

# A Comprehensive Review of Opinion Mining: Methodological Innovations, Applications, and Emerging Challenges

Arshad Hussain<sup>1</sup>, Dr. Garima Tyagi<sup>2</sup>

<sup>1</sup>Research Scholar, Career Point University, Kota

<sup>2</sup>Professor, Career Point University, Kota

**Abstract**—Opinion mining, commonly referred to as sentiment analysis, has gained significant importance with the rapid growth of user-generated content across social media platforms, e-commerce websites, and online discussion forums. By automatically extracting opinions, emotions, and attitudes from textual data, opinion mining supports informed decision-making in domains such as business intelligence, governance, healthcare, and education. Recent advances in artificial intelligence, particularly deep learning, natural language processing (NLP), and large language models (LLMs), have substantially improved the accuracy and adaptability of sentiment analysis systems.

Despite these advancements, several challenges remain unresolved. Contemporary models often struggle with sarcasm, contextual ambiguity, domain adaptation, and multilingual sentiment understanding. Additionally, the opaque nature of many deep learning and LLM-based approaches raises concerns regarding interpretability, fairness, and reliability, especially in high-stakes applications. Traditional sentiment models are also limited in capturing fine-grained and aspect-level opinions beyond coarse polarity classification.

This review presents a comprehensive analysis of recent methodological innovations in opinion mining, including deep neural networks, hybrid architectures, graph-based models, prompt learning strategies, and large language models such as ChatGPT. The paper systematically examines their applications across diverse domains, including social media analytics, healthcare informatics, education, and software engineering. Furthermore, it identifies critical research gaps and emerging challenges related to explainable AI, low-resource and multilingual settings, and ethical deployment. By offering a structured taxonomy and comparative insights, this review aims to guide future research toward more robust, transparent, and scalable opinion mining systems.

## I. INTRODUCTION

In the contemporary digital ecosystem, individuals actively express their thoughts, opinions, and emotions across a wide range of online platforms, including social media networks, e-commerce portals, blogs, and review forums. This massive volume of user-generated textual data has emerged as a valuable source for understanding public sentiment, behavioral patterns, and decision-making trends. Opinion mining, also referred to as sentiment analysis, focuses on automatically identifying, extracting, and classifying subjective information from textual content, typically categorizing opinions into positive, negative, or neutral sentiments.

Opinion mining plays a crucial role across multiple application domains. Businesses leverage sentiment analysis to assess customer satisfaction and improve products and services, while governments and public organizations use it to analyze citizen feedback and policy perception. In healthcare, sentiment analysis assists in monitoring patient experiences and public health awareness, whereas in education it supports feedback analysis and learning experience evaluation. The evolution of opinion mining techniques has been marked by significant methodological shifts. Early approaches relied on rule-based systems and lexicon-driven polarity detection. These methods were later replaced by machine learning models using handcrafted features and classifiers such as Support Vector Machines and Naïve Bayes. More recently, deep learning architectures including recurrent neural networks, convolutional neural networks, and transformer-based models have demonstrated superior performance by capturing contextual and semantic

relationships within text. The emergence of large language models (LLMs), such as GPT-based systems, has further transformed sentiment analysis by enabling zero-shot and few-shot learning, multilingual processing, and flexible adaptation across domains.

Despite substantial progress, opinion mining continues to face critical challenges. Accurately detecting sarcasm, irony, and context-dependent sentiment remains difficult, particularly in informal and short-text scenarios. Many languages and application domains suffer from limited labeled data, restricting the generalizability of trained models. Additionally, the black-box nature of deep learning and LLM-based systems raises concerns regarding transparency, interpretability, and ethical reliability—especially when deployed in sensitive areas such as healthcare, governance, and public opinion monitoring.

Unlike earlier review studies that primarily focus on traditional machine learning models or transformer-based architectures, this review emphasizes recent advancements that integrate large language models, prompt learning strategies, graph neural networks, and explainable AI into opinion mining frameworks. The paper provides a structured taxonomy of modern techniques, compares their strengths and limitations across multiple application domains, and highlights emerging research challenges aligned with post-2024 trends. By bridging methodological innovation with real-world deployment considerations, this review offers a forward-looking perspective on building robust, transparent, and scalable opinion mining systems.

#### Scope

This review paper aims to provide a comprehensive overview of recent advances in opinion mining (sentiment analysis), with a focus on both technological innovations and real-world applications. The paper surveys the evolution of opinion mining methods from traditional machine learning approaches to modern deep learning, large language models (LLMs), and graph-based frameworks. It also explores the growing use of prompt learning, transfer learning, and multilingual models, which have expanded the capability of opinion mining across various platforms and languages.

The scope of this review includes the analysis of sub-areas within opinion mining, such as:

- Aspect-Based Sentiment Analysis (ABSA)

- Contextual and Sarcasm Detection
- Explainable and Interpretable Sentiment Systems
- Multilingual and Low-Resource Language Processing
- Applications across domains like Social Media, Healthcare, Education, and Software Engineering

By covering these areas, the paper addresses both technical dimensions (e.g., model architectures, training strategies, data handling) and application domains, showcasing how sentiment analysis is used in practice to derive insights from user-generated content.

#### Objective of Study

1. To survey and classify recent opinion mining techniques, including hybrid models, LLMs, and GNNs.
2. To compare strengths and limitations of different approaches using a structured framework.
3. To highlight major application domains, particularly in social media, healthcare, education, and software development.
4. To identify key research gaps such as sarcasm detection, data imbalance, interpretability, and cross-lingual challenges.
5. To suggest future directions for building more robust, ethical, and scalable opinion mining systems.

## II. REVIEW OF LITERATURE

Recent advances in opinion mining reflect a clear transition from conventional deep learning models toward large-scale, context-aware, and explainable frameworks. Heo et al. (2025) investigated the capability of large language models (LLMs), particularly GPT-4, for multi-turn and context-sensitive sentiment classification. Their findings demonstrated that LLMs can effectively perform zero-shot and few-shot sentiment analysis when guided by well-designed prompts, reducing dependency on large labeled datasets.

Hybrid neural architectures continue to play an important role in sentiment classification tasks. Hidri and Khalfi (2025) proposed a hybrid Bidirectional Gated Recurrent Unit (Bi-GRU) and Long Short-Term Memory (LSTM) model that captures both sequential and long-range contextual dependencies in unstructured textual data.

Similarly, Qutab et al. (2024) introduced a BiLSTM-RNN-based framework tailored for sentiment analysis in software engineering contexts, demonstrating improved performance in handling domain-specific technical vocabulary from developer discussions and repositories.

With the growing adoption of conversational and informal content platforms, recent studies have explored the application of LLMs in real-world opinion mining tasks. Wang et al. (2024) conducted a case study using ChatGPT to analyze sentiments from YouTube captions, highlighting the effectiveness of generative language models in processing short, conversational, and noisy text. These findings emphasize the growing relevance of prompt-driven sentiment analysis in social media environments.

Aspect-Based Sentiment Analysis (ABSA) has received significant attention due to its ability to capture fine-grained opinions. Li et al. (2025) proposed an aspect-aware opinion decomposition approach that improves interpretability by separating sentiment expressions into domain-specific components. In a comprehensive survey, Hua et al. (2024) systematically reviewed ABSA methods and concluded that transformer-based architectures dominate current research, although challenges remain in domain adaptation and interpretability.

Several studies have also explored alternative modeling paradigms. Pattern-based sentiment analysis approaches, such as those presented in AIP Conference Proceedings (2024), demonstrate efficiency in structured datasets but exhibit limited robustness in handling sarcasm and contextual ambiguity. Optimization-driven methods, including the Enhanced Grey Wolf Optimization (EGWO) approach proposed by Gupta (2024), have shown promising results in aspect extraction tasks, particularly in e-commerce review analysis.

Beyond text-only sentiment analysis, multimodal opinion mining has emerged as a growing research direction. Link et al. (2024) reviewed multimodal sentiment analysis frameworks that integrate textual, audio, and visual signals, demonstrating substantial performance improvements in video-centric platforms. Additionally, hybrid data mining and fuzzy logic-based methods explored by ACM BDEIM (2023) highlight the effectiveness of combining rule-based reasoning with machine learning for real-time social media sentiment tracking.

Overall, the literature indicates a shift toward hybrid, explainable, and domain-adaptive sentiment analysis systems. While LLMs and deep neural networks achieve state-of-the-art performance, issues related to transparency, fairness, and generalizability across languages and domains remain open research challenges.

Recent research in opinion mining has increasingly focused on addressing limitations related to interpretability, adaptability, and ethical reliability of sentiment analysis systems. While deep learning and large language models (LLMs) have significantly improved performance, their deployment in real-world applications necessitates additional considerations beyond accuracy alone.

#### (A) Explainable Sentiment Analysis

One of the major challenges associated with modern sentiment analysis models is their lack of interpretability. Deep neural networks and LLM-based sentiment classifiers often operate as black-box systems, making it difficult to understand the reasoning behind their predictions. This limitation is particularly critical in sensitive domains such as healthcare, public policy analysis, and decision support systems, where transparency and trust are essential.

Recent studies have therefore emphasized the integration of explainable artificial intelligence (XAI) techniques into sentiment analysis frameworks. Zhang et al. (2024) provide a comprehensive survey of explainable sentiment analysis methods, highlighting approaches such as attention visualization, feature attribution, and rule-based explanation mechanisms. Their findings suggest that explainability not only improves user trust but also helps in identifying model bias and improving robustness, making XAI a key research direction in contemporary opinion mining.

#### (B) Instruction-Tuned and Prompt-Based Large Language Models

The emergence of large language models has introduced a paradigm shift in opinion mining, moving away from traditional supervised learning toward instruction-tuned and prompt-based learning approaches. Unlike conventional models that require extensive labeled datasets, prompt-driven LLMs can perform sentiment analysis in zero-shot and few-shot

settings, enabling rapid adaptation across domains and languages.

Zhao et al. (2024) demonstrated that instruction-tuned LLMs achieve competitive sentiment classification performance across multiple domains by leveraging carefully designed prompts. These models reduce annotation costs and exhibit strong generalization capabilities, particularly in low-resource environments. The growing adoption of prompt learning strategies indicates a transition toward more flexible and scalable sentiment analysis systems, positioning prompt-based LLMs as a central component of future opinion mining research.

### (C) Bias, Fairness, and Ethical Challenges in Opinion Mining

As opinion mining systems become increasingly embedded in societal decision-making processes, concerns related to bias, fairness, and ethical responsibility have gained significant attention. Sentiment analysis models, especially those trained on large-scale web data, may inadvertently reflect or amplify societal biases, leading to unfair or misleading outcomes.

Bender et al. (2023) critically examined the ethical risks associated with large language models, emphasizing issues such as data bias, overgeneralization, and unintended harmful behavior. Their work underscores the importance of responsible model design, transparent evaluation, and ethical deployment practices in sentiment analysis. Incorporating fairness-aware learning strategies and bias mitigation techniques is therefore essential to ensure that opinion mining systems remain reliable, inclusive, and socially responsible.

## III. RESEARCH METHODOLOGY

This review adopts a qualitative and structured literature analysis approach to examine recent developments in opinion mining and sentiment analysis. The primary objective of the methodology is to ensure comprehensive coverage of state-of-the-art techniques, emerging research trends, and practical applications reported in recent scholarly work.

### Data Sources

Relevant studies published between 2023 and 2025 were collected from well-established and reputable

digital libraries and indexing platforms, including IEEE Xplore, ACM Digital Library, SpringerLink, Elsevier ScienceDirect, MDPI, arXiv, and the Journal of Big Data. These sources were selected to ensure coverage of both peer-reviewed research and high-quality preprints reflecting the latest advancements in the field.

### Search Strategy

A systematic keyword-based search strategy was employed using combinations of terms such as “opinion mining,” “sentiment analysis,” “aspect-based sentiment analysis,” “large language models,” “prompt learning,” “explainable AI,” “multilingual NLP,” and “graph neural networks.” Boolean operators and advanced filtering options were applied to refine results based on relevance, publication year, and research focus.

### Inclusion and Exclusion Criteria

To maintain quality and relevance, studies were selected based on the following criteria:

#### Inclusion Criteria

- Articles published between 2023 and 2025
- Studies proposing, evaluating, or reviewing sentiment analysis techniques
- Research addressing theoretical advances or real-world applications of opinion mining
- Publications written in English

#### Exclusion Criteria

- Non-peer-reviewed articles with insufficient technical depth
- Studies not directly related to opinion mining or sentiment analysis
- Duplicate publications or papers with overlapping datasets lacking novel insights

### Analysis and Classification Framework

The selected studies were systematically analyzed and categorized based on multiple dimensions to facilitate comparative evaluation. These dimensions include:

- Technique Type: Deep learning models, hybrid architectures, large language models (LLMs), graph neural networks (GNNs), and optimization-based approaches
- Sentiment Scope: Binary, multi-class, and aspect-based sentiment analysis

- Application Domain: Social media analytics, healthcare, education, software engineering, and e-commerce
- Research Challenges Addressed: Sarcasm detection, multilingual and low-resource settings, interpretability, bias, and data imbalance

This structured classification framework enables a coherent comparison of methodologies and highlights gaps in existing research. While the review does not claim full systematic review compliance, the adopted methodology ensures transparency, reproducibility, and balanced coverage of contemporary opinion mining research.

IV. COMPARATIVE ANALYSIS TABLE

Technique	Strengths	Limitations	Explainability	Scalability
Bi-GRU + LSTM Hybrid Models	Effectively capture sequential and contextual dependencies	High training time and memory usage	Low	Moderate
GPT-4 / ChatGPT (LLMs)	Strong zero-shot and few-shot performance; rich contextual understanding	Sensitive to prompt design; limited transparency	Low–Moderate	High
Prompt Learning (e.g., PL-FGSA)	Reduced dependency on labeled data; fast domain adaptation	Limited handling of nuanced or implicit sentiment	Moderate	High
Graph Neural Networks (GNNs)	Structure-aware sentiment analysis; strong for ABSA	Complex training; dependency parsing overhead	Low	Moderate
Multilingual Models (XLM-R, mBERT)	Enable cross-lingual sentiment analysis	Performance drops for dialects and informal language	Low	High
Pattern-Based Approaches	Simple, efficient, and interpretable	Weak performance on unstructured or sarcastic text	High	High
Hybrid Optimization Models (EGWO)	High precision in aspect extraction	Domain-specific; limited generalization	Moderate	Moderate
Multimodal Sentiment Systems	Improved accuracy by integrating text, audio, and visual cues	Requires high-quality multimodal datasets	Low	Low–Moderate

The comparative analysis highlights that no single opinion mining technique provides a complete solution across all evaluation dimensions. While large language models and deep neural architectures demonstrate superior performance and scalability, they suffer from limited interpretability. In contrast, pattern-based and optimization-driven approaches offer greater transparency but lack robustness in complex linguistic scenarios. These trade-offs indicate a growing research trend toward hybrid and explainable sentiment analysis frameworks that balance accuracy, scalability, and transparency.

V. DISCUSSION AND RESULTS

The comprehensive analysis of recent literature indicates that opinion mining has undergone a significant transformation with the integration of advanced deep learning techniques and large language models (LLMs). Compared to traditional machine learning and lexicon-based approaches, modern architectures demonstrate superior performance in sentiment polarity classification, aspect-based sentiment analysis, and multilingual opinion extraction. These improvements are primarily

attributed to enhanced contextual modeling capabilities, attention mechanisms, and semantic representation learning, which allow models to capture nuanced sentiment expressions in unstructured and noisy textual data.

Large language models such as GPT-based systems represent a major paradigm shift in opinion mining research. Their ability to perform zero-shot and few-shot sentiment analysis enables effective sentiment classification even in scenarios with limited labeled data. This adaptability has proven particularly valuable for cross-domain and multilingual sentiment analysis tasks. However, despite their strong performance, LLM-based sentiment systems are highly sensitive to prompt design and input formulation. Variations in prompt structure can lead to inconsistent sentiment predictions, highlighting the need for standardized evaluation protocols and robust prompt engineering strategies.

While performance gains have been substantial, the widespread adoption of deep neural networks and LLMs introduces critical challenges related to interpretability and transparency. Most state-of-the-art sentiment analysis models operate as black-box systems, making it difficult to explain how specific predictions are generated. This lack of interpretability poses serious concerns in high-stakes application areas such as healthcare analytics, policy evaluation, and educational assessment, where understanding the reasoning behind sentiment predictions is essential for trust and accountability. Consequently, recent research has increasingly emphasized the integration of explainable artificial intelligence (XAI) techniques into opinion mining frameworks to balance accuracy with transparency.

Another prominent trend observed in the reviewed studies is the growing use of prompt learning and transfer learning strategies. These approaches significantly reduce the dependence on large annotated datasets and improve model generalization across domains and languages. Prompt-based learning has shown particular promise in low-resource and multilingual settings, enabling sentiment analysis systems to be deployed in underrepresented languages and cultural contexts. Nevertheless, challenges such as prompt sensitivity, response variability, and limited control over generative outputs remain open research problems that require further investigation.

Graph-based sentiment analysis methods, including Graph Neural Networks (GNNs), have also gained attention due to their ability to model syntactic and semantic relationships between words and aspects. These approaches are especially effective for aspect-based sentiment analysis, where capturing dependency structures and contextual relationships is critical. However, the computational complexity of graph-based models and their reliance on accurate dependency parsing limit their scalability in real-time and large-scale applications.

From an application perspective, opinion mining has expanded well beyond traditional domains such as product reviews and social media analytics. Recent studies demonstrate its increasing relevance in healthcare, education, and software engineering. In healthcare, sentiment analysis supports patient feedback interpretation and public health monitoring; in education, it facilitates the analysis of learner feedback and engagement; and in software engineering, it aids in understanding developer sentiment and software quality assessment. Despite these advancements, domain-specific challenges—such as technical jargon, contextual ambiguity, and data sparsity—continue to hinder model generalizability, underscoring the need for hybrid and domain-adaptive sentiment analysis solutions.

Overall, the findings suggest that no single opinion mining technique offers a complete solution across all performance dimensions. While LLMs and deep learning models provide high accuracy and scalability, they suffer from limited interpretability and ethical concerns. In contrast, rule-based and optimization-driven approaches offer greater transparency but lack robustness in complex linguistic scenarios. These trade-offs indicate a growing research shift toward hybrid, explainable, and ethically responsible sentiment analysis frameworks that balance performance, transparency, and real-world applicability.

## VI. CONCLUSION

Opinion mining has emerged as a vital research area for extracting meaningful insights from the rapidly growing volume of user-generated textual content across digital platforms. Over the past decade, the field has evolved from rule-based and traditional machine learning techniques to sophisticated deep learning

architectures and large language models (LLMs), enabling more accurate, context-aware, and scalable sentiment analysis. These advancements have significantly enhanced the capability of sentiment analysis systems to process complex linguistic structures, informal expressions, and multilingual data.

This review presented a comprehensive examination of recent methodological innovations in opinion mining, including hybrid deep learning models, graph-based frameworks, prompt learning strategies, and large language models. The comparative analysis highlighted the strengths and limitations of existing approaches and demonstrated that no single technique provides an optimal solution across all dimensions. While LLMs and deep neural networks offer superior performance and adaptability, challenges related to interpretability, bias, and ethical reliability remain critical barriers to their widespread adoption in sensitive application domains.

The findings of this review emphasize the importance of balancing performance with transparency and trustworthiness. As opinion mining systems are increasingly deployed in areas such as healthcare analytics, education, governance, and software engineering, ensuring explainability, fairness, and accountability becomes essential. Hybrid and explainable sentiment analysis frameworks are therefore likely to play a central role in the next generation of opinion mining research.

## VII. FUTURE SCOPE

Despite significant progress, several open research challenges and opportunities remain in the field of opinion mining. Addressing these challenges will be crucial for developing robust, inclusive, and ethically responsible sentiment analysis systems.

**Sarcasm and Contextual Understanding:** Future research should focus on improving the detection of sarcasm, irony, and context-dependent sentiment by incorporating contextual cues, discourse-level analysis, and external knowledge sources.

**Multilingual and Low-Resource Language Processing:** Expanding sentiment analysis capabilities to support underrepresented languages and dialects remains a major challenge. Leveraging multilingual

embeddings, cross-lingual transfer learning, and prompt-based LLMs offers promising directions for reducing language barriers.

**Explainable and Ethical Opinion Mining:** As sentiment analysis systems increasingly influence decision-making processes, integrating explainable AI techniques and fairness-aware learning strategies is essential. Future models must ensure transparency, bias mitigation, and accountability, particularly in high-stakes domains.

**Real-Time and Stream-Based Sentiment Analysis:** The growing demand for real-time opinion monitoring across social media streams, live events, and news platforms highlights the need for scalable and efficient sentiment analysis architectures capable of handling high-velocity data.

**Multimodal Sentiment Analysis:** Incorporating multiple modalities such as text, audio, and visual information can significantly enhance sentiment detection accuracy. Future research should focus on developing unified multimodal frameworks and standardized evaluation benchmarks.

**Standardized Evaluation and Benchmarking:** The lack of consistent benchmarks and evaluation protocols for modern LLM-based sentiment analysis systems remains an open issue. Establishing standardized datasets and metrics will be crucial for fair comparison and reproducibility of results.

By addressing these directions, future opinion mining research can move toward more intelligent, transparent, and socially responsible sentiment analysis systems that effectively support real-world applications.

## REFERENCES

- [1] S. Heo, J. Park, E. Kim, and J. Lee, "Can large language models be effective online opinion miners?" arXiv preprint, arXiv:2505.15695, 2025.
- [2] L. Hidri and B. Khalfi, "Opinion mining and analysis using hybrid deep neural networks," *Technologies*, vol. 13, no. 5, Art. no. 175, 2025.
- [3] ACM MSR, "A four-dimension gold standard dataset for opinion mining in software

- engineering,” in Proc. 21st Int. Conf. Mining Software Repositories (MSR), 2024.
- [4] J. Wang, K. Sun, and J. Luo, “Opinion mining from YouTube captions using ChatGPT: A case study,” arXiv preprint, arXiv:2304.03434, 2024.
- [5] D. Li, H. Wang, Y. Zhang, and T. Liu, “Aspect-aware decomposition for opinion summarization,” arXiv preprint, arXiv:2501.17191, 2025.
- [6] X. Hua, X. Chen, W. Zhu, and Z. Li, “A systematic review of aspect-based sentiment analysis: Domains, methods, and trends,” *Artificial Intelligence Review*, vol. 57, no. 3, pp. 1157–1184, 2024.
- [7] AIP Conference Proceedings, “An efficient opinion mining using polarity with a pattern-based approach,” *AIP Conf. Proc.*, vol. 3180, Art. no. 030001, 2024.
- [8] R. Gupta, “A novel approach for aspect-based opinion mining using enhanced grey wolf optimization,” *Int. J. Intelligent Systems and Applications in Engineering*, vol. 12, no. 1, pp. 1–9, 2024.
- [9] ACM BDEIM, “Research on hybrid data mining methods for social media sentiment analysis,” in Proc. Int. Conf. Big Data Engineering and Intelligent Management (BDEIM), 2023.
- [10] M. Link, D. Sharma, and Z. Ahmed, “Multimodal aspect-based sentiment analysis: A survey of datasets, methods, and applications,” *Information Processing & Management*, vol. 61, no. 4, Art. no. 103565, 2024.
- [11] Y. Zhang, S. Liu, and X. Wang, “Explainable sentiment analysis: A survey of methods, evaluation, and applications,” *Artificial Intelligence Review*, vol. 57, no. 6, 2024.
- [12] W. Zhao, J. Zhou, and Y. Li, “Instruction-tuned large language models for sentiment analysis across domains,” in Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP), 2024, pp. 1123–1135.
- [13] E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, “On the risks of stochastic parrots: Can language models be too big?” in Proc. ACM Conf. Fairness, Accountability, and Transparency (FAccT), 2023, pp. 610–62