International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(23s) 2026 pp. 1067-1087
https://doi.org/10.70917/ijcisim-2026-5675
Article

Transformer-Based Sentiment Analysis in Social Networks: A PRISMA-Based Systematic Literature Review

Nisha[1], Bal kishan[2]

[1]Research Scholar, Department of Computer Science & Applications, Maharshi Dayanand University, Rohtak, Haryana, India
[2]Associate Professor, Department of Computer Science & Applications, Maharshi Dayanand University, Rohtak, Haryana, India
balkishan@mdurohtak.ac.in
*Corresponding author E-mail: nisha.rs24.dcsa@mdurohtak.ac.in
ORCID: 0009-0003-5460-9815

Abstract: Sentiment analysis plays a significant role in understanding public opinion from the huge amount of data generated on review websites, social media platforms, and online forums. Recent advances in Natural Language Processing (NLP), particularly transformer-based models, have significantly enhanced sentiment classification by capturing contextual and semantic relationships from the data. This study presents a PRISMA based framework to examine recent research published during 2021 to 2026 on transformer based sentiment analysis. The outcomes indicate a strong shift toward fine-tuned pre-trained transformer models, such as BERT, RoBERTa, and domain-specific variants, along with increasing interest in large language model-based sentiment analysis. This study highlights the contributions of recent research, classifies the transformer architectures, contrast, reveals several critical challenges and future research directions, existing research gaps, trends, explainable, and scalable transformer-based sentiment analysis systems. This study provide up-to-date reference for researchers and practitioners seeking to develop scalable, accurate and effective sentiment analysis systems.
Keywords: Sentiment Analysis, Transformer Models, Social Networks, Natural Language Processing, Systematic Literature Review, Machine Learning, Deep Learning, Text Mining, BERT, RoBERTa

1. Introduction

The expansion of social networking sites has led to tremendous growth of user generated textual content such as daily post or reviews generated, comments, tweets and discussion forums as well. These have become one of the main area through which people can express their opinions, feelings, and even sentiments about products, services, social activities and policies in the society. Subsequently, the sentiment analysis has become a crucial research domain in gaining insight into the opinions of the population and contributing to decision making processes in such fields as marketing, business, healthcare, finance, and social governance [19]. The early sentiment analysis systems largely relied on lexicon-based and conventional machine learning algorithms, including Naïve Bayes, Support Vector Machines, and Logistic Regression [2][4]. Although the algorithms performed reasonably on structured and small-scale data, they failed to reflect the contextual meaning, implicit sentiment and sarcasm as well as long-range dependencies of social media text. Furthermore, Transformer-based architectures, such as BERT and Roberta, have recently been adopted [43]. Models such as BERT, GPT, RoBERTa, XLNet and its variations show improved performance in various sentiment analysis benchmarks [17]. Domain-specific transformer models, such as BERTweet, SciBERT, and BioBERT, were also used and improved the performance of sentiment analysis using task and domain-relevant linguistic patterns [18], [34]. In spite of their success, transformer-based sentiment analysis models have various challenges that are yet to be tackled especially in the social network setting. The data of social media are quite noisy, informal, and situational and most of them are full of emojis, slang, acronyms, and code-mixed language [26]. Besides, transformer models continue to have difficulties in identifying sarcasm, irony, and implicit sentiment, which is widespread in online communication [24]. Continued conversation chains and growing sentiment over time also present a challenge to typical transformer architecture with fixed input covering, leading to the introduction of long-sequence models, including Transformer-XL, Longformer, and BigBird [48].

Secondly, there has been recent research starting to consider multimodal sentiment analysis, which involves text and images or video to gain a deeper perception into the user's sentiments on the social platforms [10]. Multimodal transformer models like UNITER, CLIP, VideoBERT, and ERNIE-ViL have demonstrated promising performance [10]. Based on a PRISMA methodology, this study systematically studies the contributions of recent research, classifies the transformer architectures, contrasts their benefits and drawbacks, and reveals several critical challenges and future research directions.

1.1 Historical Evolution of Sentiment Analysis

Sentiment analysis has developed significant in the last twenty years, with the growing availability of textual data on online and offline resources. The early methods were mainly based on lexicon methods in which pre-existing sentiment dictionaries were used to determine the polarity of words or phrases. By these limitations, the traditional machine learning algorithms like Naive Bayes, Support Vector Machines and Logistic Regression were introduced. Such models made use of manually engineered features, such as n-grams and term frequency-inverse document frequency (TF-IDF), to enhance the accuracy of classification [4], [5]. Convolutional Neural Networks and Recurrent Neural Networks. Models such as LSTM or GRU were also providing improved sequence contextual understanding than the classical methods [7], [14]. Transformer-based models have established better effectiveness in sentiment analysis tasks, for example opposed to RNN-based models, as they also support parallel processing, and more importantly, they have the ability to keep contextual information [43]. Pretrained transformer models like BERT, GPT, RoBERTa, and XLNet have since formed the basis of current sentiment analysis systems, particularly for social network data [17], [18], [26]. Figure 1 presents the historical evolution of sentiment analysis techniques. It demonstrates how research has progressed from rule-based approaches to advanced transformer architectures and emerging intelligent methods, providing the foundation for the literature discussed in the following sections.

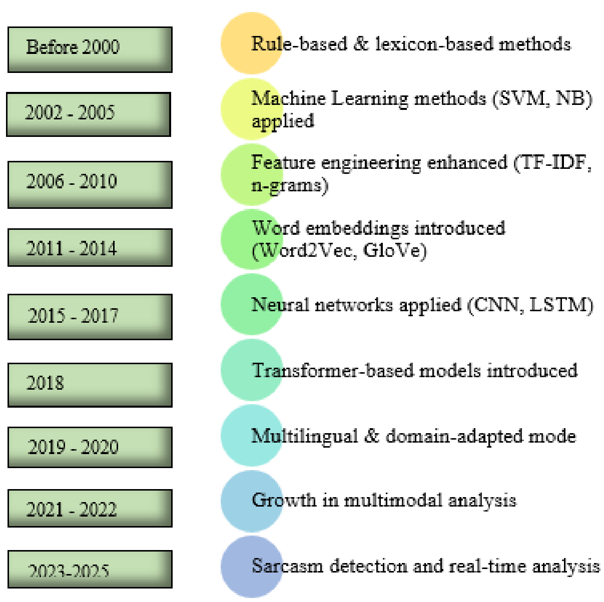

Figure 1 Foundation of Sentiment Analysis.

1.2 Research goal

The prime goal of this study is to systematically review recent advances in transformer-based sentiment analysis. Following the PRISMA methodology, a set of research questions was developed to describe the scope of the review and guide the systematic analysis of the selected studies. These questions focus on the key features of transformer-based sentiment analysis and are presented in Table1.

Table 1 Research Questions Related to Transformer-Based Sentiment Analysis

RQ No.	**Research Questions**	**Motivation / Purpose**
RQ1	What are the types of transformer-based models used in sentiment analysis?	To understand major transformer variants (BERT, RoBERTa, GPT, XLNet, etc.) and how improvements in architecture enhance sentiment interpretation.
RQ2	Why are transformer-based models preferred over traditional machine learning methods for sentiment analysis?	This question evaluates performance advantages such as contextual understanding, sarcasm detection, and deeper emotional analysis compared to traditional machine learning techniques.
RQ3	What challenges arise when applying transformer-based models to sentiment analysis in real-world scenarios?	To identify limitations such as high computational cost, domain adaptation issues, noisy user-generated data, and scarcity of high-quality labeled datasets.
RQ4	What are the emerging research directions for transformer-based sentiment analysis?	To highlight upcoming trends including lightweight models, low-resource languages, explainable AI, domain-specific transformers, and real-time large-scale deployment.

1.3 Contribution of Systematic literature review (SLR)

This study presents a systematic literature review of transformer-based sentiment analysis approaches applied to social network data. The main contributions of systematic review are summarized as follows:

- This study presents a comprehensive systematic literature review of transformer-based sentiment analysis models applied to social network data, following the PRISMA methodology to ensure transparency and reproducibility.
- The review provides a comparative analysis of state-of-the-art transformer models, including BERT variants, autoregressive models, long-context transformers, and multimodal transformers.
- Key application domains such as social media analysis, sarcasm detection, mental health assessment, and multilingual sentiment analysis are systematic review.
- Open challenges and future research directions are identified, with emphasis on explainability, ethical considerations, and low-resource language sentiment analysis.

2. Methodology of Systematic Review

This study follows a systematic literature review methodology to identify, evaluate, and synthesize existing research on transformer-based sentiment analysis. Figure 2 illustrates review process which was designed to be transparent, systematic and transparent, and reproducible by defining clear research objectives, selecting reliable academic databases, and applying appropriate keywords with Boolean operators. The relevance and quality of the selected studies are ensured with well-defined inclusion and exclusion criteria. Final selection of studies are all well recorded in order to be consistent and provide evidence-based conclusions. Every step of the review process, including study screening, eligibility assessment, and final inclusion is well documented to avoid any bias and maintain consistency [11], [12]. The selected studies were critically analyzed and helps to make conclusions and provide recommendations based on the collected evidence.

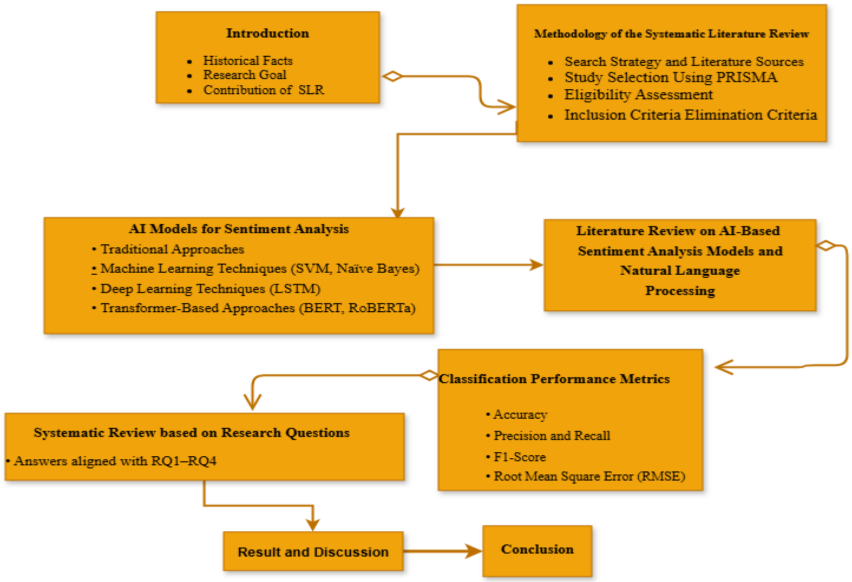

Figure 2 Systematic review of sentiment analysis

2.1 Search Strategy and Literature Sources

The study present a systematic search on studies on AI-based sentiment analysis. The four major databases are chosen through the use of the inclusion and exclusion criteria including IEEE Xplore, Scopus, ACL Anthology, and Google Scholar. Keywords like "BERT, sentiment analysis, deep learning, and transformer models" are combined in the form of a Boolean search query. Relevance and quality peer-reviewed English-language articles that were published during 2021-2026 were used to narrow down the search. This plan allowed covering all the latest trends and developments in sentiment analysis based on transformers in various areas of application. Figure 3 shows the database-based representation of the chosen studies, with the highlighting on the interdisciplinary scope of the sentiment analysis studies.

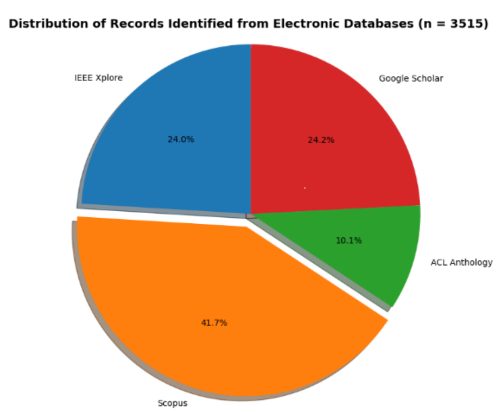

Figure 3 Database-wise distribution

2.2 Study Selection Using PRISMA

The process of data collection was started after the selection of keywords. To capture the current developments in transformer-based sentiment analysis, the literature published within the past six years (2021-2026) was considered in the review. The entire search process of the study, in accordance with PRISMA, is represented in Table 2.

Table 2 PRISMA Flow Summary for Study Selection

Stage	Process	Records	Exclusion Criteria	Excluded (n)	Remaining (n)
Identification	Records identified from IEEE Xplore, Scopus, ACL Anthology, and Google Scholar	3515	Non-English publications	120	3395
	Records imported into EndNote/Mendeley	3395	Duplicate records	535	2860
Screening	Title screening	2860	Not related to sentiment analysis, NLP, or transformer models	1450	1410
	Abstract screening	1410	Outside research scope, review/editorial, insufficient relevance	860	550
Eligibility	Full-text articles evaluated	550	Not primarily focused on transformer-based sentiment analysis, unavailable full text, or insufficient performance analysis	495	55
Inclusion	Studies included in the systematic review	**55**			**55**

2.2.1 Identification

A systematic search was conducted to find studies related to AI-based sentiment analysis. Four online databases, namely IEEE Xplore, Scopus, ACL Anthology, and Google Scholar, were searched, resulting in the identification of 3,515 records. Among these, 120 non-English publications were excluded. The remaining records is 3395 were imported into EndNote/Mendeley, where 535 duplicate records were identified and removed. As a result, 2,860 unique records were engaged for the screening phase. Boolean search strings were made using the AND operator by merging keywords such as "BERT", "sentiment analysis", "deep learning", and "transformer models". The search was classified to peer-reviewed publications written in English and published between 2021 and 2026, confirming the inclusion of recent and proper studies for the following screening and eligibility assessment [4, 5].

2.2.2 Screening

During the title screening stage, 2,860 records were estimated, and 1,450 studies were omitted because they were not related to sentiment analysis, transformer models or NLP, leaving 1,410 studies for further assessment. Then, abstract screening omitted 860 studies that were outside the research area, were review or editorial articles, or had insufficient relevance, resulting in 550 studies for the next stage of analysis.

2.3 Eligibility Assessment

During the eligibility stage, the full texts of 550 articles were estimated to determine their relevance and suitability for the study [7]. A total of 495 articles were omitted because they were not mainly focused on transformer-based sentiment analysis, had unavailable full texts, or provided insufficient performance analysis, resulting in 55 eligible studies for the final review.

2.4 Inclusion Criteria and Exclusion Criteria

To select high-quality and relevant and methodologically sound studies to include in this systematic review, inclusion and exclusion criteria have been developed. These parameters were used to get the process of screening and eligibility assessment where only recent and peer-reviewed studies were retained to match transformer-based sentiment analysis objectives.

Table 3 Inclusion and Exclusion Criteria Used in the Study Selection Process

Aspect	Inclusion Criteria	Exclusion Criteria
Research Focus	Studies applying ML, DL, or transformer-based models for sentiment analysis	Studies addressing topics unrelated to sentiment analysis
Methodology	Research with implemented models and experimental validation	Studies lacking experiments or performance evaluation
Evaluation Metrics	Articles reporting quantitative metrics such as accuracy, precision, recall, or F1-score	Articles without measurable evaluation results
Publication Type	Peer-reviewed journal or conference papers	Editorials, blogs, opinion pieces, or non-peer-reviewed sources
Language	Published in English	Published in languages other than English
Publication Period	Published between 2021 and 2026	Published outside the defined time range
Originality	Unique and original studies	Duplicate records or repeated publications
Technical Depth	Studies with sufficient technical and methodological detail	Purely theoretical discussions or surveys without methodology or implementation [8], [9]

3. Literature Review on Sentiment Analysis Models

This section reviews the sentiment analysis techniques, starting from with traditional approaches, machine learning, deep learning, and transformer-based models. The strengths, limitations, and research trends critically discussed to provide the foundation for the comparative analysis presented later in this review. Figure 5 illustrate sentiment analysis techniques from traditional approaches to machine learning, deep learning, and transformer-based models. It classifies the major AI techniques and representative models highlighting their progression over time. The figure also demonstrates the transition toward advanced transformer architectures that have become the state-of-the-art for contextual sentiment understanding.

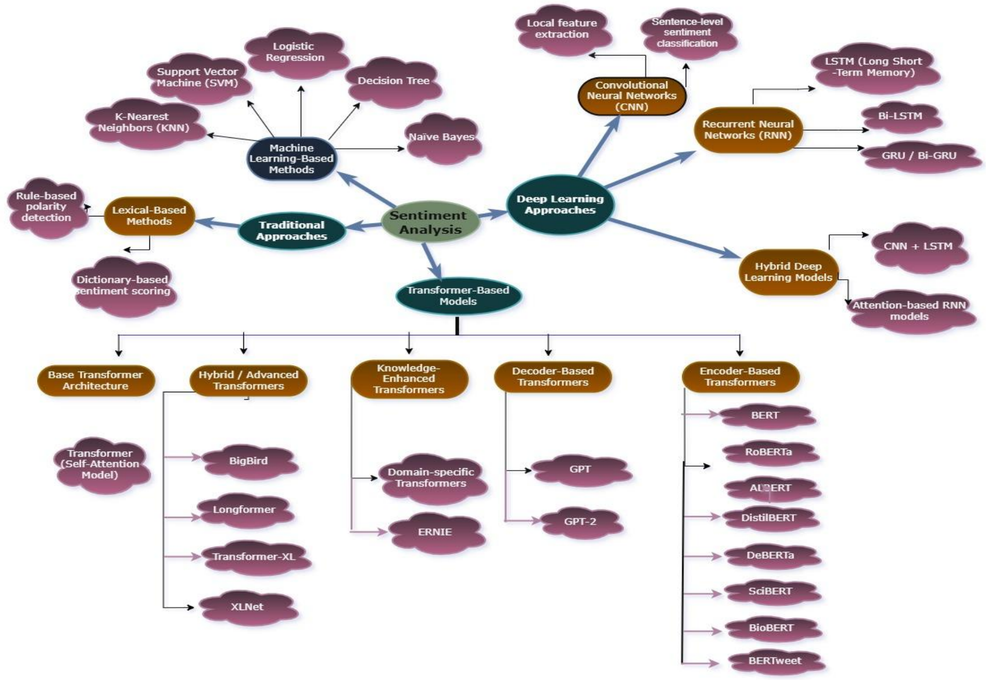

Figure 5 Structure of models related to sentiment analysis

3.1 Traditional Approaches: Lexicon-Based Methods

Sentiment analysis using lexicons is based on pre-existing dictionaries of words that are classified as positive, negative, or neutral. In dictionary-based scoring, the sentiment of a text is determined using a process of matching words in text to the sentiment lexicon and adding their scores together. This is further enhanced by rule-based polarity detection by using linguistic rules to modify sentiment values, e.g., negations and intensifiers. The methods are not based on training data and are not hard to implement. Nevertheless, they usually find it hard to contextualize, be sarcastic, and use domain-specific words and expressions.

3.2 Machine Learning Models

Traditional machine learning based sentiment classification utilizes preprocessed text and it relies on cleaned text and manually designed features, making it easy to use and requiring less inexpensive computational power. Widely used algorithms include Logistic Regression (LR), Random Forest (RF), Naïve Bayes (NB), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN), which have been broadly applied to sentiment classification tasks because of their efficiency in high-dimensional text processing and binary classification. Though models usually provide satisfactory performance on structured and short-text datasets, they depend only on manually and display limited capability in capturing sarcasm, long-range dependencies, contextual understanding and complex linguistic structures. Accordingly, their performance often reduces when applied to large-scale and dynamic social media data. [4, 6, 13, 14].

3.3 Deep Learning Techniques

Deep learning (DL) methods learn to find complex language structures automatically extract hierarchical and abstract information in data. Convolutional Neural Networks (CNNs) learn convolutional features from text using convolutional filters, and work well with short sentiment texts [14]. Recurrent Neural Networks (RNNs) used to process text one word at a time, which enables to capture contextual data with regard to past words, but they experience vanishing gradient issues when dealing with long sequences [7]. Long Short-Term Memory (LSTM) networks overcome this problem since they use memory cells and gating to model the sentiment flow in long documents more effectively [15]. GRUs represent simplified versions of LSTMs, but use fewer parameters and achieve similar performance and can be trained similarly and more quickly [16].

3.4 Literature review on Transformer

Transformer-based models become the foundation of modern sentiment analysis due to their ability to understand context effectively. Unlike earlier neural models, transformers use self-attention mechanisms to capture relationships between words regardless of their position in a sentence. Models such as BERT, RoBERTa, and GPT etc. are pre-trained on large text corpora and then fine-tuned for sentiment classification tasks. These models achieve strong performance on social media data but require high computational resources and large training datasets. Transformer models discuss in Table 5.

Researchers have managed to come up with a number of transformer-based approaches to study how people behave, feel, and engage with stuff in the digital world as it changes [14]. It is important to point out that digitization, especially through social media, has changed the way businesses and customers deal with each other. More and more individuals, especially Generation Z and Alpha, use the internet every day. Businesses can use social media to talk to certain groups of people through text, emoticons, photographs, and videos. This can help them enhance their marketing and build customer loyalty. Transformer-based models may also help firms figure out how customers feel about their product's comments by discovering sentiment in it.[22]. The DynaPR model fills a gap in current recommendation systems by taking into account how product categories affect users' changing interests. DynaPR outperforms other models by employing hierarchical LSTM networks to combine product and attribute information into one space and then uses pairwise interest features to find patterns in behavior and provide reliable recommendations. [23]. RoBERTa was better than Chat-GPT in classifying sentiment and showed important factors that affect accuracy, which helped businesses get better insights. [24]. Think about sarcasm as a multi-modal thing and utilize better encoders to assess the MUStARD++ dataset through text, speech, and pictures. These models are good at find out how people feel in code-mixed languages. Recent studies in 2026 additional extend transformer-based sentiment analysis by focusing on aspect-level opinion mining and effective lightweight architectures.

Table 5 Literature review: Methods, Datasets, and Findings in Sentiment Analysis Research

Study / Research Work	Method / Model Used	Dataset / Application	Key Findings / Contribution	Identified Research Gaps / Limitations	Reference
AI-based models to study online behavior	General AI-based sentiment & behavior study	online platforms, specially social media	Demonstrated how digitization reshapes business–customer interactions, particularly among Gen Z and Alpha users	Lacks fine-grained sentiment modeling and transformer-based analysis	[14]
Dynamic Product Recommendation (DynaPR)	Hierarchical LSTM	E-commerce recommendation systems	Improved personalization by jointly modeling products and attributes	Limited exploration of transformer architectures and explainability	[38]

Sentiment Analysis on Amazon Reviews (Samsung A53)	RoBERTa transformer model	Amazon customer reviews	Achieved higher sentiment classification accuracy compared to ChatGPT	Focused on a single product category; multimodal signals not considered	[16]
Sarcasm Detection (MUStARD++)	Multimodal encoders (text, speech, images)	MUStARD++ dataset	Enhanced sarcasm detection performance using multimodal fusion	High computational cost; limited generalization beyond scripted dialogues	[18]
Multilingual Sentiment Analysis of Slurs	Transformer-based encoders with ML classifiers	Code-mixed social media text data	Effectively handled multilingual and code-mixed sentiment patterns	Limited dataset size and lack of real-time deployment analysis	[15]
Aspect-Based Sentiment Analysis for Product Review Mining (2026)	Transformer-based models (BERT/RoBERTa)	Product review datasets	Improved fine-grained sentiment analysis by identifying sentiments at aspect level	Requires large labeled datasets; challenges in handling implicit aspects	[70]
Lightweight Transformer Models (2026)	DistilBERT / Efficient Transformers	Real-time sentiment systems	Reduced model size and faster inference with comparable accuracy	Slight drop in accuracy compared to large models	[71]

4. Classification Performance Metrics

To measure the performance of the reviewed transformer-based models, a number of performance measures of classification that are commonly accepted are taken into consideration. These metrics give a quantitative measure of the model performance and allow making an unbiased comparison of the performance of various transformer architectures, including BERT, RoBERTa, XLNet, because sentiment analysis activities may be characterized by the imbalance of the classes and light contextual differences. It is necessary to use several assessment scales to evaluate the performance in a comprehensive manner. The most recurring measures in the studies is evaluated by the accuracy, precision, recall, F1-score and Root Mean Square Error (RMSE). These measures are calculated based on the usual elements of a confusion matrix, which are true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

Table 6: Evaluation Metrics for Sentiment Analysis

Metric	Formula (Word-friendly format)	Description
Accuracy	(TP + TN) / (TP + FP + TN + FN)	Measures the proportion of correctly classified instances among all instances. It provides an overall

		performance indicator but may be misleading for imbalanced datasets [33].
Precision	TP / (TP + FP)	Indicates how many of the predicted positive instances are actually positive. It reflects the reliability of positive sentiment predictions [15].
Recall	TP / (TP + FN)	Measures how many actual positive things are correctly identified by the model. It is important for detecting critical sentiments. [15].
F1-Score	2 × (Precision × Recall) / (Precision + Recall)	Harmonic mean of precision and recall. It balances both metrics and used for evaluating transformer models on imbalanced sentiment datasets [18].
Root Mean Square Error (RMSE)	$\sqrt{[(1/n) \times \Sigma (yi - \hat{y}i)^2]}$	Measures the average deviation between predicted sentiment probabilities. It is mostly used when probabilistic outputs are produced by transformer models [42].

5. Analysis of Selected Studies Based on the Research Questions

This section offers a complete discussion based on the research questions identified. These questions aims to enhance understanding of detection methods, and feature extraction techniques, highlight the evaluation metrics by identifying gaps, and propose directions for future work.

RQ1: What are the types of transformer-based models?

5.1. Base Transformer Architecture: The base Transformer architecture is a neural network model designed for sequence processing tasks. It uses a self-attention mechanism to understand the relationship between words in a sentence. Unlike RNNs, it processes all words in parallel, which makes training faster. This design helps the model learn long-range dependencies efficiently. The Transformer architecture serves as the foundation for many modern NLP models. Transformer (2017) model added relationship between the words within a sentence irrespective of distance through the self-attention mechanism. The model efficiently acquires non-recurring long-range dependencies. Most of the current NLP architectures are based on it because of its efficiency and performance [31].

5.2 Encoder-based transformers:

This makes use of the only encoder portion of the transformer architecture. They are primarily intended to learn text as opposed to produce text. These models process the whole sentence simultaneously with the help of bi-directional self-attention. BERT, RoBERTa and ALBERT are popular models. BERT is a transformer-based encoder and is created to acquire deep contextual representations of text. It takes into account left and right contexts simultaneously based on two-way self-attention. BERT assist in the transfer learning by fine-tuning over task-oriented datasets. It enhances sentence level as well as document level interpretation. The architecture captures fine semantic and syntactic patterns. BERT has been able to perform well in sentiment analysis, classification, and question answering [33]. RoBERTa is an adaptation of BERT that is optimized and has better pretraining strategies. It eliminates the next-sentence prediction goal to concentrate on the masked language modeling. It is always able to achieve better performance in sentiment analysis and text classification tasks than BERT [36]. ALBERT is a light transformer model that aims at minimizing the use of memory. ALBERT has rival accuracy, even though it has a smaller number of parameters. It enhances effective training and does not compromise performance. ALBERT does well with a number of NLP benchmark datasets [37]. DistilBERT is a knowledge extracted version of BERT that is compressed. This leads to a reduction in speed and memory usage. DistilBERT preserves almost 95 percent of the BERT performance. It can be used in real time and resource controlled applications. Sentiment analysis systems utilize wide usage of DistilBERT [53]. Improved semantic understanding is achieved with improved attention mechanisms. Better masking techniques are also applied in the training of DeBERTa. It has greater accuracy on various NLP standards. DeBERTa is especially effective in sentiment classification [66]. SciBERT is a domain-specific transformer that is trained on scientific papers. It makes use of corpora of computer science and biomedical research articles. The model can comprehend technical terms and these belongs to the domain. SciBERT improves the citation analysis and text classification. It records complicated patterns of scientific language. SciBERT has many academic NLP applications [63]. BioBERT has been trained on big biomedical literature. The model enhances the knowledge of medical terms. BioBERT is an improvement in the tasks of named entity recognition and relation extraction. The design enhances text analysis in biomedical studies. BioBERT can be applied in sentiment analysis in healthcare [64]. BERTweet is twitter trained. It deals with casual language, slangs, emojis, and hashtags. A contextual representation of the short texts is enhanced by pretraining on tweets. It works on noisy and unstructured data. BERTweet enhances accuracy of sentiment detection in social sites. It finds application in sentiment analysis of social media [62].

5.3. Decoder-Based Transformers: Transformers that have decoder-only architecture are called decoder-based transformers. Their processing is left-to-right that is, each word relies on the preceding words. Mainly text generating tasks are performed with these models. They make predictions based on the past context on the next word in a sequence. Transformers based on decoders are pretrained with huge text corpora. Fine-tuning enables adaptation of them to other tasks. Well-known ones are GPT and GPT-2. GPT is an autoregressive transformer model based on decoding. It is used to predict the word that is coming out following prior context. GPT-2 is an extension of GPT that uses a bigger architecture and dataset. It has a powerful language modeling. Both models facilitate the transfer learning between tasks. Text generation and text analysis The GPT models have found a lot of application [39].

5.4. Hybrid / Advanced Transformers: The advanced or hybrid transformers are designed to eliminate the drawbacks of simple transformer models. They enhance the management of long line of text and complicated dependencies. Such models make use of changed attention mechanisms or memory methods. Other models are a mixture of concepts of other transformer architectures. They lower the cost of computation at the expense of performance. The hybrid transformers are good in long documents. They are XLNet, Transformer-XL, Longformer and BIGBird. XLNet is a bi-directional transformer that is an autoregressive transformer. It does not conceal tokens in training. The model incorporates the permutation-based language modelling. XLNet combines characteristics of Transformer-XL and BERT. XLNet is also faster than a number of transformer models on NLP benchmarks [35]. Transformer-XL is an extension of the existing transformer architecture. It adds the capability of segment-level recurrence to model the long sequences. Positional awareness is enhanced by relative positional encoding. The model helps to keep context in several segments. Transformer-XL enhances the work of language modeling. It can be used in the analysis of long documents [34]. Longformer is an engine that is created to efficiently handle long documents. The model does not lose the valuable contextual information. Longformer approves local and international coverage. It manages thousands of tokens in effect. Longformer can be used in document-level sentiment analysis [60]. BigBird Extension of transformers with sparse attention patterns is provided by BigBird. It is mixture of global, local, and random attention. This saves on memory usage by long sequences. It has a high degree of scaling to large text inputs. [61].

5.5. Knowledge-Enhanced Transformers: Transformers that have been enhanced with knowledge mix textual information and external knowledge. They apply in the form of entities, concepts and relationships in order to enhance understanding. These models do not just learn through plain text but use structured knowledge in the process of learning. This assists the model to have a better understanding of real world meaning and context. The transformers which are knowledge enhanced work well with tasks which are knowledge intensive. They enhance meaning expression and reasoning capabilities. An example of this kind that is popular is ERNIE. It augments transformer models using knowledge. It includes entity-level information and semantic information. The model enhances relationships between entities. Knowledge integration can be used to grasp real-life context. ERNIE enhances the learning of representation of complex texts. It is good at tasks that are knowledge-intensive. ERNIE is applicable in higher-level NLP [57].

5.6. Multimodal Transformer Models: Multimodal transformer models are models that are developed to handle more than one type of data simultaneously. They are usually used in conjunction with text and images or videos. The models employ attention processes in learning intermodal relationships. Multimodal transformers combine the visual information with the textual information. This enhances the comprehension of complicated content. They are popular in such applications as sentiment analysis, image captioning, and video understanding. The most famous ones are CLIP, UNITER, ERNIE-ViL and VideoBERT. VideoBERT is an expansion of transformers to text and video. It is a joint model of visual frames and language tokens. Video sequences are considered to be the tokenized inputs. This enhances the learning of multimodal representation. VideoBERT helps in video comprehension. It is used in video captioning and analysis [59]. CLIP acquires joint representations of text and images. It involves contrastive learning on massive datasets. CLIP does not need fine-tuning on tasks. It is very universal in its application. The architecture has facilitated zero-shot learning. Multimodal sentiment analysis can make use of CLINE [68]. ERNIE-ViL combines visual, textual and knowledge features. It enhances intermodal reasoning. The complex semantic relationship is reflected on the model. It is good in visual question answering. The model is appropriate to multimodal sentiment analysis [69]. UNITER acquires the integrated image and text representations. It records the interactions between visual and linguistic features. The model involves joint embedding learning. UNITER improves multimodal understanding tasks. It is a good performer on benchmark data [65].

5.7. Encoder–Decoder Transformers: Encoder Decoder transformers are encoder- decoder transformers that utilize encoder and decoder parts of the transformer architecture. The encoder is followed by converting the input text into meaningful representations. Decoder is the one that produces output text on the basis of these representations. The models are mostly applied in text generation and translation. They are able to process both lengthy input and output sequences. Encoder-decoder transformers are applied to facilitate various NLP tasks with the same type of architecture. One of the most famous models of this model is T5. T5 converts all NLP tasks to text-to-text format. Input and output are both in form of text. One architecture is shared by a number of tasks. The model behavior is guided by task specific prompts [67].

Table 7 Overview of Transformer-based Models and their applications

Model	Year	Core Idea / Architecture	Key Strength	Typical Sentiment Analysis Use	Reference
ULMFiT	2018	Fine-tuned language model with gradual unfreezing	Strong transfer learning for small datasets	Review classification, suggestion mining	[30]
Transformer	2017	Self-attention without recurrence	Parallel processing, long-range dependency modeling	Base architecture for all modern sentiment models	[31]
GPT	2018	Autoregressive transformer decoder	Context-aware text generation	Opinion generation, sentiment-aware text synthesis	[39]
GPT-2	2019	Large-scale pretrained transformer	Fluent long-text understanding	Document-level sentiment, QA-based sentiment	[39]
BERT	2018	Bidirectional encoder transformer	Deep contextual understanding	Sentence & document sentiment classification	[33]

RoBERTa	2019	Optimized BERT training strategy	Improved robustness and accuracy	Sarcasm detection, fine-grained sentiment	[36]
XLNet	2019	Permutation-based autoregressive training	Bidirectional context without masking	Long-text and contextual sentiment tasks	[35]
Transformer-XL	2019	Segment-level recurrence with relative positions	Long-term dependency modeling	Review and news sentiment analysis	[34]
ALBERT	2019	Parameter sharing and factorized embeddings	Memory-efficient transformer	Large-scale sentiment analysis with low resources	[37]
DistilBERT	2019	Knowledge-distilled BERT	Faster inference	Real-time sentiment systems	[53]
ERNIE	2019	Knowledge-aware transformer	Entity-level semantic understanding	Aspect-based sentiment analysis	[57]
VideoBERT	2019	Joint video–text transformer	Multimodal representation learning	Video comment sentiment analysis	[59]
Longformer	2020	Sparse attention mechanism	Efficient long document processing	Social media thread sentiment	[60]
BigBird	2020	Sparse + global attention	Scalable long-sequence modeling	Large review corpus sentiment analysis	[61]
BERTweet	2020	Twitter-pretrained BERT	Handles informal language & emojis	Social media sentiment analysis	[62]
SciBERT	2020	Scientific text pretraining	Domain-specific vocabulary understanding	Academic and survey sentiment analysis	[63]
BioBERT	2020	Biomedical domain transformer	Medical language understanding	Healthcare sentiment and opinion mining	[64]
UNITER	2020	Unified image–text representation	Vision–language fusion	Multimodal sentiment analysis	[65]
DeBERTa	2021	Disentangled attention mechanism	Improved contextual modeling	Fine-grained sentiment classification	[66]
T5	2021	Text-to-text unified framework	Task flexibility	Sentiment as text generation	[67]
CLIP	2021	Contrastive image–text learning	Strong cross-modal alignment	Image–text sentiment analysis	[68]

ERNIE-ViL	2021	Knowledge-enhanced multimodal transformer	Rich semantic reasoning	Advanced multimodal sentiment tasks	[69]

RQ2. Why are transformer-based models (like BERT and RoBERTa) preferred over classical ML methods such as SVM or Naïve Bayes?

Transformer-based models are preferred because they can capture the real meaning of words in a sentence by looking at the context. Unlike SVM or Naïve Bayes, which depend on word frequency or hand-crafted features such as TF-IDF, transformers use attention mechanisms that allow them to understand how words relate to each other. This simplifies their ability to deal with long sentences and identify the sentiment despite the complex wording. Overall, it is found that transformer models are far more accurate than classical modes. As indicated in the comparative overview, methods that use CNN, LSTM, BERT, ResNet, and Attention mechanisms enhance the process of extracting the features and understanding the context and making decisions. Its importance is based on the ability to bridge modalities, reduce ambiguity and improve applications like e-commerce review, social media monitoring and human-computer interaction. At the same time, such problems as model complexity, overfitting, and memory demands make the current research gaps visible, which makes MSA effective and an ongoing field of study. Recent research shows that classical machine learning methods like SVM and Naive Bayes are being actively employed as a baseline method because of their small ability to understand context and use manual feature engineering. Transformer-based models such as BERT and RoBERTa, on the other hand, use self-attention and large scale pre-training to scope deep semantic and contextual information, leading to higher performance on a variety of multilingual, sarcasm-aware and multimodal, sentiment analysis tasks.

Table 8 Comparative Features of traditional Machine Learning and Transformer-Based Models

Aspect	Classical ML Methods (SVM, NB, LR, KNN)	Transformer-Based Models (BERT, RoBERTa)	Key References
Feature Representation	Requires manual feature engineering such as TF-IDF, Bag-of-Words, and n-grams	Learns dense contextual embeddings automatically	[13]
Context Understanding	Treats words independently; limited contextual awareness	Uses self-attention to capture full bidirectional context	[17]
Semantic Meaning of Words	Static word meaning (same vector regardless of context)	Dynamic word meaning depending on sentence context	[18]
Long-Range Dependency Handling	Weak performance for long or complex sentences	Strong modeling of long-range dependencies via attention	[15]
Sarcasm & Irony Detection	Poor handling of sarcasm and implicit emotions	Significantly improved performance on sarcasm and irony	[24]
Aspect-Based Sentiment Analysis	Limited capability without complex feature design	Effectively captures aspect-level sentiment	[25]
Multilingual & Code-Mixed Text	Performs poorly on multilingual or mixed-language data	Strong multilingual and cross-lingual performance	[11]
Multimodal Sentiment Analysis	Not suitable for multimodal fusion	Easily extended to text–image–audio fusion	[10]

Training Strategy	Trained from scratch on task-specific datasets	Pre-trained on large corpora, then fine-tuned	[28]
Data Efficiency	Requires large labeled datasets	Performs well with limited labeled data	[52]
Domain Adaptability	Requires retraining for new domains	Efficient domain adaptation via fine-tuning	[29]
Scalability & Real-World Use	Limited scalability for large noisy datasets	Highly scalable for real-world applications	[40]
Current Research Role (2023–2025)	Mostly baseline or comparison models	State-of-the-art models in sentiment research	[47]

Table 8 contrasts the classical machine learning models with transformer-based models on sentiment analysis. Though tradtional methods are based on manual feature engineering and do not find out much contextual knowledge, transformer models like BERT and RoBERTa can find self-attention and pre-training to learn more deep semantic connections.

RQ3. What are the main challenges face when using transformers for sentiment analysis in the real world?

Transformer-based models, including BERT and RoBERTa, are high-accuracy sentiment analysis models, there are a number of challenges, which researchers experience when using them in the real world. These problems can be mostly connected with the quality of data, the cost of computation, linguistic diversity, interpretability and the ability to withstand complex emotional conditions. The text data in the real world is also very multi-lingual, noisy as well as sarcasm or mixed feelings which confuses the task of sentiment detection even with sophisticated transformer models.

Table 9 Key Challenges of Transformer-Based Sentiment Analysis

Challenge	Explanation	Supporting References
High Computational Cost	Transformer models require powerful GPUs, large memory, and long training time.	[17], [40]
Large Data Requirement	Even though transformers are pre-trained, fine-tuning still needs labeled domain-specific data, which is costly and time-consuming to collect.	[28], [29], [40]
Lack of Explainability	Transformer decisions are hard to interpret, making it difficult to understand why a particular sentiment was identified.	[54], [55], [27]
Sarcasm and Irony Handling	Transformers struggle with sarcasm, humor, and implicit emotions, especially in social media data.	[24], [27], [41]
Multilingual and Code-Mixed Text	Real-world data often mixes multiple languages, slang, idiom phrases and informal writing, which decreases model accuracy.	[11], [18]
Domain Dependency	A transformer trained on one domain may achieve poorly in another domain (e.g., healthcare or finance).	[19], [29], [47]
Bias in Pre-trained Models	Pre-trained models may hold social or cultural bias inherited from training data, affecting sentiment predictions.	[26], [27], [40]
Handling Noisy Real-World Data	Misspellings, abbreviations, emojis and incomplete sentences reduce sentiment detection accuracy.	[9], [26], [58]
Deployment in Real-Time Systems	Large model size make real-time sentiment monitoring challenging.	[40], [45], [51]

Even with their good performance, there are some practical challenges experienced in the development of transformer -based sentiment analysis models in real world. They have high computational demands and low explainability, which restrictions on their usage in resource-constrained settings. Also, it is hard to cope with slang, sarcasm, multilingual content, and domain-specific variances. Such challenges show that transformers are the state of the art, but more studies should be showed to enhance efficiency, conflict, and readability to real-world sentiment analysis systems.

RQ4. What are the future research guidelines for transformer-based sentiment analysis?

BERT and RoBERTa are transformer-based models that have better the performance of sentiment analysis to a large scope. However, according to recent research, there are still a number of limitations that have not been addressed. Future studies are possible to be on efficiency, multilinguality, multimodality, explainability, robustness, and real-world application. By resolving these problems, it will be possible to make high-performing models more accurate, reliable, and scalable sentiment analysis systems.

Table 10 Future Research Directions in Transformer-Based Sentiment Analysis

Future Direction	Description	Supporting References
Efficient and Lightweight Transformers	Developing smaller and faster models to reduce cost.	[47]
Explainable Sentiment Analysis (XAI)	Improving model transparency hence users can know why a sentiment decision is made.	[55]
Sarcasm and Implicit Emotion Detection	Improving transformer models to better detect slang, sarcasm, irony, and hidden emotions in social media text data.	[24]
Multilingual and Code-Mixed Sentiment Analysis	Designing models that handle code mixed-language content and low-resource languages more accurately.	[54]
Multimodal Sentiment Analysis	Combining text with images, audio, and video to improve sentiment understanding in real-world applications.	[10]
Domain-Adaptive Transformers	Improving generalization across areas such as finance, healthcare, and education without retraining from scratch.	[19]
Bias and Fairness Mitigation	Reducing cultural, social, and gender bias inherited from pre-trained language models.	[27]
Low-Resource and Noisy Data Handling	Improving robustness against emojis, slang, spelling errors, and limited labeled data.	[26]
Integration with Large Language Models (LLMs)	Exploring hybrid models combining transformers with generative LLMs for deeper sentiment.	[45]
Real-Time and Edge Deployment	Adapting transformer models for deployment on IoT, mobile, and edge devices.	[40]
Aspect-Based Sentiment Analysis	Focus on fine-grained opinion mining.	[70]
Lightweight Transformers		[71]

6. Multi-Dimensional Evaluation of Sentiment Analysis Approaches

While predictive performance measures such as recall, accuracy, precision, and F1-score are commonly used to compute sentiment analysis models, these metrics only do not deliver a complete assessment. The efficiency of a sentiment analysis method as well depends on its capability to handle challenging inputs, domains, understand contextual information and operate efficiently at scale. Thus, this study covers the conventional performance-oriented comparison by considering several evaluation dimensions.

The comparative opinion find sentiment analysis methods based on predictive performance, contextual understanding, scalability, robustness, domain adaptability, computational cost, and ethical risks. Robustness refers to the capability of a method to continue reliable performance when dealing with slang, sarcasm, irony, and noisy text, variations across datasets or ambiguous language. Domain adaptability returns the capability of a model to transfer knowledge across different application fields, while the bias and ethical dimension reflects the possibility of unfair predictions, privacy concerns, demographic bias, and limited explainability.

Table 11 illustrate multi-dimensional perspective offers a broader basis for comparing sentiment analysis methods. Rather than identifying a commonly superior method, this perspective helps determine the suitability of different approaches based on the characteristics and requirements of a specific sentiment analysis task.

Table 11 Multi-Dimensional Comparison of Sentiment Analysis Approaches

Approach	**Predictive Performance**	**Context Understanding**	**Robustness**	**Computational Cost**	**Scalability**	**Domain Adaptability**	**Bias/Ethical Considerations**
Rule-Based	Low–Medium	Low	Low	Low	High	Low	Medium
Machine Learning	Medium–High	Medium	Medium	Medium	High	Medium	Medium
Deep Learning	High	High	Medium–High	High	Medium–High	High	Medium–High
Transformer-Based	High–Very High	Very High	High	Very High	Medium–High	High	High

7. Results and Discussion

The selection process was done with the help of PRISMA, and there were 55 peer-reviewed studies published in the years between 2021 and 2026 that were included in the analysis process. The findings reveal that there is a definite trend of abandoning the classic machine learning and recurrent neural networks in favor of transformer-based sentiment analysis networks. This trend is indicative of the high capability of transformers to generalize text data along a context and long-range dependencies. The most commonly used models were those based on BERT and its variations (RoBERTa, ALBERT, DeBERTa, and XLNet) and then there were those based on LSTM- and GRU-based deep learning models. Despite the good performance of the transformer models, a number of studies shown some difficulties during the interpretability, computational cost, and imbalanced data. Finally, the transformer-based models are the existing state of art of sentiment analysis, and the future studies need to focus on efficient, explainable, and multimodal frameworks of transformers to work with real-world contexts. Recent 2026 studies show a change to aspect-based sentiment analysis and lightweight transformer architectures, highlighting the need for more efficient and fine-grained sentiment understanding in real-world applications.

8. Conclusions

This systematic literature review uses the PRISMA-based selection process ensured a systematic and transparent identification, screening, and inclusion of related studies published between 2021 and 2026. By predefined inclusion and exclusion criteria, the review provides a structured synthesis of recent 55 selected papers on transformer-based sentiment analysis. The outcomes show that models such as BERT and RoBERTa provides better contextual understanding and performance than traditional models. Overall, transformer models have suggestively improved sentiment analysis and are used across different application areas. Future research should focus on domain adaptation, bias reduction, explainable AI, multilingual sentiment analysis, multimodal approaches, and real-time deployment.

Overall, this review provides an updated synthesis of transformer-based sentiment analysis study through including current developments up to 2026. By classifying major application areas, methodological trends and emerging research directions. This study offers a consolidated reference for researchers finding to develop more efficient, interpretable and accurate applicable to sentiment analysis systems.

Acknowledgements: We would like to honestly thank all reviewers for kind and positive suggestions.

Conflict of Interest: The authors declare that they have no known competing financial interests that could have seemed to influence the work reported in this review paper.

Funding Information: No funding was received for this study. The authors did not receive financial support from any organization for the authorship, research, or publication of this manuscript.

Ethical Statement: This review paper and does not include any studies with human participants performed by the authors. Therefore, ethical approval consent were not required.

Author's Contributions: *Nisha* - Conceptualization, Methodology, Data Curation, Experimentation and Writing Original Draft Preparation and Editing.

Bal Kishan - Supervision, Validation and Review Draft

REFERENCES

1. Biswas, R., & De, S. (2022, November). A comparative study on improving word embeddings beyond word2vec and glove. In *2022 Seventh International Conference on Parallel, Distributed and Grid Computing (PDGC)* (pp. 113-118). IEEE.
2. Kholwal, R. (2023). Text-Classify: A comprehensive comparative study of logistic regression, random forest, and knn models for enhanced text classification performance. *International Journal of Advances in Engineering & Technology*, *16*(5), 415-433.
3. Ghilzai, S. A. (2025). A Corpus-Based Sentiment Analysis of AI-Generated Deepfakes in Political Discourse. *The Critical Review of Social Sciences Studies*, *3*(2), 2989-3003.
4. W. Bourequat and H. Mourad, "Sentiment Analysis Approach for Analyzing iPhone Release using Support Vector Machine," *Int. J. Adv. Data Inf. Syst.*, vol. 2, no. 1, pp. 36–44, 2021, doi: 10.25008/ijadis.v2i1.1216.
5. S. Kumar, N. Kaur, Kavita, and A. Joshi, "Tweet Sentiment Analysis using Logistic Regression," *IET Conf. Proc.*, vol. 2023, no. 11, pp. 332–336, 2023, doi: 10.1049/icp.2023.1801.
6. A. R. Isnain, J. Supriyanto, and M. P. Kharisma, "Implementation of K-Nearest Neighbor (K-NN) Algorithm For Public Sentiment Analysis of Online Learning," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 15, no. 2, p. 121, 2021, doi: 10.22146/ijccs.65176.
7. S. C. Nistor, M. Moca, D. Moldovan, D. B. Oprean, and R. L. Nistor, "Building a Twitter sentiment analysis system with recurrent neural networks," *Sensors*, vol. 21, no. 7, pp. 1–24, 2021, doi: 10.3390/s21072266.
8. I. Jackson, D. Ivanov, A. Dolgui, and J. Namdar, "Generative artificial intelligence in supply chain and operations management: a capability-based framework for analysis and implementation," *Int. J. Prod. Res.*, vol. 62, no. 17, pp. 6120–6145, 2024, doi: 10.1080/00207543.2024.2309309.
9. V. Buhas, I. Ponomarenko, O. Kazak, and N. Korshun, "AI-Driven Sentiment Analysis in Social Media Content," *CEUR Workshop Proc.*, vol. 3665, pp. 12–21, 2024.
10. L. Christ *et al.*, "The MuSe 2023 Multimodal Sentiment Analysis Challenge: Mimicked Emotions, Cross-Cultural Humour, and Personalisation," *MuSe 2023 - Proc. 4th Multimodal Sentim. Anal. Chall. Work. Mimicked Emot. Humour Pers. Co-located with MM 2023*, pp. 1–10, 2023, doi: 10.1145/3606039.3613114.
11. T. Ranjan, A. Singh, R. Kumari, S. Swain, A. Bandyopadhyay, and A. K. Parida, "Multilingual Code-Mixed Sentiment Analysis in Hate Speech," *Scalable Comput.*, vol. 24, no. 4, pp. 873–882, 2023, doi: 10.12694/scpe.v24i4.2375.
12. Q. Xin, Y. He, Y. Pan, Y. Wang, and S. Du, "The implementation of an AI-driven advertising push system based on a NLP algorithm," *Int. J. Comput. Sci. Inf. Technol.*, vol. 1, no. 1, pp. 30–37, 2023, doi: 10.62051/ijcsit.v1n1.05.
13. C. Dewi, R. C. Chen, H. J. Christanto, and F. Cauteruccio, "Multinomial Naïve Bayes Classifier for Sentiment Analysis of Internet Movie Database," *Vietnam J. Comput. Sci.*, vol. 10, no. 4, pp. 485–498, 2023, doi: 10.1142/S2196888823500100.

14. U. D. Gandhi, P. Malarvizhi Kumar, G. Chandra Babu, and G. Karthick, "Sentiment Analysis on Twitter Data by Using Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM)," *Wirel. Pers. Commun.*, pp. 1–11, 2021, doi: 10.1007/s11277-021-08580-3.
15. J. Shobana and M. Murali, "An efficient sentiment analysis methodology based on long short-term memory networks," *Complex Intell. Syst.*, vol. 7, no. 5, pp. 2485–2501, 2021, doi: 10.1007/s40747-021-00436-4.
16. W. Xu, J. Chen, Z. Ding, and J. Wang, "Text sentiment analysis and classification based on bidirectional Gated Recurrent Units (GRUs) model," *Appl. Comput. Eng.*, vol. 77, no. 1, pp. 132–137, 2024, doi: 10.54254/2755-2721/77/20240670.
17. M. S. Sayeed, V. Mohan, and K. S. Muthu, "BERT: A Review of Applications in Sentiment Analysis," *HighTech Innov. J.*, vol. 4, no. 2, pp. 453–462, 2023, doi: 10.28991/HIJ-2023-04-02-015.
18. O. Zalutska *et al.*, "Method for Sentiment Analysis of Ukrainian-Language Reviews in E-Commerce Using RoBERTa Neural Network," *CEUR Workshop Proc.*, vol. 3387, pp. 344–356, 2023.
19. J. Zhou and J. min Ye, "Sentiment analysis in education research: a review of journal publications," *Interact. Learn. Environ.*, vol. 31, no. 3, pp. 1252–1264, 2023, doi: 10.1080/10494820.2020.1826985.
20. H. Vartiainen and M. Tedre, "Using artificial intelligence in craft education: crafting with text-to-image generative models," *Digit. Creat.*, vol. 34, no. 1, pp. 1–21, 2023, doi: 10.1080/14626268.2023.2174557.
21. M. H. Abdalla *et al.*, "Sentiment Analysis Based on Hybrid Neural Network Techniques Using Binary Coordinate Ascent Algorithm," *IEEE Access*, vol. 11, no. November, pp. 134087–134099, 2023, doi: 10.1109/ACCESS.2023.3334980.
22. X. Yang and J. A. Esquivel, "Time-Aware LSTM Neural Networks for Dynamic Personalized Recommendation on Business Intelligence," *Tsinghua Sci. Technol.*, vol. 29, no. 1, pp. 185–196, 2024, doi: 10.26599/TST.2023.9010025.
23. Dewi, C., Chen, R. C., Christanto, H. J., & Cauteruccio, F. (2023). Multinomial Naïve Bayes classifier for sentiment analysis of internet movie database. *Vietnam Journal of Computer Science*, *10*(04), 485-498.
24. S. Bhosale *et al.*, "Sarcasm in Sight and Sound: Benchmarking and Expansion to Improve Multimodal Sarcasm Detection," pp. 1–14, 2023.
25. N. Song, H. Cai, R. Xia, J. Yu, Z. Wu, and X. Dai, "A Sequence-to-Structure Approach to Document-level Targeted Sentiment Analysis," *Find. Assoc. Comput. Linguist. EMNLP 2023*, pp. 7687–7698, 2023, doi: 10.18653/v1/2023.findings-emnlp.515.
26. M. S. Md Suhaimin, M. H. Ahmad Hijazi, E. G. Moung, P. N. E. Nohuddin, S. Chua, and F. Coenen, "Social media sentiment analysis and opinion mining in public security: Taxonomy, trend analysis, issues and future directions," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 35, no. 9, p. 101776, 2023, doi: 10.1016/j.jksuci.2023.101776.
27. A. D. L. Langure and M. Zareei, "Breaking Barriers in Sentiment Analysis and Text Emotion Detection: Toward a Unified Assessment Framework," *IEEE Access*, vol. 11, no. October, pp. 125698–125715, 2023, doi: 10.1109/ACCESS.2023.3331323.
28. M. Sharma, D. Patel, A. Bose, and R. Patel, "Leveraging BERT and Sentiment-LSTM for Enhanced AI-Driven Marketing Sentiment Analysis," *Int. J. AI ML Innov.*, vol. 11, no. 8, pp. 1–23, 2022.
29. M. K. Shaik Vadla, M. A. Suresh, and V. K. Viswanathan, "Enhancing Product Design through AI-Driven Sentiment Analysis of Amazon Reviews Using BERT," *Algorithms*, vol. 17, no. 2, 2024, doi: 10.3390/a17020059.
30. Howard, J., & Ruder, S. (2018). Universal language model fine-tuning for text classification. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL), 328–339. https://doi.org/10.18653/v1/P18-1031
31. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998–6008.
32. Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder–decoder for statistical machine translation. Proceedings of EMNLP, 1724–1734.
33. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of NAACL-HLT, 4171–4186. https://doi.org/10.18653/v1/N19-1423
34. Dai, Z., Yang, Z., Yang, Y., Carbonell, J., Le, Q. V., & Salakhutdinov, R. (2019). Transformer-XL: Attentive language models beyond a fixed-length context. Proceedings of ACL, 2978–2988.
35. Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R., & Le, Q. V. (2019). XLNet: Generalized autoregressive pretraining for language understanding. Advances in Neural Information Processing Systems, 32, 5753–5763.
36. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692.
37. Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2020). ALBERT: A lite BERT for self-supervised learning of language representations. International Conference on Learning Representations (ICLR).
38. F. Merlin, D. V. M, L. Divyabharathi, S. N. Murali, and T. V Ai-, "Information Technology , Research and Applications (IJITRA) AI-Driven Sentiment Classification via Combinatorial Techniques and Reasoning," 2024, doi: 10.59461/ijitra.v3i4.121.
39. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. OpenAI Technical Report.
40. P. S. Ghatora, S. E. Hosseini, S. Pervez, M. J. Iqbal, and N. Shaukat, "Sentiment Analysis of Product Reviews Using Machine Learning and Pre-Trained LLM," *Big Data Cogn. Comput.*, vol. 8, no. 12, 2024, doi: 10.3390/bdcc8120199.

41. P. D. Manuel, "Emotional and Sarcastic Sentiment Analytics - An Extreme AI Model," *AIBC 2024 - 2024 5th Int. Artif. Intell. Blockchain Conf.*, pp. 31–38, 2025, doi: 10.1145/3702359.3702364.
42. L. Nemes and A. Kiss, "Prediction of stock values changes using sentiment analysis of stock news headlines," *J. Inf. Telecommun.*, vol. 5, no. 3, pp. 375–394, 2021, doi: 10.1080/24751839.2021.1874252.
43. Saidi, R., & Jarray, F. (2026). Arabic word sense disambiguation: a survey in the era of transformer-based models. *International Journal of Data Science and Analytics*, *22*(1), 10.
44. E. S. Alamoudi and N. S. Alghamdi, "Sentiment classification and aspect-based sentiment analysis on yelp reviews using deep learning and word embeddings," *J. Decis. Syst.*, vol. 30, no. 2–3, pp. 259–281, 2021, doi: 10.1080/12460125.2020.1864106.
45. S. S. Jamee *et al.*, "Enhancing Automated Trading with Sentiment Analysis : Leveraging Large Language Models for Stock Market Predictions," vol. 07, no. 185, pp. 185–195, 2025, doi: 10.37547/tajet/Volume07Issue03-16.
46. T. Hulugh and C. O. Ikeh, "Artificial Intelligence in product management : Automating roadmap prioritization through sentiment analysis and customer feature demand modeling Artificial Intelligence in product management : Automating roadmap prioritization through sentiment analysis ," no. April, 2025, doi: 10.30574/ijsra.2025.15.1.1246.
47. J. W. Kang and S. Y. Choi, "Comparative Investigation of GPT and FinBERT's Sentiment Analysis Performance in News Across Different Sectors," *Electron.*, vol. 14, no. 6, pp. 1–26, 2025, doi: 10.3390/electronics14061090.
48. D. K. Nasiopoulos, K. I. Roumeliotis, D. P. Sakas, K. Toudas, and P. Reklitis, "Financial Sentiment Analysis and Classification : A Comparative Study of Fine-Tuned Deep Learning Models," pp. 1–27, 2025.
49. D. Smith-Mutegi, Y. Mamo, J. Kim, H. Crompton, and M. McConnell, "Perceptions of STEM education and artificial intelligence: a Twitter (X) sentiment analysis," *Int. J. STEM Educ.*, vol. 12, no. 1, 2025, doi: 10.1186/s40594-025-00527-5.
50. M. Alzahrani and F. AlGhamdi, "Social Media Sentiment Analysis for Sustainable Rural Event Planning: A Case Study of Agricultural Festivals in Al-Baha, Saudi Arabia," *Sustain.*, vol. 17, no. 9, 2025, doi: 10.3390/su17093864.
51. K. Kirtac and G. Germano, "Enhanced Financial Sentiment Analysis and Trading Strategy Development Using Large Language Models," *Proc. 14th Work. Comput. Approaches to Subj. Sentim. Soc. Media Anal.*, pp. 1–10, 2024.
52. K. I. Roumeliotis, N. D. Tselikas, and D. K. Nasiopoulos, "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," *Inf.*, vol. 15, no. 12, pp. 1–23, 2024, doi: 10.3390/info15120792.
53. Şentürk, A., Albayrak, A., & Arpacı, S. (2026). Multi-Class News Classification with BERT, DistilBERT, RoBERTa, and ELECTRA Natural Language Processing Models. *Duzce University Journal of Science and Technology*, *14*(1), 117-129.
54. R. Mabokela, M. Primus, and T. Celik, "Explainable Pre-Trained Language Models for Sentiment Analysis in Low-Resourced Languages," 2024, doi: 10.20944/preprints202409.0253.v1.
55. I. A. E. Madi, A. Redjdal, J. Bouaud, and B. Seroussi, "Exploring Explainable AI Techniques for Text Classification in Healthcare: A Scoping Review," *Stud. Health Technol. Inform.*, vol. 316, pp. 846–850, 2024, doi: 10.3233/SHTI240544.
56. Borg, A.; Boldt, M. Using VADER sentiment and SVM for predicting customer response sentiment. Expert Syst. Appl. 2020, 162, 113746.
57. Sun, Y., Wang, S., Li, Y., Feng, S., Tian, H., Wu, H., & Wang, H. (2019). ERNIE: Enhanced representation through knowledge integration. Proceedings of ACL, 1441–1451.
58. L. Nemes and A. Kiss, "Social media sentiment analysis based on COVID-19," *J. Inf. Telecommun.*, vol. 5, no. 1, pp. 1–15, 2021, doi: 10.1080/24751839.2020.1790793.
59. Sun, C., Myers, A., Vondrick, C., Murphy, K., & Schmid, C. (2019). VideoBERT: A joint model for video and language representation learning. Proceedings of ICCV, 7464–7473.
60. Wang, X., Liu, W., Yang, Y., Gao, Y., Du, Q., & Bao, L. (2026). Docsentinet: a adaptive architecture for efficient document-level sentiment analysis. International Journal of Data Science and Analytics, 21(1), 34.
61. Zaheer, M., Guruganesh, G., Dubey, K. A., Ainslie, J., Alberti, C., Ontañón, S., Pham, P., Ravula, A., Wang, Q., Yang, L., & Ahmed, A. (2020). *BigBird: Transformers for longer sequences*. Advances in Neural Information Processing Systems, 33, 17283–17297.
62. Nguyen, D. Q., Vu, T., & Tuan Nguyen, A. (2020). *BERTweet: A pre-trained language model for English tweets*. Proceedings of EMNLP, 9–14.
63. *Beltagy, I., Lo, K., & Cohan, A. (2019). SciBERT: A pretrained language model for scientific text. Proceedings of EMNLP-IJCNLP, 3615–3620.*
64. *Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., & Kang, J. (2020). BioBERT: A pre-trained biomedical language representation model for biomedical text mining. Bioinformatics, 36(4), 1234–1240. https://doi.org/10.1093/bioinformatics/btz682*
65. *Chen, Y.-C., Li, L., Yu, L., El Kholy, A., Ahmed, F., Gan, Z., Cheng, Y., & Liu, J. (2020). UNITER: Universal image–text representation learning. Proceedings of ECCV, 104–120.*
66. *He, P., Liu, X., Gao, J., & Chen, W. (2021). DeBERTa: Decoding-enhanced BERT with disentangled attention. International Conference on Learning Representations (ICLR).*
67. *Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. Journal of Machine Learning Research, 21(140), 1–67.*

68. *Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. Proceedings of ICML, 8748–8763.*
69. *Yu, J., Li, J., Yu, Z., Huang, Q., & Tian, Q. (2021). ERNIE-ViL: Knowledge enhanced vision-language representations through scene graph. Proceedings of CVPR, 12362–12372.*
70. Astri, R., Kamal, A., Sari, I. U., Anggraini, P., & Budiman, A. (2026). Aspect-Based Sentiment Analysis for Product Review Mining with Transformer-Based Models. *Journal of Electrical Engineering and Computer Science (JEECS)| E-ISSN: 3089-5952*, *2*(1), 37-43.
71. *Wu, D., Nie, L., Tohye, T. G., & Tao, J. (2025). AutoLTS: Lightweight Transformer-Based Simulation Model for Efficient Communication Using Hybrid Large Language Models. IEEE Transactions on Cognitive Communications and Networking, 12, 4376-4386.*