkan pada Gambar 2.

Pola distribusi ini menggambarkan adanya *academic lag* atau jeda waktu antara kemunculan fenomena dengan lahirnya publikasi ilmiah yang mengkajinya. ChatGPT membutuhkan sekitar satu tahun sebelum akhirnya menghasilkan gelombang publikasi akademik yang masif. Tahun 2024 menjadi puncak produktivitas dengan lebih dari separuh total artikel (53,3%), mencerminkan percepatan riset seiring tersedianya data yang lebih luas, model *pre-trained* multibahasa, serta meningkatnya ketersediaan dataset opini publik dari berbagai platform.

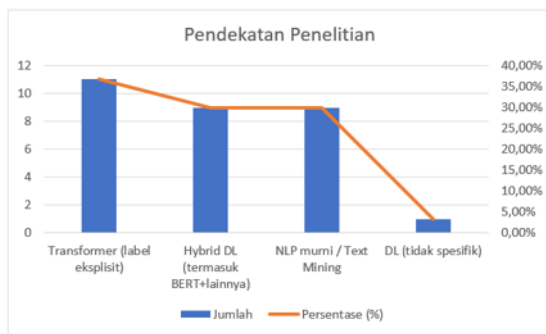


Gambar 2. Tren Publikasi Artikel Per Tahun

Sebagian besar penelitian awal berfokus pada eksplorasi opini publik di media sosial, khususnya Twitter dan Reddit (Koonchanok et al., 2024; Xu et al., 2024a). Seiring waktu, penelitian mulai berkembang dengan memanfaatkan pendekatan *deep learning*, terutama model berbasis transformer seperti BERT.

Pada periode 2024–2025, terlihat peningkatan penggunaan model berbasis BERT dan variannya dalam berbagai konteks aplikasi, seperti pendidikan, kepercayaan publik, dan evaluasi teknologi AI (Danish et al., 2025; Lyu et al., 2025). Hal ini menunjukkan bahwa tren penelitian bergerak dari sekadar analisis deskriptif menuju pendekatan yang lebih kompleks dan berbasis model *state-of-the-art*.

Selain tren tahun, RQ1 juga mengkaji distribusi pendekatan yang digunakan. Berdasarkan kolom *Model Category* dalam tabel dokumen, dari 30 artikel terpilih diperoleh distribusi sebagai berikut, yang ditunjukkan pada Gambar 3.



Gambar 3. Distribusi Pendekatan Penelitian

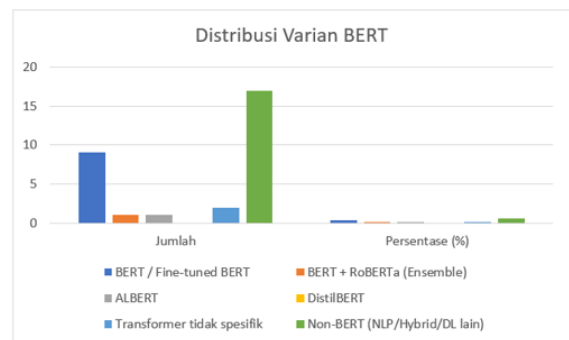
Transformer menjadi pendekatan paling dominan dengan 11 artikel (36,7%). Jika artikel *Hybrid* yang secara eksplisit mengintegrasikan komponen BERT yang menggunakan BERT+VADER ikut diperhitungkan, maka total artikel yang memanfaatkan arsitektur berbasis BERT mencapai 12 artikel (40,0%). Hal ini menunjukkan dominasi nyata pendekatan transformer, meskipun belum mencapai mayoritas absolut. Pendekatan Hybrid DL dan NLP murni masing-masing berkontribusi 9 artikel (30,0%), mengindikasikan bahwa eksplorasi metodologis masih beragam dalam penelitian analisis sentimen terhadap ChatGPT.

4.2 Varian Model BERT dan Metrik Evaluasi (RQ2)

Hasil analisis menunjukkan bahwa model berbasis BERT merupakan pendekatan yang paling dominan dalam penelitian analisis sentimen terhadap ChatGPT. Beberapa varian model yang paling sering digunakan meliputi:

- BERT sebagai model dasar (baseline) (Agus Surya Darma et al., 2023; Sudheesh et al., 2023),
- RoBERTa, yang memiliki performa lebih baik karena optimasi pelatihan (Mujahid et al., 2023)
- ALBERT, yang efektif dalam menangani teks Panjang (Safira et al., 2025)

Selain penggunaan model tunggal, beberapa penelitian juga mengadopsi pendekatan ensemble dan hybrid untuk meningkatkan akurasi klasifikasi sentimen (C & Kizhakethottam, 2025). Grafik distribusi varian BERT ditunjukkan pada gambar 4.



Gambar 4. Distribusi Varian BERT

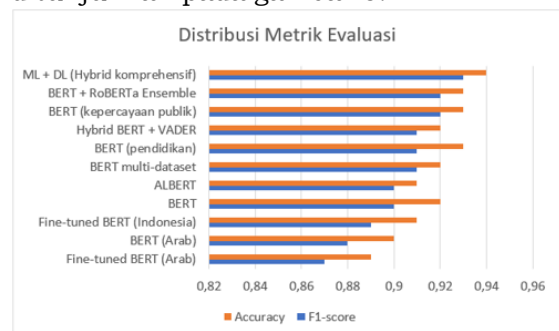
BERT dalam berbagai variannya, baik murni maupun fine-tuned mendominasi dengan 9 artikel (30,0%) dari total 30 artikel.

Dari 30 artikel yang dianalisis, hanya 11 artikel (36,7%) yang secara eksplisit melaporkan metrik evaluasi berupa *F1-score* atau *Accuracy*. Sebanyak 19 artikel (63,3%) tidak mencantumkan nilai metrik secara kuantitatif.

Dari sisi evaluasi, metrik yang paling umum digunakan adalah:

- *Accuracy*
- *Precision*
- *Recall*
- *F1-score*

Di antara metrik tersebut, *F1-score* merupakan metrik yang paling dominan karena mampu memberikan keseimbangan antara *precision* dan *recall*, terutama pada dataset yang tidak seimbang (Issa & Abdu Ibrahim, 2024). Beberapa studi juga menggunakan metrik tambahan seperti *AUC* untuk evaluasi yang lebih komprehensif (Elmadya & Oktavia, 2025). Distribusi metrik evaluasi ditunjukkan pada gambar 5.



Gambar 5. Distribusi Metrik Evaluasi

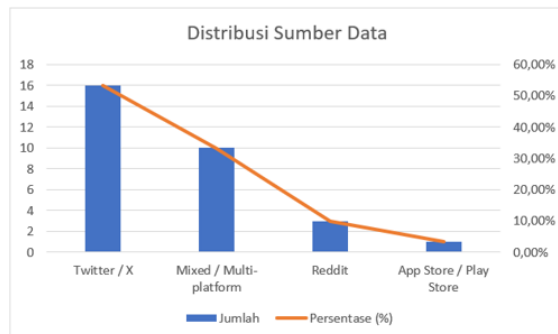
Rata-rata *F1-score* dari 11 artikel yang melaporkan metrik adalah 0,904 dengan rentang antara 0,87 hingga 0,93. Rata-rata *Accuracy* adalah 91,5% dengan rentang 89% hingga 94%. *F1-score* merupakan metrik yang paling dominan digunakan karena mampu memberikan keseimbangan antara *precision* dan *recall*, terutama pada dataset yang tidak

seimbang. Performa tertinggi dicapai oleh Elmadya & Oktavia (2025) yaitu dengan nilai F1: 0,93; *Accuracy*: 94% yang menggunakan pendekatan *hybrid* komprehensif (Elmadya & Oktavia, 2025), sementara performa terendah tercatat pada penelitian M. Hannani et al. yaitu dengan nilai F1: 0,87; *Accuracy*: 89% yang berkaitan dengan tantangan bahasa Arab sebagai bahasa dengan sumber daya terbatas (Hannani et al., 2024).

4.3 Sumber Data dan Bahasa (RQ3)

Sebagian besar penelitian menggunakan data yang berasal dari media sosial, yang mencerminkan opini publik secara langsung dan *real-time*. Platform yang paling sering digunakan meliputi:

- Twitter/X, sebagai sumber data utama karena kemudahan akses dan volume data yang besar (Alobedalla et al., 2025; Danish et al., 2025)
- Reddit, yang menyediakan diskusi yang lebih mendalam dan kontekstual (Naing & Udomwong, 2024b; Xu, Fang, Huang, & Xie, 2024b)
- Ulasan aplikasi (*App Store/Play Store*), yang digunakan untuk menganalisis pengalaman pengguna terhadap ChatGPT (Elmadya & Oktavia, 2025). Grafik distribusi sumber data ditunjukkan pada gambar 6.



Gambar 6. Distribusi Sumber Data

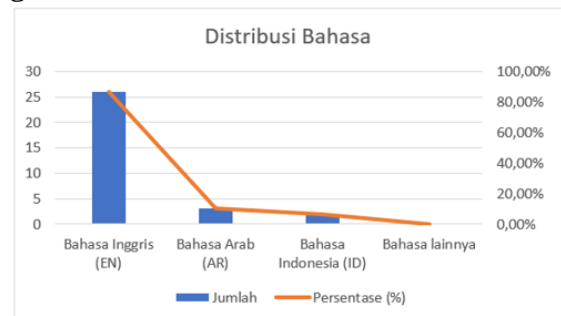
Twitter/X mendominasi sebagai sumber data dengan 16 artikel (53,3%), menjadikannya *platform* yang paling banyak digunakan dalam penelitian analisis sentimen terhadap ChatGPT. Platform *Mixed/Multi-platform* menempati posisi kedua dengan 10 artikel (33,3%), sedangkan Reddit hanya digunakan dalam 3 artikel (10,0%), dan *App Store/Play Store* hanya dalam 1 artikel (3,3%). Ketergantungan yang tinggi pada satu platform (Twitter/X) berpotensi membatasi generalisasi hasil penelitian, mengingat opini publik terhadap ChatGPT juga

berkembang secara aktif di berbagai platform lain.

Dari sisi bahasa, mayoritas penelitian menggunakan:

- Bahasa Inggris sebagai bahasa dominan
- Bahasa Arab untuk konteks regional tertentu (Alobedalla et al., 2025; Hannani et al., 2024)
- Bahasa Indonesia dalam studi local (Agus Surya Darma et al., 2023; Sinaga et al., 2025).

Grafik distribusi Bahasa ditunjukkan pada gambar 7.



Gambar 7. Distribusi Bahasa

Bahasa Inggris mendominasi secara sangat signifikan dengan 26 artikel (86,7%). Bahasa Arab menempati posisi kedua dengan 3 artikel (10,0%), sementara Bahasa Indonesia hanya digunakan dalam 2 artikel (6,7%). Tidak ditemukan penelitian yang menganalisis bahasa lain seperti Mandarin, Spanyol, Prancis, atau bahasa-bahasa Asia Tenggara lainnya. Dominasi bahasa Inggris yang mencapai 86,7% mengindikasikan adanya kesenjangan penelitian yang signifikan dalam konteks multibahasa, padahal penggunaan ChatGPT bersifat global dan melibatkan beragam latar belakang linguistik dan budaya.

Hal ini menunjukkan bahwa penelitian analisis sentimen terhadap ChatGPT bersifat global dan melibatkan berbagai konteks linguistik dan budaya.

4.4 Tantangan dan Peluang Penelitian (RQ4)

Meskipun model berbasis BERT menunjukkan performa yang tinggi dalam analisis sentimen, beberapa keterbatasan masih ditemukan dalam penelitian yang dianalisis.

Salah satu tantangan utama adalah kesulitan model dalam memahami sarkasme, ironi, dan bahasa gaul yang sering muncul dalam teks media sosial. Hal ini disebabkan oleh keterbatasan model dalam memahami konteks pragmatik dan nuansa emosional yang

kompleks (Hannani et al., 2024; Xing et al., 2024).

Selain itu, terdapat beberapa keterbatasan lain, seperti:

- Sarkasme dan Bahasa informal
- Dominasi Bahasa Inggris merupakan gap multibahasa
- Adanya penelitian yang tidak melaporkan evaluasi eksplisit
- Ketergantungan pada Twitter/X saja

Grafik distribusi keterbatasan model ditunjukkan pada gambar 7.



Gambar 7. Distribusi Keterbatasan

Dari keempat keterbatasan di atas, kesulitan dalam menangani sarkasme dan bahasa informal merupakan tantangan yang paling universal, serta diakui secara implisit oleh seluruh artikel (100%) yang menganalisis data media sosial. Hal ini disebabkan oleh keterbatasan model dalam memahami konteks pragmatik dan nuansa emosional yang kompleks. Keterbatasan kedua adalah dominasi bahasa Inggris sebesar 86,7%, yang menunjukkan bahwa sebagian besar penelitian belum mampu merepresentasikan pengguna global ChatGPT secara proporsional. Sebanyak 63,3% artikel tidak melaporkan metrik evaluasi secara eksplisit, mengindikasikan kurangnya standarisasi dalam pelaporan hasil penelitian. Terakhir, 53,3% artikel hanya mengandalkan Twitter/X sebagai satu-satunya sumber data, yang berpotensi menghasilkan bias platform dalam pemodelan opini publik.

Berdasarkan temuan kuantitatif dan kualitatif di atas, beberapa arah penelitian selanjutnya yang direkomendasikan adalah:

- Pengembangan model yang lebih kontekstual dan adaptif terhadap sarkasme, ironi, dan bahasa informal yang umum ditemukan pada media sosial, misalnya melalui pendekatan *context-aware* atau integrasi pengetahuan eksternal.
- Perluasan cakupan multibahasa dan lintas budaya, mengingat 86,7% penelitian saat ini masih berfokus

pada bahasa Inggris, sementara bahasa Arab hanya 10,0% dan bahasa Indonesia hanya 6,7%.

- Standardisasi pelaporan metrik evaluasi, mengingat 63,3% artikel tidak melaporkan F1-score maupun *Accuracy* secara eksplisit untuk meningkatkan komparabilitas dan reproduktibilitas hasil penelitian.
- Diversifikasi sumber data melampaui Twitter/X (yang saat ini mendominasi 53,3%) untuk mencakup Reddit, *App Store/Play Store*, YouTube, dan forum diskusi akademik.
- Integrasi pendekatan hybrid dan ensemble yang terbukti menghasilkan performa kompetitif.
- Eksplorasi model yang lebih ringan dan efisien seperti DistilBERT untuk mendukung implementasi pada skala besar, mengingat model ini belum ditemukan dalam satu pun artikel yang dianalisis dalam SLR ini.
- Selain itu, pemanfaatan model generatif seperti ChatGPT dalam proses anotasi data juga menjadi peluang penelitian yang menjanjikan di masa depan (Mathebula, Modupe, & Marivate, 2024).

4.5 Diskusi Temuan

4.5.1 Dominasi Model Transformer dalam Analisis Sentimen ChatGPT

Dominasi penggunaan model berbasis transformer, khususnya BERT dan variannya, menunjukkan bahwa penelitian dalam domain ini telah mengadopsi pendekatan *state-of-the-art* dalam *Natural Language Processing*. Dari 30 artikel yang dianalisis, 11 artikel (36,7%) secara eksplisit berlabel Transformer dalam kolom *Model Category*. Jika digabungkan dengan artikel *Hybrid* yang mengintegrasikan komponen BERT, total artikel berbasis BERT mencapai 12 artikel (40,0%), dengan BERT murni dan fine-tuned BERT mendominasi sebesar 30,0% (9 artikel). ALBERT dan ensemble BERT+RoBERTa masing-masing tercatat dalam 1 artikel (3,3%).

4.5.2 Pengaruh Kualitas Dataset terhadap Kinerja Model

Kualitas dataset menjadi faktor krusial yang mempengaruhi performa model, sehingga banyak penelitian masih bergantung pada data media sosial yang bersifat *noisy*, tidak terstruktur, dan memiliki ketidakseimbangan distribusi kelas. Proses anotasi data juga menjadi tantangan tersendiri karena membutuhkan konsistensi antar anotator,

sementara subjektivitas dalam penilaian sentimen sering kali sulit dihindari.

Dari sisi sumber data, dominasi penggunaan Twitter/X yang mencapai 53,3% (16 dari 30 artikel) menunjukkan bahwa penelitian masih terfokus pada satu jenis *platform*. Hal ini berpotensi membatasi generalisasi hasil, mengingat opini publik terhadap ChatGPT juga berkembang aktif pada berbagai platform lain. Reddit hanya digunakan dalam 3 artikel (10,0%), sementara *App Store/Play Store* hanya dalam 1 artikel (3,3%), menunjukkan masih besarnya ruang eksplorasi pada *platform* yang lebih beragam.

4.5.3 Kesenjangan Penelitian Multibahasa

Kesenjangan multibahasa juga menjadi temuan penting dalam SLR ini. Bahasa Inggris mendominasi 86,7% artikel (26 dari 30), sementara bahasa Arab hanya 10,0% (3 artikel) dan bahasa Indonesia hanya 6,7% (2 artikel). Tidak ditemukan satu pun artikel yang menganalisis bahasa lain seperti Mandarin, Spanyol, Prancis, atau bahasa-bahasa Asia Tenggara. Padahal penggunaan ChatGPT bersifat global dan melibatkan berbagai latar belakang linguistik dan budaya, sehingga kesenjangan ini berimplikasi langsung pada keterbatasan representasi dan generalisasi temuan.

4.5.4 Implikasi dan Arah Penelitian Masa Depan

- Berdasarkan temuan kuantitatif dan kualitatif di atas, beberapa arah penelitian selanjutnya yang dapat dikembangkan antara lain:
- Pengembangan model yang lebih kontekstual. Diperlukan model yang mampu memahami konteks pragmatik seperti sarkasme, ironi, dan bahasa informal secara lebih akurat, misalnya melalui pendekatan *context-aware* atau integrasi pengetahuan *eksternal*. Tantangan ini diakui oleh 100% artikel yang menganalisis teks media sosial.
- Pendekatan multibahasa dan lintas budaya. Penelitian perlu memperluas cakupan bahasa untuk menghasilkan model yang lebih representatif terhadap pengguna global. Saat ini 86,7% penelitian masih berfokus pada bahasa Inggris, sementara bahasa Arab hanya 10,0% dan bahasa Indonesia hanya 6,7%.
- Peningkatan kualitas dan standarisasi dataset. Pengembangan dataset dengan anotasi yang lebih konsisten

dan representatif sangat diperlukan. Sebanyak 63,3% artikel tidak melaporkan metrik evaluasi secara eksplisit, sehingga standarisasi pelaporan *F1-score* dan *Accuracy* perlu menjadi prioritas untuk meningkatkan komparabilitas dan reproduktibilitas hasil penelitian.

- Integrasi model *hybrid* dan *ensemble*. Kombinasi berbagai pendekatan terbukti menghasilkan performa kompetitif.
- Diversifikasi sumber data. Penelitian selanjutnya perlu memanfaatkan berbagai sumber data melampaui Twitter/X yang saat ini mendominasi 53,3%, untuk mencakup Reddit (saat ini hanya 10,0%), *App Store/Play Store* (3,3%), YouTube, dan forum diskusi akademik.
- Eksplorasi DistilBERT dan model efisien lainnya. Pengembangan model yang lebih ringan dan efisien penting untuk mendukung implementasi pada skala besar. DistilBERT belum ditemukan sebagai metode utama dalam satu pun artikel dari 30 artikel yang dianalisis dalam SLR ini, sehingga terdapat peluang besar untuk eksplorasi lebih lanjut.
- Pemanfaatan model generatif sebagai anotator. Pemanfaatan ChatGPT dalam proses anotasi data juga menjadi peluang penelitian yang menjanjikan di masa depan, sebagaimana ditunjukkan oleh referensi (Mathebula et al., 2024) yang menggunakan ChatGPT sebagai alat anotasi teks.

5. Simpulan

Penelitian ini menyajikan *Systematic Literature Review* (SLR) terhadap analisis sentimen publik terhadap ChatGPT menggunakan pendekatan berbasis Transformer, khususnya BERT dan variannya, dengan menganalisis 30 artikel yang memenuhi kriteria inklusi dan eksklusi. Hasil penelitian menunjukkan bahwa bidang ini mengalami pertumbuhan yang pesat setelah peluncuran ChatGPT, ditandai dengan peningkatan publikasi dari 13,3% pada tahun 2023 menjadi 53,3% pada tahun 2024. Dari sisi metodologi, model berbasis BERT mendominasi penelitian dengan proporsi 40,0%, diikuti oleh ALBERT dan pendekatan *ensemble*, sementara evaluasi menunjukkan rata-rata *F1-score* sebesar 0,904 dan *accuracy* sebesar 91,5%, yang

mengindikasikan kinerja model Transformer yang sangat baik dalam tugas analisis sentimen. Twitter/X merupakan sumber data yang paling banyak digunakan (53,3%), sedangkan bahasa Inggris mendominasi dataset penelitian (86,7%), menunjukkan masih adanya kesenjangan penelitian multibahasa. Selain itu, penelitian ini mengidentifikasi beberapa tantangan utama, yaitu kesulitan dalam menangani sarkasme dan bahasa informal, dominasi bahasa Inggris, kurangnya standarisasi pelaporan metrik evaluasi, serta ketergantungan pada satu platform data yang berpotensi menimbulkan bias. Secara keseluruhan, penelitian ini memberikan gambaran secara menyeluruh mengenai tren publikasi, pendekatan metodologi, performa model, sumber data, distribusi bahasa, serta keterbatasan yang ada dalam penelitian analisis sentimen terhadap ChatGPT. Temuan ini diharapkan dapat menjadi referensi bagi penelitian selanjutnya dalam mengembangkan model yang lebih adaptif terhadap konteks bahasa, memperluas cakupan multibahasa, meningkatkan standarisasi evaluasi, mendiversifikasi sumber data, serta mengeksplorasi varian Transformer yang lebih efisien dan belum banyak diteliti.

Daftar Pustaka

- Agus Surya Darma, I. W., Mahendra Putra, P. R., Sugiartawan, P., Waas, V., & Sutramiani, N. P. (2023). A Fine-tuned BERT-based Approach for Sentiment Analysis of Indonesian Public Towards ChatGPT. *2023 International Conference on Smart-Green Technology in Electrical and Information Systems (ICSGTEIS)*, 88–93. IEEE. <https://doi.org/10.1109/ICSGTEIS60500.2023.10424123>
- Alobedalla, A., Hatamleh, A., Alawneh, M., Alsawaftah, H., & Obeidat, R. (2025). ChatGPT and Public Emotion: A Study of Arabic Opinions on Twitter. *2025 16th International Conference on Information and Communication Systems (ICICS)*, 1–6. IEEE. <https://doi.org/10.1109/ICICS65354.2025.11072958>
- Amrullah, A. (2025). *Sentiment Analysis in the Age of Transformers and Large Language Models: A Comprehensive Review of Recent Advances and Future* (Vol. 1).
- Bin-Nashwan, S. A., Sadallah, M., & Bouteraa, M. (2023). Use of ChatGPT in academia: Academic integrity hangs in the balance. *Technology in Society*, 75, 102370. <https://doi.org/10.1016/j.techsoc.2023.102370>
- Danish, M., Amjad, M., Dyussekeyev, K., & Zhumadillayeva, A. (2025). Exploring ChatGPT's Role in Education: An NLP-Based Sentiment and Thematic Study Using Twitter Data. *Procedia Computer Science*, 259, 959–968. <https://doi.org/10.1016/j.procs.2025.04.049>
- Darma Pamuja, S., Studi Teknologi Informasi, P., Kesehatan dan Teknologi, F., & Muhammadiyah Klaten, U. (2025). *Optimasi Analisis Sentimen Ulasan Platform Pendidikan Daring Menggunakan Arsitektur ALBERT dan Teknik Augmentasi Kontekstual* (Vol. 01).
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. Retrieved from <http://arxiv.org/abs/1810.04805>
- Dewangga, B. ., & Prasetyo, R. T. (2025). *Comparative Analysis of BERT Model and DistilBERT Model for Enhanced Clickbait Headline Structure Detection in Indonesian Online News* (Vol. 13).
- El Azzouzy, O., Chanyour, T., & Andaloussi, S. J. (2025). Transformer-based models for sentiment analysis of YouTube video comments. *Scientific African*, 29. <https://doi.org/10.1016/j.sciaf.2025.e02836>
- Elmadya, D., & Oktavia, T. (2025). Comprehensive Sentiment Analysis of ChatGPT App Reviews Using Lexicon-Based Methods, Deep Learning, and Machine Learning. *2025 International Conference on Computer Sciences, Engineering, and Technology Innovation (ICoCSETI)*, 786–791. IEEE. <https://doi.org/10.1109/ICoCSETI63724.2025.11019303>
- Fahmi, M. I., & Nababan, A. A. (2025). A Comparative Study of BERT and RoBERTa for Sentiment Analysis on Twitter Data Related to Mental Health. *Mathematical Modelling of Engineering Problems*, 12(9), 3289–3295. <https://doi.org/10.18280/mmep.120931>
- Fajri, F., Tutuko, B., & Sukemi, S. (2022). *Membandingkan Nilai Akurasi BERT dan DistilBERT pada Dataset Twitter*. 8(2), 71–80.
- Hannani, M., Soudi, A., & Van Laerhoven, K. (2024). Assessing the Performance of ChatGPT-4, Fine-tuned BERT and Traditional ML Models on Moroccan Arabic Sentiment Analysis. *Proceedings of the 4th International Conference on Natural Language Processing for Digital Humanities*, 489–498. Stroudsburg, PA, USA: Association for Computational Linguistics.

<https://doi.org/10.18653/v1/2024.nlp4dh-1.47>

- Issa, E., & Abdu Ibrahim, A. (2024). Integrating BERT for Nuanced Sentiment Analysis: A Detailed Examination of Diverse Textual Datasets. *IEEE Access*, 12, 186296–186312. <https://doi.org/10.1109/ACCESS.2024.3510944>
- Koonchanok, R., Pan, Y., & Jang, H. (2024). Public attitudes toward chatgpt on twitter: sentiments, topics, and occupations. *Social Network Analysis and Mining*, 14(1), 106. <https://doi.org/10.1007/s13278-024-01260-7>
- Leiter, C., Zhang, R., Chen, Y., Belouadi, J., Larionov, D., Fresen, V., & Eger, S. (2024). ChatGPT: A meta-analysis after 2.5 months. *Machine Learning with Applications*, 16, 100541. <https://doi.org/10.1016/j.mlwa.2024.100541>
- Lyu, Z., Lin, J., & Cheng, W. (2025). Public Trust in ChatGPT: A BERT-Based Framework for Analyzing Social Media Discourse. *International Journal of Human-Computer Interaction*, 1–20. <https://doi.org/10.1080/10447318.2025.2545454>
- Mathebula, M., Modupe, A., & Marivate, V. (2024). ChatGPT as a Text Annotation Tool to Evaluate Sentiment Analysis on South African Financial Institutions. *IEEE Access*, 12, 144017–144043. <https://doi.org/10.1109/ACCESS.2024.3464374>
- Muhammad, I., & Rospocher, M. (2025). On Assessing the Performance of LLMs for Target-Level Sentiment Analysis in Financial News Headlines. *Algorithms*, 18(1), 46. <https://doi.org/10.3390/a18010046>
- Mujahid, M., Kanwal, K., Rustam, F., Aljedaani, W., & Ashraf, I. (2023). Arabic ChatGPT Tweets Classification Using RoBERTa and BERT Ensemble Model. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 22(8), 1–23. <https://doi.org/10.1145/3605889>
- Naing, S. Z. S., & Udomwong, P. (2024a). Public Opinions on ChatGPT: An Analysis of Reddit Discussions by Using Sentiment Analysis, Topic Modeling, and SWOT Analysis. *Data Intelligence*, 6(2), 344–374. https://doi.org/10.1162/dint_a_00250
- Naing, S. Z. S., & Udomwong, P. (2024b). Public Opinions on ChatGPT: An Analysis of Reddit Discussions by Using Sentiment Analysis, Topic Modeling, and SWOT Analysis. *Data Intelligence*, 6(2), 344–374. https://doi.org/10.1162/dint_a_00250
- Padarha, S., & Vijaylakshmi, S. (2024). A Multifaceted Approach at Discerning Redditors Feelings Towards ChatGPT. *EAI Endorsed Transactions on Internet of Things*, 10. <https://doi.org/10.4108/eetiot.6447>
- Safira, W., Prabaswara, B., Stevens Karnyoto, A., & Pardamean, B. (2025). Leveraging ALBERT for Sentiment Classification of Long-Form ChatGPT Reviews on Twitter. *International Journal of Computing and Digital Systems*, 17(1), 1–12. <https://doi.org/10.12785/ijcds/1570999256>
- Sinaga, F. M., Pangaribuan, J. J., -, K., -, F., & Widjaja, A. E. (2025). Dynamic Sentiment Analysis on the Emergence of Pre-Trained Generative Model-Based Applications in Indonesia. *International Journal of Advanced Computer Science and Applications*, 16(12). <https://doi.org/10.14569/IJACSA.2025.01612111>
- Stryker, C., & Bergmann, D. (n.d.). What is a transformer model? Retrieved June 8, 2026, from <https://www.ibm.com/think/topics/transformer-model>
- Su, Y., & Kabala, Z. J. (2023). Public Perception of ChatGPT and Transfer Learning for Tweets Sentiment Analysis Using Wolfram Mathematica. *Data*, 8(12), 180. <https://doi.org/10.3390/data8120180>
- Sudheesh, R., Mujahid, M., Rustam, F., Shafique, R., Chunduri, V., Villar, M. G., ... Ashraf, I. (2023). Analyzing Sentiments Regarding ChatGPT Using Novel BERT: A Machine Learning Approach. *Information (Switzerland)*, 14(9). <https://doi.org/10.3390/info14090474>
- Triawan, P., Tahyudin, I., & Purwadi, P. (2025). Impact of NLP Algorithms on Sentiment Analysis Efficiency and Accuracy. *Journal of Information Systems and Informatics*, 7(3), 2684–2709. <https://doi.org/10.51519/journalisi.v7i3.1222>
- Xing, Y., Zhang, J. Z., Teng, G., & Zhou, X. (2024). Voices in the digital storm: Unraveling online polarization with ChatGPT. *Technology in Society*, 77, 102534. <https://doi.org/10.1016/j.techsoc.2024.102534>
- Xu, Z., Fang, Q., Huang, Y., & Xie, M. (2024a). The public attitude towards ChatGPT on reddit: A study based on unsupervised learning from sentiment analysis and topic modeling. *PLoS ONE*, 19(5 May). <https://doi.org/10.1371/journal.pone.0302502>
- Xu, Z., Fang, Q., Huang, Y., & Xie, M. (2024b). The public attitude towards ChatGPT on reddit: A

study based on unsupervised learning from sentiment analysis and topic modeling. *PLOS ONE*, 19(5), e0302502. <https://doi.org/10.1371/journal.pone.0302502>

Zhang, F., Guo, R., Xing, W., Zhu, W., Kim, T., & Liu, Z. (2024). WIP: From Tweets to Trends: Tracing the Public's Perception of AI in Education Post-ChatGPT. *2024 IEEE Frontiers in Education Conference (FIE)*, 1–5. IEEE. <https://doi.org/10.1109/FIE61694.2024.10893041>