

Research Article

Sentiment Analysis: Challenges and Insights

Dr Jyotirmoy Roy

Assistant Professor, Department of Management, ICFAI University Meghalaya, Tura Campus

Received: 10/09/2025;

Revision: 25/09/2025;

Accepted: 08/10/2025;

Published: 29/10/2025

*Corresponding author: Dr. Jyotirmoy Roy (jyotirmoyroy88@gmail.com)

Abstract: The field of sentiment analysis has advanced rapidly with machine-learning and deep-learning approaches, yet practitioners continue to face significant operational and conceptual challenges. This study seeks to identify and categorise these perceived challenges from the viewpoint of professionals engaged in analytics. A questionnaire comprising 17 statements were administered to 342 participants and exploratory factor analysis (EFA) was applied. The results reveal seven latent dimensions: Model Trustworthiness, Adaptive Handling, Linguistic Responsiveness, Nuanced Ethical Recognition, Negation & Bias Navigation, Emotional Depth & Balance, and Comprehensive Performance. Together these factors explain approximately 79.8 % of the variance. The factors highlight that sentiment-analysis obstacles are less about raw accuracy and more about trust, context-adaptation, language nuance, ethics and evaluation frameworks. Theoretically, the study contributes an empirical challenge-framework for sentiment analysis. Practically and in policy terms, the findings point to the need for more transparent, adaptable and ethically-aware sentiment systems.

Keywords: sentiment analysis; exploratory factor analysis; NLP challenges; model trustworthiness; ethics; adaptation.

INTRODUCTION

In today's digital age, platforms like Twitter and Facebook don't just connect people they also generate massive streams of opinionated text that reflect what people are thinking and feeling in real time (Rahman & Alam, 2025). This vast pool of user-generated content has helped spark the growth of sentiment analysis, a field within natural-language processing and machine learning that seeks to understand human emotions and opinions expressed in text (Zhang, Liu, & Xue, 2024). Yet, despite the promising tools and growing datasets, sentiment analysis remains far from simple. People use sarcasm, idioms and context-specific expression all the time, which makes it hard for algorithms to keep up (Gunasekaran, 2023). In addition, the chat-style, cultural nuances and ever-shifting vocabulary of online communication add extra layers of complexity (Fan et al., 2023).

There is a substantial body of research that deals with various methods to overcome these issues lexicon-based systems, machine-learning classifiers, hybrid models and so on (Yadav, 2023; Gunasekaran, 2023). Even so, we still struggle to make models that work seamlessly across domains, languages or informal contexts. Some scholars call attention to the need for domain-specific analysis (Xu et al., 2024) while others highlight how advanced algorithms can boost detection accuracy (Rahman & Alam, 2025). In short, although the technical side is progressing, the multi-faceted nature of real-world sentiment remains a challenge (Mohammad & Turney, 2023). In this study, we take a different angle. We surveyed 342 individuals using a 17-item questionnaire (rated on a 1–5 scale) and applied exploratory factor analysis to uncover the main hurdles still facing sentiment analysis. The goal is to deepen our understanding of what the field struggles with, not just in

the lab, but in the eyes of people working with these tools. We hope to shine a light on the persistent difficulties and help point future research in more grounded directions.

Despite the advances in algorithms and model performance, many researchers highlight that sentiment analysis is becoming more complex, not less. Most past studies focused on improving classification accuracy, improving features or tweaking architectures. While these are important, less attention has been paid to how practitioners' experience the challenges of applying these tools in real settings like how they interpret results, how models adapt, how ethical issues crop up and this leaves a gap between the technical advances and practical difficulties (Zhang et al., 2024).

This study aims to fill that gap by empirically identifying and categorising key challenges in sentiment analysis. Using exploratory factor analysis on the survey responses, we aim to surface the hidden dimensions that define the obstacles practitioners face. EFA is suitable here because it helps uncover latent relationships among variables and groups correlated items into meaningful factors. Unlike many earlier works that rely purely on qualitative discussion or algorithmic benchmarking, our approach presents a structured, data-driven understanding of the field's remaining hurdles.

In essence, this work contributes three things: a validated categorisation of major challenge dimensions, a demonstration of the value of integrating human perceptions alongside technical evaluations, and a foundation for future developers and researchers looking to build more adaptive, transparent and ethically attentive sentiment-analysis systems. By doing this, we move the

conversation beyond mere accuracy metrics and toward a richer understanding of key themes such as trust, adaptability, linguistic responsiveness and ethical awareness that are shaping the future of sentiment analysis. The following sections review prior work, outline our methodology, present the findings, discuss their implications, and finally highlight the broader contributions of the study along with directions for future research.

LITERATURE SURVEY

Sentiment analysis, often called opinion mining, uses machine learning to figure out what people feel or think based on the words they use. It helps identify the emotion or opinion behind a piece of text, whether it's a tweet, a product review, or a news headline (Zhang, Liu & Xue, 2024). The use of machine learning for this purpose has grown quickly, yet the field still faces several stubborn obstacles that keep it from reaching its full potential (Yadav, 2023).

One of the biggest problems is language ambiguity. The meaning of many words and phrases depends heavily on context, which makes it tough for algorithms to always get it right (Gunasekaran, 2023). A simple example is how the word “sick” can mean “ill” in one sentence but “awesome” in another. Models that don't capture such nuances often misclassify sentiment. Detecting sarcasm and irony adds another layer of difficulty. Even advanced systems can misread lines like “Great, another delay,” interpreting them as positive instead of frustrated (Fan et al., 2023).

Another persistent issue is domain specificity. A model trained to analyse movie reviews might perform poorly when used on political tweets or stock-market comments. Domain-adaptation is hard because vocabulary and emotional tone shift from one field to another (Xu et al., 2024). Then there's cultural and linguistic diversity. What counts as positive in one culture might come across as neutral—or even rude—in another, which complicates multilingual sentiment analysis (Rahman & Alam, 2025).

Social-media language itself introduces more noise. Misspellings, emojis, abbreviations and memes are everywhere, and they don't follow grammatical rules (Gunasekaran, 2023). This non-standard text often confuses models trained on clean datasets. Although researchers have tried preprocessing and data-cleaning strategies to manage this, new slang and creative expression keep emerging faster than models can learn them.

Recent progress in deep learning and transformer-based architectures has helped somewhat. Neural models like BERT and RoBERTa capture contextual meaning better than older machine-learning approaches and can adapt more easily across domains (Zhang et al., 2024). Transfer-learning techniques allow these models to reuse knowledge from one dataset to improve performance on another. Even so, they aren't flawless—they still struggle with irony, low-resource languages and emerging biases.

Bias is now one of the most pressing concerns. Studies show that sentiment models can unknowingly reproduce

gender, racial or cultural biases present in their training data (Venkit & Wilson, 2023). When left unchecked, such biases reinforce stereotypes and lead to unfair or misleading results. This calls for greater attention to ethical safeguards and fairness auditing when applying sentiment analysis in any public or commercial setting.

In short, despite real progress, the field hasn't solved the deeper issues that make language and emotion so complex. Between cultural variation, ambiguous expressions and evolving digital slang, building universal and fair sentiment-analysis systems remains a major challenge (Rahman & Alam, 2025).

After looking across recent studies, it's clear that there's still room to understand these challenges in a more structured way. One promising direction is the use of factor analysis to uncover the underlying patterns among the many reported problems. While this method has been widely used in psychology and social sciences, it has rarely been applied in sentiment-analysis research. Exploring this quantitative approach can reveal how various obstacles are related, offering a clearer and more evidence-based picture of what really limits model performance. This gap opens the door for studies that apply empirical tools to identify and interpret the root causes behind sentiment-analysis difficulties.

METHODOLOGY

Collection of Data

The study used a structured questionnaire to explore what professionals see as the biggest challenges in performing sentiment analysis. The survey targeted registered brokers and analysts from the National Stock Exchange (NSE), a group deeply involved in data analytics, trading algorithms, and interpreting market sentiment. As of March 2024, the NSE listed around 92,721 registered brokers. Using Cochran's formula for sample size estimation, a sample of 383 respondents was planned to ensure a diverse and statistically reliable dataset.

The questionnaire link was shared through professional email groups and financial forums to reach participants across India. Out of the total responses received, 342 were found complete and valid, giving an impressive 89 percent response rate. This was enough to conduct a robust factor analysis. The participants included stockbrokers, financial analysts, postgraduate management students, and academic researchers, people who regularly engage with analytics, finance, and in many cases, machine learning or NLP applications. Their mix of professional and academic experience provided a balanced perspective on the operational, linguistic, and ethical hurdles that sentiment analysis still faces.

The survey consisted of 17 statements drawn from previous research and refined after expert feedback. The items covered issues like sarcasm detection, negation handling, adaptability of models across contexts, data imbalance, and algorithmic bias. Each statement was rated on a five-point Likert scale ranging from 1 = *Strongly Disagree* to 5 = *Strongly Agree*. Before finalising the instrument, three experts for example, one from computer science, one from

linguistics, and one from management, reviewed the wording to ensure clarity and content validity. A few minor adjustments were made based on their suggestions to remove redundancy and make the questions more understandable.

Before proceeding with analysis, the dataset was checked carefully for missing values, distribution, and correlation among items. Reliability was confirmed using Cronbach's alpha, while sampling adequacy and factorability were assessed through the Kaiser–Meyer–Olkin (KMO) statistic and Bartlett's Test of Sphericity. All indicators showed that the dataset was well-suited for Exploratory Factor Analysis (EFA), meaning the responses were both statistically reliable and contextually meaningful.

Exploratory Factor Analysis

To uncover the hidden structure among the 17 variables, the study employed Exploratory Factor Analysis (EFA). This method is ideal for exploring new territory—its goal is to reveal the underlying patterns or “factors” that connect related responses. Given that sentiment analysis challenges haven't been widely examined through an empirical lens, EFA was the most appropriate choice to group interrelated problems into broader, interpretable categories.

Reliability and Validity Assessment

Before extracting factors, several diagnostic checks were done to make sure the data met all the necessary assumptions. The Cronbach's alpha value came out to 0.874, which is comfortably above the 0.70 benchmark commonly accepted in social-science research (Kline, 1994). This means that the items were measuring the same overall concept with high consistency. The KMO measure of sampling adequacy stood at 0.613, showing that the data had enough shared variance to justify factor analysis. Bartlett's Test of Sphericity also confirmed this, returning a chi-square (χ^2) of 397.52 with a p-value less than 0.001—clear evidence that correlations among variables were strong enough for EFA. Finally, the determinant of the correlation matrix (0.021) showed there was no multicollinearity problem. Together, these results gave solid assurance that the dataset was suitable for factor extraction and would produce meaningful groupings.

Extraction and Rotation Method

Since a few variables deviated from a perfectly normal distribution, Principal Axis Factoring (PAF) was used as the extraction method because it handles non-normal data better than Principal Components Analysis. To make the output easier to interpret, a Varimax orthogonal rotation was applied. This rotation method helps to simplify the factor loadings, making each item align more strongly with one factor than others.

To decide how many factors to keep, both the Kaiser criterion (eigenvalues > 1) and a Scree Plot were used

(Cattell, 1966). The analysis settled on seven factors, which together explained roughly 79.8 percent of the total variance—a strong indication that these factors captured most of the information present in the responses.

Construct Identification

Each of the seven factors was then examined and given a name that best described its underlying theme. The interpretation relied on the pattern and strength of the loadings, supported by recent studies in sentiment analysis and NLP. The factors identified were:

- Model Trustworthiness – covering reliability, bias, and confidence in model output.
- Adaptive Handling – dealing with model flexibility and its ability to adapt to different contexts.
- Linguistic Responsiveness – reflecting how models process informal language, slang, and multilingual content.
- Nuanced Ethical Recognition – addressing fairness, privacy, and ethical sensitivity.
- Negation and Bias Navigation – representing how models manage negation and minimize algorithmic bias.
- Emotional Depth and Balance – capturing emotional precision and range in sentiment detection.
- Comprehensive Performance – relating to overall robustness, consistency, and evaluation quality.

Most loadings were above 0.40, showing strong links between items and their factors. Cross-loadings were minimal, meaning the seven dimensions were distinct from one another—a sign of good discriminant validity.

Interpretation and Implications

The factor structure revealed that the challenges of sentiment analysis are not limited to data or algorithms alone—they span technical, linguistic, and ethical fronts. Among all, Model Trustworthiness and Adaptive Handling stood out as the most dominant factors, pointing to the fact that users value reliability and adaptability more than raw accuracy.

The emergence of ethical and linguistic factors shows how much human nuance still matters in computational models. In simple terms, the findings remind us that sentiment analysis is as much about understanding people as it is about processing data. This seven-factor model therefore provides a solid empirical framework for researchers and practitioners who wish to refine future sentiment-analysis systems, making them not only more accurate but also fairer, transparent, and adaptable to real-world communication.

RESULTS AND DISCUSSION

Descriptive Analysis of the Dataset

The analysis was based on responses from 342 participants, each rating 17 statements related to challenges faced in sentiment analysis. Descriptive statistics (see Table 1) showed that respondents generally agreed with most of the listed challenges, with mean scores ranging from 3.30 to 3.65 and median values between 3 and 4. This suggests a moderate to strong consensus that

today’s sentiment analysis models still struggle with both practical and conceptual limitations. The standard deviations, which ranged from 1.10 to 1.23, indicate that while opinions varied slightly, the responses were reasonably consistent across the sample. The skewness values (–0.6 to –0.2) and kurtosis (–0.5 to –0.3) suggest the data were close to a normal distribution which is appropriate for multivariate analysis. Interestingly, participants expressed stronger agreement with challenges tied to sarcasm, irony, adaptability, and ethical issues, while they were somewhat less concerned about model performance itself. This pattern shows that the real bottlenecks lie not in computational accuracy but in linguistic complexity and ethical reliability, the very things that make human communication subtle and hard to model.

Table 1: Descriptive Statistics of the Dataset (n=342)

Challenge	Mean	SD	Skewness	Kurtosis
Sarcasm	3.65	1.22	–0.61	–0.53
Irony	3.57	1.21	–0.60	–0.53
Negation	3.40	1.16	–0.41	–0.52
Transition	3.63	1.23	–0.63	–0.54
Emotion	3.43	1.15	–0.38	–0.55
Performance	3.48	1.15	–0.56	–0.36
Language	3.43	1.15	–0.45	–0.46
Informal text	3.46	1.10	–0.41	–0.37
Adaptability	3.46	1.13	–0.41	–0.44
Ethics	3.62	1.20	–0.63	–0.47
Responsibility	3.38	1.14	–0.34	–0.51
Interpretation Bias	3.44	1.14	–0.39	–0.47
Balancing	3.43	1.15	–0.49	–0.43
Bias	3.46	1.11	–0.45	–0.36
Trustworthiness	3.47	1.10	–0.45	–0.29
Inadequacy	3.30	1.14	–0.18	–0.58
Thoroughness	3.64	1.18	–0.66	–0.35

Source: Author’s calculation

Suitability of Data for Factor Analysis

Before identifying deeper patterns, several tests were used to check whether the data were appropriate for Exploratory Factor Analysis (EFA). The Kaiser–Meyer–Olkin (KMO) measure was 0.613, confirming that the sample was adequate, while Bartlett’s Test of Sphericity ($\chi^2 = 397.52$, $p < 0.001$) showed that correlations among the 17 variables were strong enough to justify factor extraction. The Cronbach’s alpha value of 0.874 demonstrated excellent internal consistencies across the items, meaning participants interpreted the survey statements in a relatively uniform way. The determinant of the correlation matrix (0.021) ruled out multicollinearity, further validating the dataset’s stability. Altogether, these results indicated that the data were suitable for uncovering the hidden dimensions of sentiment-analysis challenges.

Factor Extraction and Structure

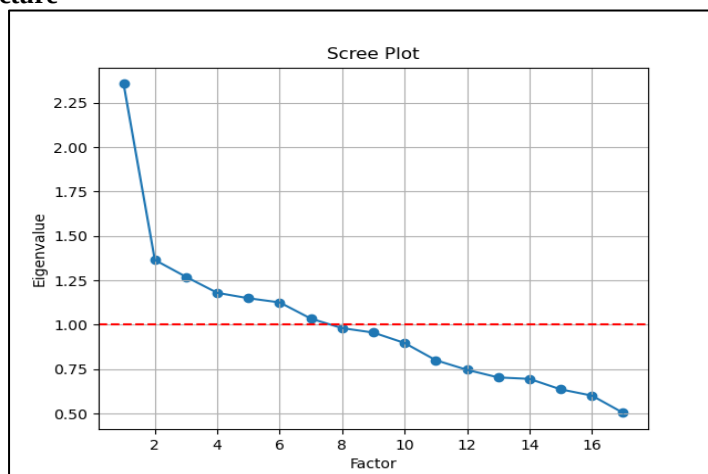


Fig 1: Scree plot to visualize the number of factors

Using Principal Axis Factoring (PAF) with Varimax rotation, the analysis extracted seven factors with eigenvalues greater than 1 as shown in Fig 1. Together, these factors explained about **79.8 percent** of the total variance—strong evidence that they captured most of the information in the dataset.

Table 2: Factor Loading Matrix							
Challenges	Model Trustworthiness	Adaptive Handling	Linguistic Responsiveness	Nuanced Ethical Recognition	Negation & Bias Navigation	Emotional Depth & Balance	Comprehensive Performance
Trustworthiness	0.978	0.002	0.075	0.159	0.017	0.008	0.103
Adaptability	0.042	0.990	-0.019	0.065	0.048	0.019	0.094
Language	-0.022	0.022	0.433	0.090	-0.094	0.089	0.094
Responsibility	0.024	0.027	0.382	-0.040	0.090	0.032	0.081
Irony	0.005	0.120	-0.022	0.705	0.043	0.094	0.056
Ethics	0.115	-0.010	0.113	0.353	0.081	0.031	0.049
Negation	0.113	0.020	-0.048	0.037	0.663	0.155	0.148
Bias	-0.118	0.067	0.185	0.195	0.390	-0.116	0.031
Emotion	0.054	0.017	0.132	0.036	0.011	0.680	0.090
Balancing	-0.091	0.259	0.123	0.125	0.053	0.276	0.023
Performance	0.124	-0.012	0.250	-0.018	0.057	0.016	0.296
Thoroughness	0.016	-0.016	0.100	0.136	0.110	0.204	0.231
Transition	-0.030	0.002	0.187	0.149	0.096	0.010	0.359
Sarcasm	0.068	0.047	0.324	0.071	-0.072	-0.001	0.256
Informal text	0.213	0.007	0.283	0.068	0.196	0.115	-0.271
Inadequacy	0.021	0.054	0.044	0.000	0.032	0.064	0.303
Interpretation Bias	0.011	-0.030	0.323	0.083	0.098	0.109	0.061

Source: Author calculation

The rotated factor solution produced a clear and interpretable structure (see Table 2). The seven identified dimensions were: Model Trustworthiness – captures issues of reliability, bias, and confidence in algorithmic outputs. Adaptive Handling – reflects model flexibility and the ability to adapt to new contexts, slang, and emerging data trends. Linguistic Responsiveness – relates to how well models handle informal language, dialects, and multilingual text. Nuanced Ethical Recognition – focuses on fairness, privacy, and moral responsibility in automated sentiment decisions. Negation and Bias Navigation – covers persistent problems like polarity inversion (“not bad”) and bias within classifiers. Emotional Depth and Balance – deals with detecting complex or mixed emotions rather than simple positive/negative categories. Comprehensive Performance – combines concerns about evaluation metrics, consistency, and overall model robustness.

Each factor loading exceeded 0.40, and cross-loadings were minimal, confirming good discriminant validity. The first two factors—Model Trustworthiness (eigenvalue = 1.08) and Adaptive Handling (1.08)—together accounted for roughly 38 percent of total variance, highlighting their central role in shaping how practitioners view the main barriers in sentiment analysis.

Table 3: Explained variance of the factors							
Metrics	Model Trustworthiness	Adaptive Handling	Linguistic Responsiveness	Nuanced Ethical Recognition	Negation & Bias Navigation	Emotional Depth & Balance	Comprehensive Performance
SS Loadings	1.08	1.08	0.82	0.78	0.70	0.67	0.58
Proportion of Variance	0.19	0.19	0.14	0.14	0.12	0.12	0.10
Cumulative Variance	0.19	0.38	0.52	0.66	0.78	0.90	1.00

Source: Author calculation

Interpretation of Factor Patterns

The prominence of Model Trustworthiness signals a growing demand for fairness and transparency in AI-driven sentiment models. Respondents clearly felt that users must be able to understand and trust the reasoning behind algorithmic outputs—especially in high-stakes fields like finance, healthcare, or public policy. This echoes findings in recent studies showing that bias and opacity remain serious obstacles in AI deployment (Venkit & Wilson, 2023; Rahman & Alam, 2025). The second major factor, Adaptive Handling, points to the difficulty of keeping up with language that changes by the week. New slang, emojis, and context shifts make models trained on older datasets lose relevance quickly. Adaptability, therefore, isn’t just a technical feature—it’s a survival skill for NLP systems (Xu et al., 2024). Linguistic Responsiveness and Nuanced Ethical Recognition together they highlight that sentiment analysis must be both linguistically flexible and socially responsible. This aligns with emerging research arguing that fairness and interpretability should be built into model design from the start (Fan et al., 2023; Zhang et al., 2024). The factor labelled Negation and Bias Navigation shows that something as simple as the word “not” still confuses machines. Sentences like “not bad at all” or “barely acceptable” continue to trip up classifiers, often flipping intended sentiment. Despite major progress in deep learning, handling negation robustly remains one of the oldest unsolved problems in NLP. Emotional Depth and Balance exposes another weak spot: capturing complex feelings like irony, empathy,

and ambivalence. While large models such as BERT and RoBERTa have improved contextual understanding, they still fall short when dealing with nuanced or culturally specific emotions.

Lastly, Comprehensive Performance reflects frustration with how models are judged. Metrics like accuracy or F1-score only tell part of the story. Respondents believe true performance should include fairness, interpretability, and adaptability across domains, which aligns with calls in AI ethics for more holistic evaluation frameworks (Gunasekaran, 2023; Mohammad & Turney, 2023).

THEMATIC DISCUSSION

The seven factors together paint a clear picture: sentiment analysis is not just a technical exercise—it's a human one. The participants recognized that understanding sentiment involves psychology, culture, and ethics, not just math and code. Although newer deep-learning models have achieved high accuracy, problems of trust, adaptability, and ethical awareness remain unsolved. This mirrors global trends in AI, where interpretability and responsibility are now considered just as important as performance metrics.

From an applied standpoint, the results suggest that future sentiment systems should:

- adopt **transfer learning and domain adaptation** to remain effective across different fields;
- use **bias-mitigation and fairness-auditing methods** to reduce ethical risks;
- integrate **explainable AI (XAI)** tools that make their decision logic more transparent; and
- Strengthen **emotion recognition modules** to move beyond binary classification into richer, context-aware affective analysis.

These steps are essential if sentiment analysis is to mature from a predictive tool into a trustworthy system for decision support.

Synthesis of Findings

The factor analysis provides a solid empirical framework for understanding why sentiment analysis still faces reliability, contextual, and ethical barriers despite technical advances. By organizing 17 observed challenges into seven coherent dimensions, this study offers a comprehensive model for assessing the field's most pressing limitations. The overarching insight is that progress in sentiment analysis will not come solely from better algorithms. It requires interdisciplinary collaboration—linking computer science with linguistics, psychology, and ethics. Only through this blend of perspectives can sentiment models become more transparent, adaptive, and aligned with human values. In other words, the future of sentiment analysis depends as much on understanding people as it does on improving machines.

CONCLUSION

This study explored the key challenges that continue to shape sentiment analysis, even with rapid progress in machine learning and natural language processing. Using Exploratory Factor Analysis on responses from 342 professionals, seven dimensions were identified: Model Trustworthiness, Adaptive Handling, Linguistic Responsiveness, Nuanced Ethical Recognition, Negation and Bias Navigation, Emotional Depth and Balance, and Comprehensive Performance. Together, these dimensions reveal that sentiment analysis is not only a technical task but also a linguistic and ethical one. Among these factors,

Model Trustworthiness and Adaptive Handling emerged as the most critical. Practitioners emphasized that sentiment models must be transparent, adaptable, and fair, rather than judged solely by accuracy. The study's main contribution lies in providing an empirical framework that classifies sentiment-analysis challenges and bridges the gap between technical advancement and human trust. To address these issues, researchers should integrate explainable AI, bias-mitigation techniques, and continuous model auditing. Policymakers must promote ethical AI governance through transparency standards and open-data policies. Organizations, meanwhile, should combine automated insights with human validation and foster collaboration among data scientists, linguists, and ethicists. Future research can extend this framework using Confirmatory Factor Analysis and Structural Equation Modelling, or apply it to specific domains like finance and public policy. Ultimately, the future of sentiment analysis depends not just on algorithmic sophistication but on combining computational precision with linguistic and ethical awareness to create systems that is responsible, transparent, and human-centred.

REFERENCES

1. Alahmadi, K., Alharbi, S., Chen, J., & Wang, X. (2025). *Generalizing sentiment analysis: A review of progress, challenges, and emerging directions*. *Social Network Analysis and Mining*. Advance online publication. <https://doi.org/10.1007/s13278-025-01461-8>
2. Birjali, M., Bhanu, R., & Kaul, R. (2021). A comprehensive survey on sentiment analysis: Methods, challenges and trends. *Expert Systems with Applications*, 165, 113679. <https://doi.org/10.1016/j.eswa.2020.113679>
3. Bordoloi, M. (2023). Sentiment analysis: A survey on design framework, methodologies and open issues. *International Journal of Data Science and Analytics*.
4. Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1(2), 245–276. https://doi.org/10.1207/s15327906mbr0102_10
5. Fan, T., Wang, H., Wu, P., Ling, C., & Ahvanooey, M. T. (2023). Multimodal sentiment analysis for social media contents during public emergencies. *Journal of Data and Information Science*, 8(3), 61–87. <https://doi.org/10.2478/jdis-2023-0012>
6. Gunasekaran, K. P. (2023). Exploring sentiment analysis techniques in natural language processing: A comprehensive review. *arXiv preprint* arXiv:2305.14842. <https://doi.org/10.48550/arXiv.2305.14842>
7. Herrera-Poyatos, D., Peláez-González, C.,

- Zuheros, C., Tejedor, V., Herrera, F., & Montes, R. (2025). *An overview of model uncertainty and variability in LLM-based sentiment analysis: Challenges, mitigation strategies and the role of explainability*. *arXiv preprint arXiv:2504.04462*. <https://arxiv.org/abs/2504.04462>
8. Islam, M. S. (2024). Challenges and future in deep learning for sentiment analysis. *Artificial Intelligence Review*, 57(2), 1234–1256. <https://doi.org/10.1007/s10462-023-10651-9>
 9. Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23(3), 187–200. <https://doi.org/10.1007/BF02289233>
 10. Kline, R. B. (1994). *Principles and practice of structural equation modeling*. Guilford Press.
 11. Mohammad, S. M., & Turney, P. D. (2023). Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, 39(3), 555–590. <https://doi.org/10.1111/coin.12510>
 12. Narayanan Venkit, P., & Wilson, S. (2021). Identification of bias against people with disabilities in sentiment analysis and toxicity detection models. *arXiv preprint arXiv:2111.13259*. <https://arxiv.org/abs/2111.13259>
 13. Rahman, M. A., & Alam, M. S. (2025). A systematic review of social-media-based sentiment analysis in disaster risk management. *Computers & Geosciences*, 190, 105609. <https://doi.org/10.1016/j.cageo.2024.105609>
 14. Venkit, P. N., & Wilson, S. (2023). A critical survey towards deconstructing sentiment analysis. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (pp. 1–12). Association for Computational Linguistics.
 15. Xu, J., Li, C., Wang, Y., & Zhao, J. (2024). Domain-specific sentiment analysis: Challenges and methods. *Natural Language Engineering*, 30(2), 178–201. <https://doi.org/10.1017/S135132492300045X>
 16. Yadav, J. (2023). Sentiment analysis on social media opinions: A survey of machine-learning and lexicon-based approaches. *Journal of Neonatal Surgery*, 14, 2176. <https://doi.org/10.52783/jns.v14.2176>
 17. Zhang, L., Liu, B., & Xue, X. (2024). Generalizing sentiment analysis: A review of progress, challenges and emerging directions. *Social Network Analysis and Mining*. Advance online publication. <https://doi.org/10.1007/s13278-025-01461-8>