A Review of Artificial Intelligence for Sentiment Analysis in Social Media Data

Manish Kumar Singla[1,*], Amel Ali Alhussan

[1]Department of Electrical Engineering, Thapar University, Patiala, Punjab, India

Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh 11671, Saudi Arabia

Emails: msingla0509@gmail.com; aalhussan@pnu.edu.sa

Abstract

Social media sentiment analysis has benefited from the miracle of artificial intelligence (AI), mainly how it can handle large, conflated data sets and distill valuable insights. In this review, the authors consider the positive impact of AI in business, health care, politics, and social justice, including marketing, mental health screening, misinformation, and multilingualism. Using ML and NLP, artificial intelligence technologies empower real-time analysis of the social trends and behaviors that affect decision-making and social interactions. However, many challenges are still reflected in data imbalance, ethical concerns relating to privacy and consent, and difficulties in processing dynamic content and several modalities, languages, and emotional states. Such limitations call for interdisciplinary collaboration and comprehensible ethical guidelines, including the FAIR principles for bettering data stewardship and ensuring no biases in AI systems. When developed as scalable, context-aware, and equitable systems, opinion mining may help solve social dilemmas and build an inclusive digital environment. Based on current trends, challenges, and suggested future directions, this review underlines the need for ethical, interdisciplinary, and culturally sensitive approaches to unlock the proper potential of AI in SA and social media sentiments.

Keywords: Artificial Intelligence; Sentiment Analysis; Social Media; Machine Learning; Natural Language Processing; Ethical AI

1. Introduction

Social media has produced vast volumes of information that Artificial Intelligence (AI) can only effectively analyze and interpret. As a result, AI is opening up new revolutionary avenues in new-age applications like sentiment analysis, informational and mental health monitoring, and multilingual information processing. With further developments, such technologies will also solve social problems, such as distributing prohibited materials and diagnosing mental disorders early. Still, the increased Application of AI in social media analysis raises challenges such as imbalance sampling, ethical issues, and scalability challenges. To better understand this interaction between AI and the world of social media, this section elucidates the focus areas, trends, and challenges to help lay out the advantages and disadvantages of using AI in social media.

1.1 AI and ML in Social Media Sentiment Analysis: Impacts and Implications

Technological advancement has played a significant role in influencing the growth of social media by allowing large volumes of text data to be produced and disseminated daily. This type of transformation has

DOI: https://doi.org/10.54216/MOR.020201
Received: May 24, 2024 Revised: September 10, 2024 Accepted: December 06, 2024

created new opportunities to use AI and ML for the proper analysis of social media sentiment efficiently. These advanced technologies have enabled researchers and practitioners to draw usable information from challenging large-scale unstructured data sources to meet the need to address the issues related to misinformation, public sentiments and the modeling of behaviors. Besides helping us understand societal trends, AI sentiments help us make better decisions in domains such as marketing, health, or political campaigning [1].

The work related to employing AI and ML for sentiment analysis has significant tendencies that affect societal, economic, and informational aspects. These technologies can help decide the general narrative and thus answer scaled-up questions, such as mental health surveillance or misinformation identification. AI tools benefit businesses by allowing them to gain insight into the consumer market, thus enabling them to develop strategies to address this area and use available resources in the best way possible. Informationally, sentiment analysis is beneficial in determining the trend and patterns of social media interactions to fashion public opinion. Such developments demonstrate that AI and ML are cardinal to turning social media into a tool for mastering modern processes [2].

1.2 AI in Social Media: Misinformation, Sentiment, Mental Health, and Multilingual Processing

Identifying new and relevant topics for using artificial intelligence (AI) and machine learning (ML) in social media analysis is as follows: Active misinformation management, sentiment classification, mental health detection, and multilingual processing. Fake news is a social problem because it can easily be found and circulated online. NLP is one of the most applied AI technologies for tracking and preventing fake news dissemination. Sentiment analysis, another important domain, is used to analyze the sentiments of people and their interactions through social media platforms for helping in business and political decision-making besides helping healthcare organizations. In addition, mental health detection can be counted as an important part of detection with the help of artificial intelligent systems that can detect depressive tendencies and other mental health disorders from the posts shared on social media to help start early interventions [3].

The Application of AI is also relevant in these areas, where multilingual processing is an essential component of globalization and digital integration. With the help of text analysis, AI systems can analyze texts in several languages and, therefore, obtain insights concerning various cultures. These focus areas signal the interdisciplinary nature of AI applications, which involves inputs from linguistics, psychology, data sciences and social sciences, among others. The synthesis of different fields helps AI technologies solve various problems in different areas in a comprehensive way as advances in contemporary digital issues [4].

1.3 Advancements in Large Language Models: Detecting Tobacco Marketing and Mental Health Indicators

Breakthroughs in large language models have greatly improved AI's capacity to identify subtle social phenomena, such as the marketing of tobacco products and signs of mental illness on social media. Such models, trained on terabytes of textual information and the most advanced machine learning algorithms, can analyze language texture, polarity, and references necessary for identifying trick aims at manipulation in a marketing campaign or signs of an upset mind. Fine-tuned LLMs accurately detect promotional content associated with toxic products such as tobacco for better vigilance and regulatory action. Similarly, LLMs perform a similar task of helping in the identification of Mental Health Indicators such as Depressive Language or Suicidal Intent, which could be extremely valuable inputs for Mental Health Care Professionals and may assist in the early identification of the possible disorder [5].

These technological advancements also provide potentially exciting approaches for other societal issues, including detecting fake news and recognizing toxic users. The increasing ability of LLMs to manage significant amounts of social media data in real time is helping to control trends, comprehend underlying stories that are not visible, and respond to problems such as cyberbullying, hate speech, or the proliferation of dangerous content. With the advancement of Artificial intelligence in every sector of society, the Application of law-leveraging technologies will provide socio-economic solutions to different societal problems and change the overall avenue of digital marketing while making some unforeseeable ethical questions relating to privacy, data bias, and algorithm examination [6].

DOI: https://doi.org/10.54216/MOR.020201
Received: May 24, 2024 Revised: September 10, 2024 Accepted: December 06, 2024

1.4 Trends and Challenges in AI for Social Media: Data Imbalance, Ethics, and Scalability

While artificial intelligence (AI) has increasingly become integrated into several areas of social media analysis, several challenges that affect the appropriate use of the technology have been realized. One of the main problems is data distribution; the data used to train machine learning often needs to be revised to simulate real situations accurately. This causes the results to be skewed, such as when some sentiments are repeated numerous times while others are skipped in comparison. Secondly, social factors are becoming more of a concern where ethical issues, especially privacy and consent and AI technologies, are concerned. As AI systems are used for analyzing personal data, like signs of mental health or political opinions – people start wondering how personal information will be misused and what will happen [7].

Another great difficulty is related to the evolution of artificial intelligence models for digital spaces. Since social networks are developing and producing vast amounts of data, AI systems' real-time data processing also raises significant issues. Although recent breakthroughs have boosted the processing capacity of machines, developing models scalable for global multilingual programs is still a challenge. Moreover, digital ecosystems are highly complex—compared to text, content can be dynamic, multimodal, and varied—and this causes an additional challenge in maintaining both the efficiency of the AI tool and its ability to maintain accuracy and fairness across different contexts. These challenges make it clear that using AI in large-scale solutions will require constant work on creating new civil solutions that incorporate AI, assuring that any AI solution will be built efficiently while also being built ethically and without prejudice [8].

In conclusion, AI has important applications in social media sentiment analysis. The analysis shows how it can provide new and creative approaches to such socially complex issues. Although there have been remarkable strides, some areas remain to be overcome, like data imbalance, ethical issues, and the problem of scope to be addressed to make it possible to use those advanced insightful AI techs. Understanding such issues helps us remember the significant requirement for integrating interindustry approaches to enhance AI models and adoption and work on more accuracy and equity. However, it will become evident that as new and improved AI technologies continue to break the surface, there is a strong need to carry out further studies in these areas to address the issue of biases and to find possible advancements in scaling up the proposed solutions across different interfaces of digital platforms. In surmounting these challenges, AI can contribute to constructing a sophisticated, fair, and moral information society in cyberspace.

2. Literature Review

The role and Application of social media and digital technologies in the production, dissemination and consumption of knowledge: the opportunities and challenges. In the current shifting environment, the increasing use of such problems as misinformation, sentiment analysis, mental health detection and multilingual processing shows the level of technological development required. Increased use of machine learning and artificial intelligence has provided fresh opportunities for tackling these issues or providing sound analyses and recommendations in an ever-widening range of fields. This sizeable literature review aims to present the latest trends and developments in AI, specifically machine learning and the impact created while enhancing the overall societal, economic and informational matrices. This section describes advanced practices and results as a basis for recognizing how AI can solve modern digital issues.

As detailed in [9], the study addresses the challenge of monitoring tobacco promotion on social media by developing an artificial intelligence model that leverages a fine-tuned large language model to identify promotional content with high precision (81.1%) and recall (87.8%). Through the analysis of 177,684 Turkish-language tweets, mechanisms such as modeling behavior, expressing positive attitudes, recommending use, and marketing brands or vendors were identified as key strategies of tobacco promotion. This AI tool facilitates effective surveillance and enforcement of tobacco control policies. It creates opportunities for health promotion by enabling research, strengthening community actions, and fostering supportive environments to combat tobacco marketing in dynamic digital spaces.

As outlined in [10], social media awareness campaigns could decrease youth vaping prevalence, but practical implementation requires addressing limitations in AI-generated materials. A mixed-methods study in Queensland, Australia, evaluated AI-generated campaign materials using two approaches: A basic method involving zero-shot prompting, automatic text-image integration, and a refined method incorporating few-shot prompting, multi-prompt usage, and manual text-image integration based on youth feedback. Phase 1 focus groups (n=10) highlighted technical issues such as misaligned text and images, awkward language, and

DOI: https://doi.org/10.54216/MOR.020201
Received: May 24, 2024 Revised: September 10, 2024 Accepted: December 06, 2024

outdated visuals, while Phase 2 interviews (n=9) revealed six key characteristics of effective messaging: visual appeal, focus on immediate consequences, relevance to youth, clear calls to action, avoidance of fear-mongering, and tailored content for diverse youth. The findings emphasize the need for co-designing with young people to refine AI-generated materials and ensure they resonate effectively with their intended audience.

As outlined in [11], membrane engineering involves designing suitable membrane processes for specific purposes and advancing the operation of membrane technologies, with reverse osmosis (RO) being a significant focus area. An analysis of 1424 articles from the Scopus database revealed that research in this domain has gained momentum since 2009, with thin-film composite (TFC) polymeric materials dominating the field through 550 publications. Additionally, using nanomaterials and polymers has resulted in 821 articles, while challenges such as fouling, biofouling, and scaling were addressed in 324 studies. Key contributors include Wang J., with the highest number of publications (73), and Gao C., excelling in other metrics. The Journal of Membrane Science emerged as the leading publication platform, and social network analysis identified five core research clusters. Sentiment analysis indicated that abstracts in this domain are positively perceived, objective, and emotionally neutral, providing a robust foundation for future advancements.

As outlined in [12], the rise of social media as an integral part of daily life has opened avenues for analyzing political sentiments through user comments. This study analyzed over 14,000 political comments using machine-learning models such as logistic regression, linear support vector classification (SVC), random forest, decision tree classification, and naive Bayes. Linear SVC demonstrated the highest accuracy at 91.18%, followed closely by logistic regression at 90%. The research offers actionable insights into model suitability by addressing data imbalance and evaluating models using performance metrics like accuracy, precision, recall, and F1 scores. The findings introduce an optimized approach to political sentiment analysis, enhancing robustness and accuracy, and providing a strong foundation for understanding sentiments on social media platforms.

As outlined in [13], emotional artificial intelligence plays a pivotal role in recognizing emotions through text mining, particularly in detecting depression-related content on social media. This study analyzed 983 Facebook posts using natural language processing techniques such as stemming, stop word removal, and feature extraction methods like TF-IDF, stylometric features, and word embedding. Machine learning models, including LSTM, GRU, support vector machines, and Naive Bayes, were employed to classify depressive text, with performance evaluated using metrics such as accuracy and F1-score. The research highlights the potential of sentiment analysis to aid psychologists in identifying depression through social media posts, ultimately contributing to early diagnosis and treatment to mitigate adverse behaviors and improve mental health outcomes.

As detailed in the paper [14], branding has evolved into a multidisciplinary domain influenced by digital technologies, artificial intelligence, and advanced linguistic methods. This study, utilizing tools such as VOSviewer, Biblioshiny, and CiteSpace, explores the intersection of branding and linguistics, analyzing data from Scopus to uncover shifts in traditional branding strategies driven by machine learning, sentiment analysis, and social media. Key contributions from leading authors, primary publication sources, and international collaborations were identified, while thematic and co-occurrence analyses revealed a growing emphasis on databased and computational approaches. Additionally, research gaps were noted in areas such as the long-term effects of digital branding, ethical considerations of AI-driven photo strategies and multilingual branding, providing a foundation for future research and innovation in the field.

As outlined in [15], branding has evolved into a complex multidisciplinary field shaped by integrated digital technologies, artificial intelligence, and sophisticated linguistic strategies. This study employs tools such as VOSviewer, Biblioshiny, and CiteSpace to analyze the intersection of branding and linguistics, using data from Scopus to identify trends, themes, and collaboration networks. It examines how machine learning, sentiment analysis, and social media have transformed traditional branding strategies and highlights contributions from leading authors, key publication sources, and patterns of international collaboration. Thematic and co-occurrence analyses reveal a significant shift toward data-driven and computational approaches. At the same time, research gaps are identified in areas such as the long-term impacts of digital branding, ethical concerns surrounding AI-generated visuals, and multilingual branding techniques. These findings provide a valuable foundation for advancing research and practical innovation in branding.

DOI: https://doi.org/10.54216/MOR.020201
Received: May 24, 2024 Revised: September 10, 2024 Accepted: December 06, 2024

In research discussed in [16], sentiment analysis evaluates user opinion toward Telkomsel internet network services using Twitter data collected from Jul. 7, 2020, to Dec. 31, 2022, totaling 30004 data and sample 3900 data-labeled. The Machine Learning models used were K-nearest Neighbour (KNN), Support Vector Machine (SVM), and Ensemble KNN-SVM based on majority vote and average. The results show the performance of the KNN model with K=15 achieved an 83.21% accuracy, SVM using C=100 achieved 84.33%, the Ensemble KNN-SVM majority vote gets an accuracy of 83.26%, and the Ensemble KNN-SVM average reached a highest accuracy of 84.79%. The four models also applied to the unlabeled data, showing the predicted majority sentiments as unfavorable. Thus, this finding indicates that opinions of society about the quality of Telkomsel internet service are collected primarily negatively. Overall, the KNN, SVM, and Ensemble KNN-SVM models perform efficiently for sentimental analysis with labeled and unlabeled data, thus making them a robust evaluation method of sentiment evaluation among users. In this study, sentiment analysis was conducted to evaluate user perceptions against the internet network service of Telomer, utilizing tweets within the range from Jul. 7, 2020, to Dec. 31, 2022, consisting of 30004 raw data, 3900 of which labeled data were used as a sample. Models for machine learning classification include K-Nearest Neighbour (KNN), Support Vector Machine (SVM), and Ensemble KNN-SVM based on majority vote and average. The KNN-MODEL with K=15 germinated the result of 83,21%, while SVM with C=100 matured with 84.33% rate of accuracy, Ensemble KNN-SVM majority vote gave origin accuracy of 83,26%, and the Ensemble KNN-SVM average scorched the highest with 84.79%. All four models were also applied to the unlabeled data, with the majority prediction of the sentiment as unfavorable. This finding indicates that public opinion towards Internet service quality provided by Telkomsel is primarily negative. Overall, KNN, SVM, and Ensemble KNN-SVM perform well in sentiment analysis for labeled and unlabeled data, making it a robust method for sentiment evaluation among users.

The growing access and usage of social media have significantly contributed to the growth of text data. This makes text mining an advanced area within data mining, as discussed in [17]. This multidisciplinary field borrows techniques from information retrieval, text analysis, information extraction, and clustering. A key aspect of text mining is text classification, which assigns documents to predetermined categories such as politics, economics, sports, and more. The study compared the performance of various classification algorithms, wherein Naive Bayes performs with an accuracy of 77.78% and run time of 0.046 seconds, SVM reaching the highest accuracy of 80.60% with 0.43 seconds runtime, Random forest achieving 70.1% accuracy in 150 seconds, while KNN had 24.88% accuracy within 15 seconds. These findings suggest that statistical methods, such as Naive Bayes and SVM, can be effective and efficient for text emotion classification, especially for large-scale datasets.

The custom architecture for feature extraction via pre-trained models BERT and RESNET50V2 was created in the research presented in [18] to solve an extensive issue in cyberbullying on social networks concerning multimodal content, combining text and distinct images. Overall, the model was trained on 149,823 instances of image-text pairs and achieved a testing accuracy of 78%. Performance improvements were also made with cross-attention, usage of dropout, and focal loss, specifically in rare classes. In comparison, a separate text-only model of the best-in-class BERT achieved a slightly lower accuracy of 75% when trained on the same data. Again, it shows that the multimodal model has a superior capacity to detect cyberbullying, mainly when dealing with complex integration in social media text accompanied by images.

Currently, crowdfunding and, more importantly, digital crowdfunding have established their foothold in the financial technology sector. According to [19], the relationship between a project and its backers tends to finance the project through social networks since the project creator contacts them; therefore, this paper introduces a machine-learning-enabled framework for analyzing the numerical and textual Kickstarter data to predict the success and delivery of crowdfunding projects. The proposed methodology is beneficial to the stakeholders in evaluating the credibility of the creator, lower risk of projects, and higher confidence for backers. The main aspects include preprocessing for preparing the dataset, feature extraction for prediction, and comparing training for five numerical machine-learning models and three text mining-based techniques. Results suggest that the SVM model has high performance, including accuracy, precision, recall, F1 scores, and training latency for numerical data. When it comes to textual analysis, BERT advances on what is perceived as complex datasets, while Word2vec handles the most straightforward features. These findings paved the way for machine learning applications to optimize outcomes from crowdfunding projects.

In the study, cited as [20], challenges concerning language identification in the code-mixed texts, particularly about Persian-English social media messages, were addressed by creating the PinLID dataset. Language

DOI: https://doi.org/10.54216/MOR.020201
Received: May 24, 2024 Revised: September 10, 2024 Accepted: December 06, 2024

identification is important for other applications like sentiment analysis and question answering. However, it could be more robust concerning multilingual and code-mixed contexts such as Twitter texts with heavy language mixing. PinLID provides a collection annotated at both sentence and token levels, thus enabling accurate identification of languages in the case of Persian-English code-mixed tweets. The dataset was evaluated with many machine learning classification algorithms, such as Classical Support Vector Machine (SVM), multilingual BERT, XLM-RoBERTa, ParsBERT, AriaBERT, and PersianLLaMA. The performance showed the same incredibly high effectiveness with an F1 score of 99.59% at both the sentence and token levels. This would enhance intelligent systems that rely on natural language processing in multilingual and code-mixed conditions.

The present article also chronicles the broader evaluation undertaken in [21] a social and economic appraisal of current digital ecosystems, extending the concept to capture a "Circular Digital Ecosystem," where users are expected to participate actively by sharing data and digital content, whereby they transform themselves from consumers into producers. A special focus falls on advanced AI algorithms and the data that trains them, such as those empowered by generative AI models like LLMs. The study notes that shared data quality hugely affects algorithm outcomes; hence, bias in data could lead to potentially harmful or inappropriate results. The comparison of how scientific research data is managed according to FAIR (Findable, Accessible, Interoperable, Reusable) principles with the unstructured variety daily found in internet data brings into perspective the urgent need for much more effective data management at the societal level. The authors encourage the incorporation of FAIR principles into much broader data-sharing practices, with an eye on the data manager and curator in risk reduction and improvement of the reliability of AI outcomes being emphasized in this regard. This approach is intended to elevate the awareness of users in the digital ecosystem regarding their part in enhancing data quality while at the same time lowering the potential dangers brought forth by the misuse of data.

In the study described in [22], the significant problem of detecting depression from social media content is discussed, stressing the fact that if left untreated, depression can lead to suicide. The data mining study with machine learning examined 10,000 data posts from Facebook, comments section and YouTube comments to locate symptoms of depression. This paper adapted six classifiers to differentiate between the depressed and non-depressed, and our result indicated that SVM gave the best result in the process. This research paves the way for subsequent research indicating that social media is effective in assisting in identifying users who may be struggling with their mental health and that machine learning should be pursued as an effective tool in detecting and helping those who need help.

As mentioned earlier in [23], social media has become an integral part of everybody's life, and as a result, social media comments are a rich source for sentiment analysis of political issues. This study used a dataset containing over 14,000 comments related to political issues to analyze sentiments using five models: logistic regression, linear support vector classification (SVC), random forest, decision tree classification, and Naïve Bayes. Comparing these, the one that performed best was Linear SVC, with an accuracy level of 91.18%, with logistic regression coming second with an accuracy level of 90%. The research also discussed issues such as data imbalance and presented recommendations concerning the applicability of each algorithm. Thus, the given research introduces a more developed approach toward model selection, thus providing the methodological foundation for examining political sentiments in social media technologies and enhancing the sentiment analysis methods.

Similar to the work described in [24], novel instances of disseminating such fake news have emerged secondary to advancements in social media platforms to disseminate text and images. Given this, the study proposes a self-learning multimodal model for fake news classification that uses contrastive learning for feature extraction without using labeled data and including the features of Large Language Models (LLMs). Using different language data, LLMs also outcompete simultaneously, allowing the model to handle text and image inputs concurrently. Performance analysis on a public benchmark dataset shows that the proposed method outperforms existing classification approaches with an accuracy of more than 85%, coupled with precision, recall, and F1-score. The document emphasizes the usefulness of this model for addressing the further development of false information identification mechanisms in a multimedia environment.

Table 1 presents an evidence-based manuscript of the reviewed literature, which addresses multiple focus areas, methodological approaches, and key findings across the studies. The competition offers information about specific use cases related to machine learning and artificial intelligence, including text mining and

DOI: https://doi.org/10.54216/MOR.020201
Received: May 24, 2024 Revised: September 10, 2024 Accepted: December 06, 2024

cyberbullying detection, sentiment analysis, multilingual language identification, and fake news detection. The table also points to the efficiency of the selected models, such as SVM and BERT, and the progress made in the field of multimodal analysis. It indicates the role of frame webs, such as FAIR principles, in evaluating the digital ecosystem. Thus, these results highlight the further possibilities of AI in solving modern and future problems and contribute to the further development of studies in social, economic and technological fields.

Table 1: Summary of Literature Review

Study ID	Focus Area	Methodology	Key Findings
[9]	Tobacco Promotion on Social Media	AI model leveraging fine-tuned large language models analyzing 177,684 Turkish-language tweets.	Achieved precision of 81.1% and recall of 87.8%. Identified key promotion strategies and supported tobacco control policy enforcement.
[10]	Youth Vaping Campaigns	Mixed-method study in Queensland using AI-generated materials evaluated by focus groups and interviews.	Highlighted six characteristics for effective messaging emphasized co-designing materials with youth for better resonance.
[11]	Membrane Engineering	Analysis of 1,424 articles using bibliometric tools focusing on reverse osmosis and thin-film composite materials.	Identified key contributors, challenges like fouling, and five core research clusters. Positive sentiment in abstracts provided a robust research foundation.
[12]	Political Sentiment Analysis	Machine learning models (SVC, logistic regression, random forest, etc.) analyzing 14,000+ political comments.	Linear SVC achieved 91.18% accuracy, offering insights into model suitability and robustness in sentiment analysis.
[13]	Emotional AI and Depression	Natural language processing with machine learning (LSTM, GRU, SVM, Naïve Bayes) analyzing 983 Facebook posts.	Identified depressive text with high accuracy, aiding early diagnosis and treatment of depression.
[14]	Branding and Linguistics	Bibliometric tools analyzing branding trends in machine learning, sentiment analysis, and social media.	Found gaps in digital branding is long-term impacts and ethical issues in AI-driven strategies, highlighting areas for future research.

DOI: https://doi.org/10.54216/MOR.020201
Received: May 24, 2024 Revised: September 10, 2024 Accepted: December 06, 2024

[15]	Branding in the Digital Age	Similar bibliometric tools as [14], focusing on collaboration networks and computational branding approaches.	Reinforced findings from [14] while emphasizing shifts to data-driven branding strategies.
[16]	Sentiment Analysis of Internet Services	Analysis of 30,004 tweets using KNN, SVM, and Ensemble KNN-SVM models for sentiment evaluation.	SVM showed highest accuracy (84.33%), with most sentiments being unfavorable towards Telkomsel internet services.
[17]	Text Mining and Classification	Comparison of classification algorithms (Naïve Bayes, SVM, etc.) for text emotion analysis.	SVM performed best with 80.60% accuracy, demonstrating effectiveness in text classification for large datasets.
[18]	Cyberbullying Detection	Multimodal model combining BERT and RESNET50V2 trained on 149,823 image-text pairs.	Achieved 78% accuracy; superior to text-only models, especially for complex multimodal data.
[19]	Crowdfunding Analysis	Machine-learning-enabled framework analyzing Kickstarter data with numerical and textual techniques.	SVM excelled in numerical data, while BERT was most effective for textual analysis, supporting better decision-making for crowdfunding stakeholders.
[20]	Code-Mixed Text Language Identification	Development of PinLID dataset with machine learning models like multilingual BERT and ParsBERT.	Achieved an F1 score of 99.59%, enhancing NLP applications in multilingual and code-mixed contexts.
[21]	Circular Digital Ecosystem	Evaluation of FAIR data principles for digital content sharing and user participation.	Stressed quality data sharing practices to improve AI outcomes and mitigate risks associated with biased data.
[22]	Depression Detection on Social Media	Machine learning classifiers analyzing 10,000 posts and comments from social media.	SVM provided best results, highlighting potential for aiding mental health detection and interventions through social media data.

DOI: https://doi.org/10.54216/MOR.020201
Received: May 24, 2024 Revised: September 10, 2024 Accepted: December 06, 2024

[23]	Political Sentiment on Social Media	Similar to [12], analyzing 14,000+ comments with various machine-learning models.	Linear SVC had top performance, reinforcing applicability for sentiment analysis in political discourse.
[24]	Fake News Classification	Self-learning multimodal model using contrastive learning and LLMs for text-image inputs.	Outperformed existing approaches with >85% accuracy, offering advancements in identifying false information in multimedia environments.

The literature showcased proves that AI and machine learning can work wonders and solve the various emergent problems in the digital world for SA, including sentiment analysis and mental health detection, language identification, and combating fake news. Although it is clear that significant progress has been made, problems persist, ranging from lack of data and data bias to the ethical concerns that revolve around the use of AI. The FAIR principles and user awareness connected to the practical use of AI technologies call for further innovation to maintain proper and efficient usage of AI. The staking present subsequent knowledge that makes a way toward designing more comprehensive, credible, and revolutionary digital environments. This review emphasizes the need for multidisciplinary efforts to allow artificial intelligence to solve some of society's challenges.

3. Discussion

Based on the sentiment analyses in social media platforms using artificial intelligence (AI), advancements across different fields, such as public health, politics, marketing, and extreme content regulation, are experienced. However, this field also presents severe challenges and ethical issues that need to be solved for the practical realization of this area [25].

3.1 Key Impacts and Opportunities

The main advancement in sentiment analysis has arisen from the Application of AI technologies such as Machine learning and natural language processing. These insights are instrumental in several areas [26]:

- Business and Marketing: Consumer preferences are best known in business through analysis of the use of artificial intelligence to assist in marketing strategies and the allocation of resources.
- Mental Health Monitoring: Modern machine learning techniques identify possible depressive shifts or other mental disorders in the text of the user's posts and provide an opportunity to intervene at an early stage.
- Fake News Mitigation: Misinformation is detected and prevented through NLP, and therefore, information assurance is improved.
- Multilingual Processing: AI systems that deal with content in two or more languages improve globalization and international communication.
- Regulatory Oversight: Identifying promotion messages about risky goods, including tobacco products, benefits health initiatives and other enforcement processes.

3.2 Emerging Trends

LLMs made AI powerful in processing language complexity, allowing it to analyze text with images and work with real-time social media content streams. First, what these models can do is they can find toxic behavior in users and, at the same time, determine the sentiment in the message regardless of the language it uses. Moreover, these models are valuable tools for fighting against cyberbullying, hate speeches, and spreading fake news [27], [28].

DOI: https://doi.org/10.54216/MOR.020201
Received: May 24, 2024 Revised: September 10, 2024 Accepted: December 06, 2024

3.3 Concerns, Issues and Implications

Despite its promise, AI's integration into social media sentiment analysis is fraught with challenges [29], [30]:

- Data Imbalance: The problem is increased reliance on selected training datasets that contain slight variations in accuracy and biases in sentiment identification.
- Scalability: Currently, there is still a technical challenge in managing dynamic, multimodal, and high-volume social media content in real-time.
- Ethical Concerns: Those are all rather significant problems that need to be solved with ethical rules and regulation of personal data.
- Complex Digital Ecosystems: The interaction of textual, visual, and contextual data has been analyzed, complicating the construction of a fair and accurate AI model [31], [32].

3.4 Future Directions

To address these challenges, researchers emphasize:

- Interdisciplinary Approaches: Close integration of knowledge about natural languages, psychological and computational foundations for their use, and techniques from computer science for the integration into complete AI models [33].
- Fairness and Ethics: Utilizing principles such as FAIR to enhance data direction and decrease biases [34].
- Model Refinement: Improving AI algorithms for real-time, enabling scalable sentiment analysis relevant to contexts [35].

AI for SMS remains a revolutionary idea for assessing social media sentiment analysis; however, constant attempts to minimize its drawbacks are necessary. Further work and development need to happen to understand how to build a system that is efficient, fair, and moral, following the users' expectations needed by society.

4. Conclusion

AI application in Social Media, particularly sentiment analysis, is one of the most prominent breakthroughs of current innovative technologies of use to humans, especially when handling substantial textual data. Moreover, AI's capability to analyze social media data for patterns, trends and user sentiment has been handy in various fields such as business, health, politics, and social causes. For example, corporations use it to tailor their marketing approaches and manage resource distribution efficiently and more effectively. At the same time, public health takes advantage of it to track the indices of mental health and look for depressive language for early action. In addition, those pieces help determine people’s sentiments and behaviors regarding political matters that are important for policy-making and enhancing democracy.

However, the use of AI in sentiment analysis experiences several constraints that hold the potential Application in its inefficient stage. One of the key problems is that data distribution and training datasets often need to be more profound. Privacy issues, consent, and data use abuse are ethical and vital hurdles. For instance, AI for processing personal social media posts brings issues surrounding the ownership and security of individuals' data into the spotlight. Furthermore, the issue of scalability is still a technical one, with the current intelligent systems satisfying stability, adaptability, heterogeneity, non-stationarity and real-time learning concerning multimodal dynamic and high-volume data, particularly in multilingual and multicultural settings.

These challenges must be overcome by collaborating in an interdisciplinary manner. Computer science, linguistics, psychology, and ethics all have to come together to ensure that the models developed for AI are sturdy and comprehensive, as well as the cooperative environment and context awareness. It is possible and highly necessary to enhance data quality, including less bias or equitable results, by applying ethical models such as FAIR principles (Findable, Accessible, Interoperable, and Reusable). The continuation of AI algorithms and their improvement to solve other challenging problems when working with enormous volumes of real-time data such as text, image, or multimodal content becomes essential for scalability and

DOI: https://doi.org/10.54216/MOR.020201
Received: May 24, 2024 Revised: September 10, 2024 Accepted: December 06, 2024

efficiency. AI models must also be developed sensitively: many application circumstances and environments would be crass or inappropriate.

In conclusion, AI has already shown the enormous opportunities it can offer to enhance the circumstances around which sentiment analysis in social media depends. Ethical issues, technical bottlenecks, and cultural sensitivity should be targeted for further research and innovations. AI technologies can enhance the possibilities of searching for fresh, feasible, and sustainable methods to fulfill societal demands with the spirit of obtaining multidisciplinary fairness and equity. Ultimately, these endeavors are set towards helping construct a just, connected, moral digital society where AI is used to expand human cognition, not harm.

References

[1] H. Küçükali and M. S. Erdoğan, "AI for tobacco control: identifying tobacco-promoting social media content using large language models," Nicotine & Tobacco Research, Nov. 2024, doi: 10.1093/NTR/NTAE276.

[2] M. S. Hossain, M. R. Islam, Dr. B. R. Riskhan, M. M. H. HASAN, and R. I. ISLAM, "Political sentiment analysis using natural language processing on social media," International Journal of Applied Methods in Electronics and Computers, vol. 12, no. 4, pp. 81–89, Dec. 2024, doi: 10.58190/IJAMEC.2024.108.

[3] S. Ahmed, S. Rakin, M. W. I. Waliur, N. B. Islam, B. Hossain, and Md. M. Akbar, "Depression detection from Social Media Bangla Text Using Recurrent Neural Networks," Dec. 2024, Accessed: Dec. 13, 2024. [Online]. Available: https://arxiv.org/abs/2412.05861v1

[4] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," 1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings, 2013.

[5] H. Küçükali and M. Sarper Erdoğan, "AI for Tobacco Control: Identifying Tobacco-promoting Social Media Content Using Large Language Models," Nicotine Tob Res, Nov. 2024, doi: 10.1093/NTR/NTAE276.

[6] S. Ahmed, S. Rakin, M. W. I. Waliur, N. B. Islam, B. Hossain, and Md. M. Akbar, "Depression detection from Social Media Bangla Text Using Recurrent Neural Networks," Dec. 2024, Accessed: Dec. 13, 2024. [Online]. Available: https://arxiv.org/abs/2412.05861v1

[7] T. Joseph, "Natural Language Processing (NLP) for Sentiment Analysis in Social Media," International Journal of Computing and Engineering, vol. 6, no. 2, pp. 35–48, Jul. 2024, doi: 10.47941/IJCE.2135.

[8] H. Chen et al., "A Self-Learning Multimodal Approach for Fake News Detection," Dec. 2024, Accessed: Dec. 13, 2024. [Online]. Available: https://arxiv.org/abs/2412.05843v1

[9] H. Küçükali and M. Sarper Erdoğan, "AI for Tobacco Control: Identifying Tobacco-promoting Social Media Content Using Large Language Models," Nicotine Tob Res, Nov. 2024, doi: 10.1093/NTR/NTAE276.

[10] T. Sun et al., "Title: Co-designing AI-generated vaping awareness materials with adolescents and young adults: A qualitative study."

[11] E. Aytaç, N. K. Khanzada, Y. Ibrahim, M. Khayet, and N. Hilal, "Reverse Osmosis Membrane Engineering: Multidirectional Analysis Using Bibliometric, Machine Learning, Data, and Text Mining Approaches," Membranes 2024, Vol. 14, Page 259, vol. 14, no. 12, p. 259, Dec. 2024, doi: 10.3390/MEMBRANES14120259.

[12] M. S. Hossain, M. R. Islam, Dr. B. R. Riskhan, M. M. H. HASAN, and R. I. ISLAM, "Political sentiment analysis using natural language processing on social media," International Journal of Applied Methods in Electronics and Computers, vol. 12, no. 4, pp. 81–89, Dec. 2024, doi: 10.58190/IJAMEC.2024.108.

[13] S. Ahmed, S. Rakin, M. W. I. Waliur, N. B. Islam, B. Hossain, and Md. M. Akbar, "Depression detection from Social Media Bangla Text Using Recurrent Neural Networks," Dec. 2024, Accessed: Dec. 12, 2024. [Online]. Available: https://arxiv.org/abs/2412.05861v1

[14] A. Lukose, R. S. Cleetus, H. Divya, T. M. Saravanakumar, and J. Jose, "Exploring the Intersection of Brands and Linguistics: A Comprehensive Bibliometric Study," International Review of Management and Marketing, vol. 15, no. 1, pp. 257–271, Dec. 2025, doi: 10.32479/IRMM.17538.

DOI: https://doi.org/10.54216/MOR.020201
Received: May 24, 2024 Revised: September 10, 2024 Accepted: December 06, 2024

[15] A. Lukose, R. S. Cleetus, H. Divya, T. M. Saravanakumar, and J. Jose, "Exploring the Intersection of Brands and Linguistics: A Comprehensive Bibliometric Study," International Review of Management and Marketing, vol. 15, no. 1, pp. 257–271, Dec. 2025, doi: 10.32479/IRMM.17538.
[16] M. F. Fachrudin, C. V. Angkoso, and D. A. Fatah, "Analisis Sentimen Pada Sosial Media Twitter Terhadap Kualitas Jaringan Internet Telkomsel Menggunakan Ensemble K-Nearest Neighbour - Support Vector Machine," Jurnal Teknologi Informasi dan Ilmu Komputer, vol. 11, no. 6, pp. 1253–1264, Dec. 2024, doi: 10.25126/JTIIK.1168713.
[17] G. Wijaya, D. Irawan, Z. Arifin, H. Oktavianto, M. Rahman, and G. Abdurrahman, "STUDI KLASIFIKASI TOPIK BERITA DENGAN ALGORITMA MACHINE LEARNING," J-ENSITEC, vol. 11, no. 01, pp. 10202–10206, Dec. 2024, doi: 10.31949/JENSITEC.V11I01.12037.
[18] J. Wandeto, I. Musyoka, and B. Kituku, "Integrating BERT and RESNET50V2 for Multimodal Cyberbullying Detection," 2024 Sixth International Conference on Intelligent Computing in Data Sciences (ICDS), pp. 1–6, Oct. 2024, doi: 10.1109/ICDS62089.2024.10756337.
[19] S. Pourroostaei Ardakani et al., "Identifying crowdfunding storytellers who deliver successful projects: a machine learning approach," The Journal of Supercomputing 2024 81:1, vol. 81, no. 1, pp. 1–30, Dec. 2024, doi: 10.1007/S11227-024-06785-4.
[20] A. Ghafouri, H. Naderi, and M. Firouzmandi, "PinLID: a dataset for Pinglish language identification based on code-mixing sentence on unstructured resources," Language Resources and Evaluation 2024, pp. 1–27, Dec. 2024, doi 10.1007/S10579-024-09783-3.
[21] S. Cozzini and M. de Luca, "Living in Digital Ecosystems: Are We Aware of This?," pp. 113–132, 2024, doi: 10.1007/978-3-031-76961-0_6.
[22] Z. N. Vasha, B. Sharma, I. J. Esha, J. Al Nahian, and J. A. Polin, "Depression detection in social media comments data using machine learning algorithms," Bulletin of Electrical Engineering and Informatics, vol. 12, no. 2, pp. 987–996, Apr. 2023, doi: 10.11591/EEI.V12I2.4182.
[23] M. S. Hossain, M. R. Islam, Dr. B. R. Riskhan, M. M. H. HASAN, and R. I. ISLAM, "Political sentiment analysis using natural language processing on social media," International Journal of Applied Methods in Electronics and Computers, vol. 12, no. 4, pp. 81–89, Dec. 2024, doi: 10.58190/IJAMEC.2024.108.
[24] H. Chen et al., "A Self-Learning Multimodal Approach for Fake News Detection," Dec. 2024, Accessed: Dec. 12, 2024. [Online]. Available: https://arxiv.org/abs/2412.05843v1
[25] X. Shen, M. Huang, Z. Hu, S. Cai, and T. Zhou, "Multimodal Fake News Detection with Contrastive Learning and Optimal Transport," Front Comput Sci, vol. 6, p. 1473457, Nov. 2024, doi: 10.3389/FCOMP.2024.1473457/BIBTEX.
[26] X. Shen, M. Huang, Z. Hu, S. Cai, and T. Zhou, "Multimodal Fake News Detection with Contrastive Learning and Optimal Transport," Front Comput Sci, vol. 6, p. 1473457, Nov. 2024, doi: 10.3389/FCOMP.2024.1473457/BIBTEX.
[27] H. Küçükali and M. S. Erdoğan, "Identification and classification of tobacco-promoting social media content at scale using deep learning: a mixed-methods study," Popul Med, vol. 5, no. Supplement, Apr. 2023, doi: 10.18332/POPMED/164106.
[28] M. S. Hossain, M. R. Islam, Dr. B. R. Riskhan, M. M. H. HASAN, and R. I. ISLAM, "Political sentiment analysis using natural language processing on social media," International Journal of Applied Methods in Electronics and Computers, vol. 12, no. 4, pp. 81–89, Dec. 2024, doi: 10.58190/IJAMEC.2024.108.
[29] J. Wandeto, I. Musyoka, and B. Kituku, "Integrating BERT and RESNET50V2 for Multimodal Cyberbullying Detection," 2024 Sixth International Conference on Intelligent Computing in Data Sciences (ICDS), pp. 1–6, Oct. 2024, doi: 10.1109/ICDS62089.2024.10756337.
[30] S. Ahmed, S. Rakin, M. W. I. Waliur, N. B. Islam, B. Hossain, and Md. M. Akbar, "Depression detection from Social Media Bangla Text Using Recurrent Neural Networks," Dec. 2024, Accessed: Dec. 13, 2024. [Online]. Available: https://arxiv.org/abs/2412.05861v1
[31] A. Ghafouri, H. Naderi, and M. Firouzmandi, "PinLID: a dataset for Pinglish language identification based on code-mixing sentence on unstructured resources," Lang Resour Eval, Dec. 2024, doi 10.1007/S10579-024-09783-3.
[32] S. Pourroostaei Ardakani et al., "Identifying crowdfunding storytellers who deliver successful projects: a machine learning approach," The Journal of Supercomputing 2024 81:1, vol. 81, no. 1, pp. 1–30, Dec. 2024, doi: 10.1007/S11227-024-06785-4.

DOI: https://doi.org/10.54216/MOR.020201
Received: May 24, 2024 Revised: September 10, 2024 Accepted: December 06, 2024

[33] S. Cozzini and M. de Luca, "Living in Digital Ecosystems: Are We Aware of This?," pp. 113–132, 2024, doi: 10.1007/978-3-031-76961-0_6.
[34] Z. N. Vasha, B. Sharma, I. J. Esha, J. Al Nahian, and J. A. Polin, "Depression detection in social media comments data using machine learning algorithms," Bulletin of Electrical Engineering and Informatics, vol. 12, no. 2, pp. 987–996, Apr. 2023, doi: 10.11591/EEI.V12I2.4182.
[35] H. Chen et al., "A Self-Learning Multimodal Approach for Fake News Detection," Dec. 2024, Accessed: Dec. 13, 2024. [Online]. Available: https://arxiv.org/abs/2412.05843v1

DOI: https://doi.org/10.54216/MOR.020201
Received: May 24, 2024 Revised: September 10, 2024 Accepted: December 06, 2024