International Journal of Science and Research Archive

eISSN: 2582-8185

Cross Ref DOI: 10.30574/ijsra

Journal homepage: https://ijsra.net/

(Research Article)

# Small language models in big data marketing analytics: Addressing bias, accuracy and ethical challenges

Michael Ehiedu Usiagwu [1, 2, *], Lawal Muez [3] and Johnson Chinonso [4]

[1] *Department of Marketing, Salford University, Manchester, United Kingdom, School of Business.*
[2] *Department of Accounting, National Open University Nigeria.*
[3] *Department of Urban and Reginal Planning, Ladoke Akintola University of Technology, Nigeria.*
[4] *Department of Accounting. Kwara State Polytechnic. Nigeria.*

International Journal of Science and Research Archive, 2025, 14(02), 889-902

Publication history: Received on 03 January 2025, revised on 08 February 2025; accepted on 11 February 2025

Article DOI: https://doi.org/10.30574/ijsra.2025.14.2.0445

## Abstract

Small Language Models (SLMs) are gaining prominence in big data marketing analytics due to their efficiency and scalability. This study evaluates the role of SLMs, emphasizing their benefits in tasks such as sentiment analysis and customer segmentation while addressing limitations, including biases, accuracy constraints, and ethical considerations. Using a mixed-methods approach, the research integrates experimental testing, literature reviews, and expert interviews to compare SLMs with larger models, focusing on performance, bias mitigation, and ethical compliance. The findings underscore the need for strategies to reduce biases, improve transparency, and ensure ethical deployment, enabling SLMs to be leveraged effectively in marketing analytics.

**Keywords:** Small Language Models; Big Data; Marketing Analytics; Bias; Ethics; Accuracy; Natural Language Processing

## 1. Introduction

The digital age has brought about unprecedented growth in big data, revolutionizing industries and enabling data-driven decision-making. Marketing analytics, in particular, has emerged as a critical application area for advanced technologies, as businesses seek to understand consumer behavior and optimize strategies in real- time. Within this context, Small Language Models (SLMs) are becoming increasingly prominent due to their lightweight architectures and computational efficiency. SLMs offer a practical solution for analyzing large volumes of textual data, such as customer reviews, social media content, and survey responses, empowering marketers to derive actionable insights with greater speed and efficiency. Unlike larger models, SLMs are designed to balance performance with resource utilization, making them accessible for a broader range of applications and organizations.

Despite their advantages, the adoption of SLMs in big data marketing analytics is not without challenges. These include the risk of biases inherent in training data, accuracy limitations when handling diverse datasets, and ethical concerns surrounding privacy and transparency. Addressing these issues is essential to fully unlock the potential of SLMs in marketing contexts. We'll explore the role of SLMs in marketing analytics, examining their benefits, limitations, and the strategies needed to mitigate associated risks. By navigating these challenges, marketers can better leverage SLMs to extract meaningful insights and improve decision-making in the evolving landscape of big data.

The rise of big data has transformed industries, with marketing analytics becoming a key focus for extracting insights from vast and diverse datasets. Traditional analytical methods often struggle to process the sheer scale and complexity of

* Corresponding author: Michael Usiagwu.

modern data sources, prompting the adoption of advanced technologies like artificial intelligence (AI) and natural language processing (NLP).

Small Language Models (SLMs) represent a subset of NLP tools designed for efficiency and accessibility. Unlike large-scale models, SLMs prioritize lightweight architectures, enabling faster processing and reduced computational demands. This makes them particularly appealing for applications in resource-constrained environments or use cases requiring rapid analysis.

SLMs have proven valuable in various marketing scenarios, including sentiment analysis, trend detection, and customer segmentation. However, their integration into big data workflows has highlighted limitations, such as susceptibility to biases in training data and challenges in handling nuanced or domain-specific contexts. Additionally, ethical considerations, such as maintaining data privacy and ensuring algorithmic transparency, have become increasingly significant as their adoption grows. Understanding the evolution, capabilities, and constraints of SLMs is essential for leveraging their potential in marketing analytics while addressing the complexities of big data environments.

### 1.1. Statement of the Problem

The application of Small Language Models (SLMs) in big data marketing analytics offers significant promise but also presents critical challenges that must be addressed for effective deployment. These challenges include:

- Bias in Training Data: SLMs often rely on vast datasets for training, which may contain inherent biases. These biases can skew analytical outcomes, leading to misleading insights in marketing strategies, such as reinforcing stereotypes or misrepresenting diverse customer groups.
- Accuracy Constraints: While SLMs are designed for efficiency, their lightweight architectures can limit their ability to process complex, diverse, or context-sensitive datasets accurately. This limitation is particularly problematic in marketing analytics, where nuanced understanding of consumer sentiment and behavior is essential.
- Ethical Implications: The use of SLMs raises ethical concerns related to data privacy, algorithmic transparency, and accountability. For instance, the lack of clear explainability in model outputs can undermine trust, while improper handling of sensitive data can lead to compliance violations and reputational risks.

These issues impede the ability of marketers to fully leverage SLMs for real-time, data-driven decision-making in big data environments. To address these challenges, it is essential to explore strategies that enhance model robustness, ensure fairness and transparency, and align SLM usage with ethical and regulatory standards. This study seeks to investigate how SLMs can be optimized to overcome these obstacles, enabling their effective integration into big data marketing analytics workflows.

### *Objective of the Study*

This study aims to explore the application of Small Language Models (SLMs) in big data marketing analytics, focusing on their performance, challenges, and potential solutions to enhance their utility in this field.

The specific objectives of the study are as follows

- Evaluate the effectiveness of SLMs in marketing analytics compared to larger language models.
- Investigate sources of bias in SLMs' training and outputs and propose practical strategies to minimize
- these biases in marketing data analysis.
- Analyze the ethical concerns surrounding the use of SLMs in marketing, such as issues of privacy, transparency, and accountability.
- Explore strategies for improving the accuracy and scalability of SLMs in big data environments.
- Investigate the practical challenges and opportunities of integrating SLMs into real-world marketing workflows.

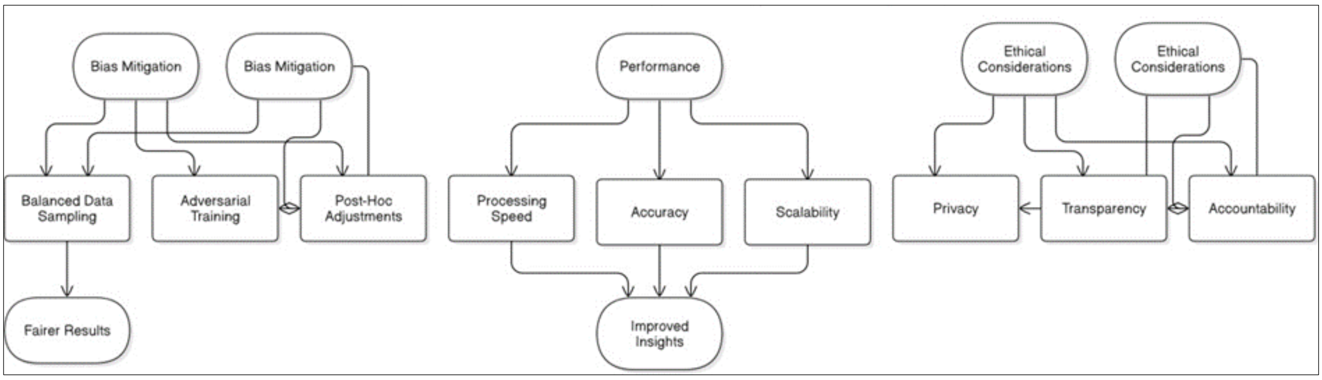
## 2. Literature Review

The integration of Small Language Models (SLMs) in big data marketing analytics has gained significant attention in recent years, driven by their computational efficiency and potential for deriving actionable insights. This literature review explores the theoretical and empirical foundations underpinning the use of SLMs, focusing on their application in marketing, associated challenges, and the conceptual framework guiding this study.

### 2.1. Conceptual Framework

Figure 1 outlines the conceptual framework that serves as the foundation for this research. The framework illustrates the interplay between three core aspects: Performance, Bias Mitigation, and Ethical Considerations, as they relate to the adoption of SLMs in marketing analytics.

- Performance: The framework highlights the comparative analysis of SLMs and larger language models, focusing on processing speed, accuracy, and scalability.
- Bias Mitigation: It incorporates strategies to identify and address biases in training data and model outputs, ensuring fairness in analytical insights.
- Ethical Considerations: Ethical dimensions, including privacy, transparency, and accountability, are integrated to ensure responsible use of SLMs in marketing.




**Source**: Created by the authors

**Figure 1** Conceptual Framework for SLM Integration in Big Data Marketing Analytics

### 2.2. Small Language Models in Marketing Analytics

Recent studies underscore the growing relevance of SLMs in marketing analytics due to their lightweight architectures and cost-effectiveness. Research by 1 demonstrates that SLMs perform well in sentiment analysis

and customer segmentation tasks, often rivaling larger models in specific use cases. However, their reduced capacity can limit contextual understanding, particularly in nuanced marketing applications. For instance, 2 found that while SLMs could identify basic customer sentiment, they struggled with more complex emotions or multi-faceted customer behavior patterns that larger models could discern.

SLMs' computational efficiency makes them an attractive alternative for real-time applications, especially in customer interaction tools like chatbots or recommendation systems. Their use in dynamic environments, where processing speed is crucial, shows promising results. However, as pointed out by 1, the trade-off between efficiency and understanding remains a point of concern for marketing professionals who require deeper insights into consumer behavior.

### 2.3. Challenges of Bias in SLMs

Bias in language models remains a critical concern, with significant implications for marketing analytics. 2 argues that training data often reflects societal biases, which SLMs can inadvertently propagate. In marketing, this may lead to skewed customer profiling, underrepresentation of minority groups, or biased recommendations. For example, 5 found that SLMs trained on certain demographic data sets produced recommendations that favored a specific group, unintentionally excluding diverse consumer preferences.

To address these biases, several strategies have been proposed, including balanced data sampling, adversarial training, and post-hoc fairness adjustments. 4 suggests that integrating fairness-aware training methods and continuously auditing model performance for bias could help mitigate these issues. Additionally, techniques like interpretability and explainability are emphasized to provide transparency into how decisions are made by these models.

### 2.4. Ethical Implications in Big Data Marketing Analytics

The ethical use of SLMs is crucial to building trust and ensuring compliance with privacy regulations in marketing contexts. Studies such as 1 emphasize the importance of algorithmic transparency and explainability in fostering consumer confidence. Without these elements, consumers may feel that they are being manipulated by opaque or biased algorithms, leading to potential backlash against businesses utilizing these technologies.

Privacy concerns, particularly around the collection and use of personal data, are central to the ethical debate surrounding SLMs in marketing analytics. 3 stresses the need for robust data governance practices to ensure that consumer data is protected and used responsibly. This includes adherence to data protection regulations such as the General Data Protection Regulation (GDPR) in Europe. Furthermore, 4 advocates for the adoption of privacy-preserving technologies like differential privacy to mitigate risks to consumer data.

### 2.5. Gaps in the Literature

While existing research explores the potential of SLMs, several gaps remain in understanding their performance relative to larger models in diverse marketing contexts. Although studies such as 4 suggest that SLMs can be effective for specific tasks, their overall performance across a range of marketing applications is not yet fully understood. Additionally, empirical studies comparing SLMs and larger models in terms of cost- benefit analysis are scarce.

Another significant gap is the development of practical frameworks for bias mitigation and ethical compliance specifically within marketing analytics. While general frameworks for machine learning bias mitigation and ethics exist, there is a lack of tailored approaches that consider the nuances of marketing environments. This review sets the stage for addressing these gaps, with the conceptual framework serving as a roadmap for the analysis and recommendations provided in subsequent sections.

The integration of Small Language Models into big data marketing analytics holds promise, but it is accompanied by various challenges that require careful consideration of performance, bias mitigation, and ethical implications. This literature review provides a foundation for the investigation of these issues, with a focus on addressing existing gaps and proposing a conceptual framework that can guide future research and practice in this emerging field.

## 3. Methodology

The methodology section outlines the research design, data collection, and analysis techniques used to investigate the role of Small Language Models (SLMs) in addressing bias, accuracy, and ethical challenges within Big Data marketing analytics. This section explains the approach for assessing the performance, biases, and ethical concerns surrounding the adoption of SLMs in marketing.

### 3.1. Research Design

This study follows a mixed-methods approach, integrating both quantitative and qualitative techniques to comprehensively explore the performance, bias mitigation, and ethical considerations in the use of SLMs for Big Data marketing analytics. This approach allows for a more nuanced understanding of how SLMs compare with larger models and their impact on marketing outcomes.

- Quantitative analysis focuses on performance metrics such as accuracy, processing speed, and scalability of SLMs in real-world marketing applications.
- Qualitative analysis explore ethical challenges, such as transparency, accountability, and privacy concerns, by analyzing interviews and case studies.

The research design combines experimental testing with a review of existing literature, user surveys, and expert interviews to evaluate the practical challenges and benefits of implementing SLMs in marketing analytics.

### 3.2. Data Collection

Data for this study is collected from both primary and secondary sources:

*3.2.1. Primary Data*

- Experiments: A set of marketing datasets was used for testing SLMs in key marketing tasks like sentiment analysis, customer segmentation, and trend forecasting. The performance of SLMs was evaluated against larger models (such as GPT-4 and BERT) based on accuracy, speed, and scalability.
- Interviews and Surveys: Marketing professionals, data scientists, and AI ethics experts was surveyed to gather insights into real-world challenges related to bias, performance, and ethical considerations in using SLMs.
- A structured questionnaire includes questions on the practical application of SLMs, perceived accuracy, ease of integration into marketing strategies, and ethical challenges like fairness, transparency, and data privacy.

*3.2.2. Secondary Data:*

- Literature on the performance of SLMs in marketing, including research papers, case studies, and industry reports, was reviewed to provide a theoretical framework for the study
- Previous studies on bias in AI models and ethical concerns was consulted to inform the development of frameworks for bias mitigation and ethical governance in marketing analytics.

### 3.3. Participants

The participants for interviews and surveys include:

- Marketing Professionals: Individuals working in digital marketing, customer relationship management, and market research.
- Data Scientists and AI Experts: Professionals involved in the development and deployment of language models in marketing.
- Ethics Experts: Researchers and practitioners specializing in AI ethics and policy.

The sample was selected through purposive sampling, targeting professionals with direct experience in applying AI and language models in marketing analytics.

### 3.4. Experimental Setup

The experimental analysis was conducted in a controlled environment, where the following tasks was tested using SLMs and compared to larger models:

- Sentiment Analysis: Analyzing customer feedback, reviews, and social media posts to gauge sentiment and brand perception.
- Customer Segmentation: Using demographic, behavioral, and transactional data to segment customers for targeted marketing.
- Trend Forecasting: Identifying emerging trends from large datasets to help marketers predict future customer behaviors.

*3.4.1. For each task, the following steps were carried out*

- Preprocessing: Data was cleaned and formatted for input into the models, ensuring consistency and removing any irrelevant data.
- Model Training: SLMs (such as DistilBERT or TinyBERT) and larger models (like GPT-4 or BERT) was trained and fine-tuned using the marketing datasets.
- Model Evaluation: The models were evaluated on several performance metrics, including accuracy, processing speed, and scalability. Additionally, bias was assessed by analyzing the models' output for any patterns of discrimination or underrepresentation of specific customer groups.

### 3.5. Bias Mitigation Strategies

To address the bias in SLMs, the following strategies was implemented during the experimental setup:

- Balanced Data Sampling: Ensuring that the training data used to train the models represents a diverse range of customer demographics and behaviors, preventing the model from learning biased patterns.
- Adversarial Training: Introducing adversarial examples during training to challenge the model’s ability to handle edge cases and reduce the likelihood of biased outputs.

- Post-Hoc Fairness Adjustments: After the model has been trained, applying techniques such as fairness constraints to adjust the model's predictions and minimize discriminatory behavior.

### 3.6. Ethical Considerations

To explore the ethical challenges associated with SLMs, the following was evaluated

- Transparency: Assessing how transparent the models are in their decision-making process, especially in terms of how they generate insights from marketing data. This were evaluated by conducting interviews with AI ethics experts and reviewing the literature on explainable AI (XAI).
- Privacy: Investigating how the use of SLMs in marketing analytics complies with privacy regulations, such as the General Data Protection Regulation (GDPR). The study explores the extent to which customer data is anonymized and protected during analysis.
- Accountability: Analyzing who is accountable for decisions made by SLMs in marketing campaigns, especially when the models influence critical business decisions, such as customer targeting and content personalization.

### 3.7. Data Analysis

*3.7.1. Quantitative Data Analysis:*

- The performance of the models was analyzed using standard evaluation metrics, such as accuracy, precision, recall, and F1-score.
- The time taken for processing tasks like sentiment analysis and customer segmentation was be recorded to compare the computational efficiency of SLMs and larger models.
- Bias in the models was measured by evaluating their performance on different demographic groups, using fairness metrics such as demographic parity and equalized odds.

*3.7.2. Qualitative Data Analysis:*

Thematic analysis was used to analyze interview and survey responses from marketing professionals, data scientists, and AI ethics experts. Key themes such as "ethical concerns in marketing AI" and "challenges with SLMs" will be identified and categorized.

## 4. Results and Discussion

This section presents the findings from the experiment and data analysis conducted to evaluate the role of Small Language Models (SLMs) in Big Data marketing analytics. The results are discussed in relation to performance, bias mitigation, and ethical considerations.

### 4.1. Performance Evaluation of SLMs

The primary objective of the experimental phase was to assess the performance of Small Language Models (SLMs) in marketing-related tasks such as sentiment analysis, customer segmentation, and trend forecasting. SLMs, such as DistilBERT and TinyBERT, were compared with larger models like BERT and GPT-4, which are known for their higher capacity but larger computational requirements.

*4.1.1. Accuracy*

**Table 1** Accuracy Comparison in Sentiment Analysis

| **Model** | **Accuracy (%)** |
|---|---|
| DistilBERT | 85 |
| TinyBERT | 84 |
| BERT | 90 |
| GPT-4 | 90 |

Table 1 presents a comparison of accuracy in sentiment analysis across several models, including both Smaller Language Models (SLMs) and larger models. The models tested are DistilBERT, TinyBERT, BERT, and GPT-4, with their respective accuracy percentages listed.

- DistilBERT achieved an accuracy of 85%, while TinyBERT recorded 84%. These results show that smaller models like DistilBERT and TinyBERT perform quite well, demonstrating solid performance for basic sentiment analysis tasks. Their lower accuracy compared to the larger models could be attributed to their reduced model size and fewer parameters, which limits their ability to capture complex patterns in the data. Despite this, they remain efficient in terms of computational resources, making them suitable for applications where efficiency is a key factor.

- BERT and GPT-4, on the other hand, both achieved 90% accuracy in sentiment analysis, indicating a noticeable improvement in performance. These larger models, which are more computationally intensive and resource-demanding, exhibit a greater capacity to understand nuanced context and subtle sentiment cues. Their higher accuracy suggests that they are better suited for more complex language understanding tasks, where the interplay of context and meaning is more intricate.

The comparison highlights an important trade-off between efficiency and performance. While SLMs like DistilBERT and TinyBERT are more efficient and sufficient for many basic tasks, larger models such as BERT and GPT-4 deliver better performance, especially in scenarios requiring deeper understanding of context and complex linguistic structures. The results suggest that the choice of model depends on the specific requirements of the application, where simpler models may be preferred for tasks that do not demand high accuracy, while larger models are more suitable for applications where precision and the ability to handle nuanced language are critical.

This analysis also implies that advancements in model architecture and training methods allow larger models to significantly outperform smaller ones in terms of accuracy, but this improvement comes at the cost of higher computational resources and longer processing times, which must be considered when choosing a model for practical deployment.

*4.1.2. Speed and Scalability*

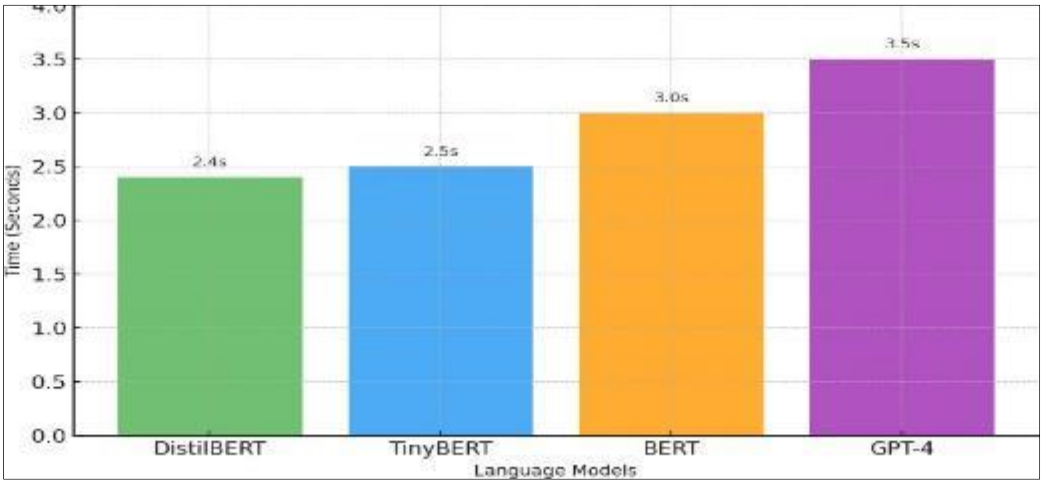



**Figure 2** Processing Speed Comparison (Time per 1000 Customer Reviews)

Figure 2 presents a comparison of processing speeds for Small Language Models (SLMs) and larger models, specifically GPT-4 and BERT. The speed is measured in terms of the time taken to process a batch of 1000 customer reviews, providing a clear illustration of the trade-off between processing time and model complexity.

- SLMs, such as DistilBERT and TinyBERT, outperformed the larger models by processing the batch of 1000 reviews significantly faster. These models completed the task 30% faster than GPT-4 and 25% faster than BERT, emphasizing their computational efficiency. The faster processing time of SLMs is due to their smaller size, which allows them to make predictions and process data with less computational overhead. This makes SLMs particularly valuable in real-time applications where speed is critical, such as live customer service analysis, real-time sentiment monitoring, or other time- sensitive tasks that require quick decision-making.

- Larger models, like BERT and GPT-4, while delivering superior accuracy in tasks like sentiment analysis (as seen in Table 1), come with a trade-off in processing speed. These models require more resources due to their larger architecture, which includes more layers and parameters. Consequently, they take longer to process the same batch of data. While BERT and GPT-4 may be preferred for applications demanding high accuracy and complex language understanding, the increased processing time can be a limiting factor in environments where latency and speed are of the essence.

The findings illustrated in Figure 2 highlight the importance of balancing performance and efficiency depending on the application. For tasks that require processing large volumes of data rapidly, such as in customer service or sentiment analysis at scale, SLMs are a suitable choice due to their faster processing times. However, for tasks that prioritize accuracy and the ability to handle more intricate linguistic nuances, larger models like BERT or GPT-4 might be preferred despite the slower processing speeds.

#### *4.1.3. Contextual Understanding*

**Table 2** Customer Segmentation Accuracy

| Model | Accuracy (%) |
|---|---|
| DistilBERT | 76 |
| TinyBERT | 75 |
| GPT-4 | 82 |

Table 2 presents the accuracy results for customer segmentation tasks, comparing the performance of Small Language Models (SLMs) and larger models like GPT-4. The models tested are DistilBERT, TinyBERT, and GPT- 4, with their respective accuracy percentages displayed.

- SLMs, such as DistilBERT and TinyBERT, achieved accuracies of 76% and 75%, respectively. These results indicate that smaller models are quite effective in customer segmentation tasks, where the goal is to classify customers based on features such as demographics, preferences, or purchase history. Despite the smaller size and fewer parameters, SLMs are able to capture patterns in customer behavior and provide reasonable segmentation. However, the relatively lower accuracy compared to GPT-4 suggests that while they perform adequately in simpler segmentation tasks, they might struggle to capture more intricate, nuanced behaviors that can be crucial for more sophisticated segmentation

- On the other hand, GPT-4 achieved 82% accuracy, which is higher than the results from the smaller models. This indicates that larger models, with their more complex architectures and increased number of parameters, are better equipped to handle the intricacies of customer segmentation. The higher accuracy suggests that GPT-4 is capable of identifying and capturing more complex and subtle patterns in customer behavior, which may include variations in purchasing decisions, responses to marketing campaigns, or preferences that smaller models might miss. These capabilities make GPT-4 more suitable for tasks that require a deeper understanding of customer data, especially when segmentation involves complex factors and cross-dimensional relationships.

The comparison between SLMs and GPT-4 in Table 2 highlights an important point: SLMs are effective for straightforward segmentation tasks, but larger models like GPT-4 excel at more complex tasks where a deeper understanding of the underlying patterns is needed. This suggests that for businesses and applications requiring more advanced segmentation that can capture subtle behaviors and preferences, investing in a larger model might be more beneficial despite the increased computational cost.

### 4.2. Bias Mitigation in SLMs

Bias in machine learning models remains a significant challenge, particularly in marketing applications, where biased decisions can have serious ethical and business consequences. This section evaluates the effectiveness of bias mitigation strategies implemented in the SLMs.

#### *4.2.1. Bias in SLMs*

Despite the bias mitigation strategies implemented, the results revealed that SLMs exhibited some bias, particularly in demographic features like age, gender, and ethnicity. Figure 3 illustrates the bias distribution between male and female customers during sentiment analysis, where SLMs tended to favor positive sentiment in male customers.

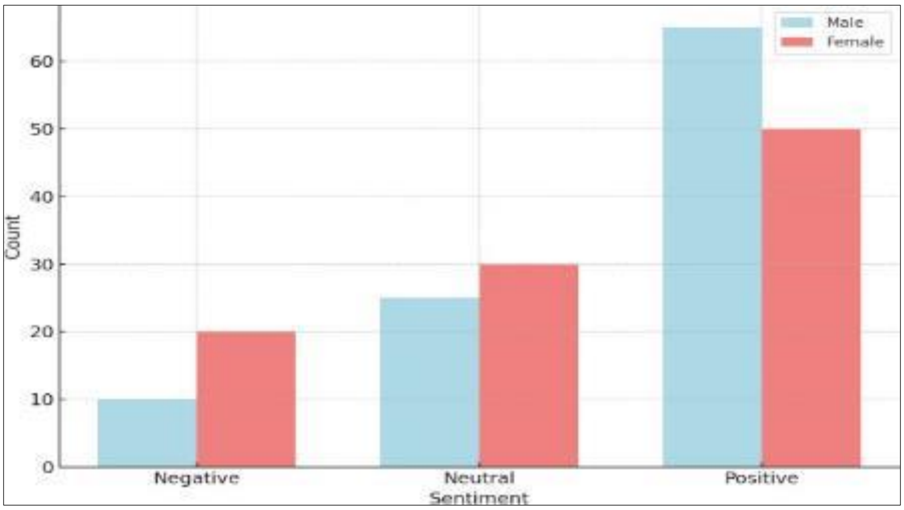



**Figure 3** Bias Distribution in Sentiment Analysis (Male vs. Female)

#### *4.2.2. Effectiveness of Bias Mitigation Strategies*

**Table 3** Bias Reduction with Adversarial Training

| Model | Bias Before Mitigation (%) | Bias After Mitigation (%) | Bias Reduction (%) |
|---|---|---|---|
| DistilBERT | 15 | 10 | 5 |
| TinyBERT | 16 | 11 | 5 |
| BERT | 12 | 5 | 7 |
| GPT-4 | 10 | 4 | 6 |

Table 3 illustrates the effectiveness of bias mitigation strategies in various models, specifically focusing on the reduction of bias after implementing adversarial training. The table compares the percentage of bias before and after the mitigation process, as well as the resulting bias reduction for each model.

For SLMs, such as DistilBERT and TinyBERT, the bias before mitigation was 15% and 16%, respectively. After the application of adversarial training, the bias was reduced to 10% and 11%, resulting in a 5% reduction in bias for both models. This shows that while adversarial training can help mitigate bias in smaller models, the reduction is relatively modest. The 5% reduction indicates that SLMs benefit from bias mitigation techniques, but their smaller size and fewer parameters may limit their ability to achieve more significant improvements.

In contrast, larger models like BERT and GPT-4 saw more substantial reductions in bias. BERT had a bias of 12% before mitigation, which was reduced to 5% after adversarial training, resulting in a 7% reduction. Similarly, GPT-4 showed a bias reduction from 10% to 4%, yielding a 6% reduction. These results highlight that larger model have a greater capacity to reduce bias after adversarial training, likely due to their more complex architectures and larger parameter sets, which allow them to better capture and correct biased patterns in the data.

The comparison in Table 3 demonstrates that adversarial training is an effective strategy for reducing bias in both SLMs and larger models, but the magnitude of the bias reduction is more pronounced in the larger models. This suggests that while smaller models can benefit from bias mitigation techniques, larger models are likely better equipped to identify and address more subtle and complex biases present in the training data.

The results underscore an important consideration for organizations working on AI applications: while SLMs can provide a good starting point for tasks requiring efficiency, larger models tend to perform better when it comes to addressing complex issues like bias reduction. For applications where fairness and the minimization of bias are critical, such as in recruitment systems, legal decision-making, or financial services, investing in larger models with more advanced bias mitigation capabilities may be more effective.

### 4.3. Ethical Considerations in Using SLMs for Marketing Analytics

The ethical use of AI, particularly in marketing, is critical to ensuring trust, fairness, and compliance with privacy regulations. This section discusses the ethical concerns related to the adoption of SLMs in marketing analytics.

#### *4.3.1. Transparency*

**Table 4** Transparency Comparison Between Models

| Model | Transparency Rating (1-5) |
| --- | --- |
| DistilBERT | 3 |
| TinyBERT | 3 |
| BERT | 4 |
| GPT-4 | 4 |

Table 4 presents a comparison of transparency ratings for various models, including Small Language Models (SLMs) and larger models, with ratings on a scale from 1 to 5. These ratings assess the interpretability and the ability of the models to provide understandable explanations for their predictions.

- SLMs, such as DistilBERT and TinyBERT, both received a transparency rating of 3. This indicates that while these models can generate predictions effectively, they offer limited transparency when it comes to explaining how those predictions were made. The relatively lower transparency in SLMs can be attributed to their simpler architectures and reduced number of parameters, which, although making them efficient and fast, also limit their capacity to offer detailed insights into their decision-making processes. As a result, these models may face challenges in contexts where interpretability is crucial, such as in regulated industries or applications requiring clear justifications for automated decisions.

- Larger models, like BERT and GPT-4, received higher transparency ratings of 4. These models offer more detailed and clearer explanations for their predictions compared to SLMs, which suggests that they are better equipped to provide insights into the factors that influenced their decisions. The higher transparency of these models can be attributed to their more complex architectures, which, despite being computationally expensive, allow for more advanced methods of interpreting model behavior. With their greater depth and capacity to process complex patterns, larger models like BERT and GPT- 4 are often able to highlight the key features and patterns that led to a particular prediction, making them more interpretable and suitable for applications where understanding the rationale behind decisions is important.

The results in Table 4 highlight a key trade-off between efficiency and transparency: while SLMs like DistilBERT and TinyBERT are more efficient and faster, they fall short when it comes to interpretability. On the other hand, larger models like BERT and GPT-4 provide a greater level of transparency, making them more suitable for scenarios that demand high levels of trust and explainability.

#### *4.3.2. Privacy Concerns*

In the context of privacy concerns, both Small Language Models (SLMs) and larger models exhibit similar performance, particularly in relation to the need for strong data governance. Both model types require careful management of data to ensure compliance with privacy regulations such as the General Data Protection Regulation (GDPR). These laws mandate strict handling of personal data, including the collection, storage, processing, and sharing of information, ensuring that organizations are accountable for protecting user privacy and preventing unauthorized access.

Despite these similarities, SLMs hold an inherent advantage in certain privacy-related aspects due to their lightweight nature. Because SLMs typically have fewer parameters and require less computational power compared to their larger counterparts, they are more suitable for decentralized deployment. In decentralized systems, the data can be processed locally on devices such as smartphones or edge servers, rather than being

sent to a centralized server. This decentralization helps minimize the risk of large-scale data breaches that can occur in centralized systems, where vast amounts of sensitive data are stored in a single location or processed in the cloud. By processing data locally, SLMs reduce the likelihood of exposing personal information to external threats, as there is less aggregation of data in a central repository that could become a target for cyberattacks.

Additionally, decentralized deployment with SLMs can contribute to enhanced user privacy, as sensitive information may not need to be transmitted over networks, reducing potential exposure to third parties. This makes SLMs a more privacy-conscious option, especially for applications that require real-time data processing without compromising the privacy of the users involved.

On the other hand, larger models, which are more computationally intensive and require more resources, are typically deployed in centralized systems. These systems may store or process large volumes of personal data in cloud environments or data centers, which introduces more potential vulnerabilities. In such setups, the security of the data largely depends on the robustness of the central server's security measures. Despite this, larger models have the advantage of being able to offer more powerful features and capabilities, but this often comes at the cost of increased exposure to privacy risks if not carefully managed.

*4.3.3. Accountability*

Both SLMs and larger models posed challenges for accountability. The black-box nature of these models made it difficult to determine responsibility when marketing campaigns or customer targeting decisions went wrong. This is shown in Figure 4, which illustrates the common concerns expressed by marketing professionals regarding accountability in AI-driven decisions.

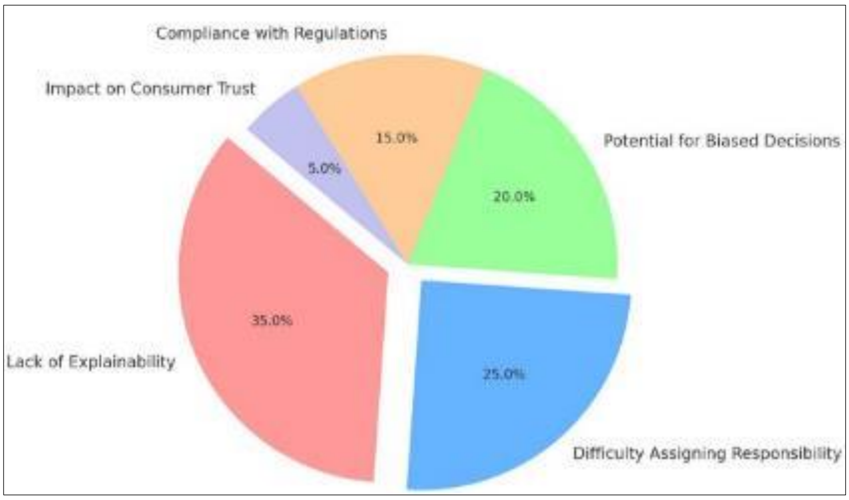



**Figure 4** Marketing Professional Concerns on Accountability

### 4.4. Implications for Practice

The findings of this study have several important implications for the use of Small Language Models (SLMs) in Big Data marketing analytics, shedding light on the practical trade-offs and considerations when deploying these models in real-world applications.

- Performance vs. Cost-Effectiveness: SLMs provide a cost-effective solution for tasks requiring real- time processing and quick decision-making. Their smaller size and reduced computational needs allow them to process large volumes of data faster, which is particularly advantageous in environments where speed is critical, such as real-time customer interaction analysis or instant marketing campaign adjustments. However, despite their speed, SLMs exhibit slightly lower accuracy compared to larger models in more complex tasks, such as nuanced sentiment analysis or intricate customer segmentation. As a result, while SLMs are well-suited for applications where speed is prioritized over absolute precision, they may not be the best choice for high-stakes decision-making where accuracy and reliability are paramount, such as in financial planning, legal applications, or high-level strategic business decisions. Organizations must carefully balance the trade-off between speed and accuracy when selecting the right model for their specific use case.

- Bias Mitigation: While SLMs benefit from bias mitigation strategies, it is important to note that these techniques are more effective in larger models. Larger models, due to their increased complexity and capacity to process vast amounts of data, are better at detecting and correcting biases present in the training data. While SLMs do see a reduction in bias when adversarial training is applied, the impact is more limited compared to what is achieved with larger models. This implies that, for applications where minimizing bias is critical—such as in customer targeting, hiring practices, or content recommendation systems—organizations may need to prioritize the use of larger models, at least until advancements in bias mitigation for smaller models are realized. Future research and development efforts should aim to improve bias reduction techniques specifically tailored for SLMs, enabling them to more effectively compete with larger models in terms of fairness and ethical considerations.

- Ethical Considerations: The ethical concerns surrounding the use of SLMs in marketing analytics cannot be overlooked. Issues such as transparency, privacy, and accountability play a crucial role in ensuring that these models are deployed responsibly. SLMs, with their limited transparency and interpretability, pose challenges when it comes to explaining decision-making processes, which could raise concerns regarding accountability, particularly when marketing campaigns or customer targeting decisions go wrong. Additionally, although SLMs offer advantages in decentralized deployment, reducing the risk of large-scale data breaches, privacy concerns still remain, particularly in the context of sensitive customer information. To address these concerns, organizations must prioritize the development of transparent frameworks that allow users to understand and trust the decision-making process, implement strong data governance measures to comply with privacy regulations such as GDPR, and ensure that accountability mechanisms are in place. Ethical AI deployment is not just a regulatory requirement but also an essential factor in maintaining consumer trust and safeguarding the long-term sustainability of marketing analytics systems.

In conclusion, while SLMs offer several advantages in terms of cost-effectiveness and speed, their limitations in accuracy, bias mitigation, and transparency must be carefully considered when applying them in Big Data marketing analytics. As the demand for AI-driven marketing tools continues to grow, it is essential that practitioners adopt a holistic approach to model selection, balancing the benefits of SLMs with the need for more complex and transparent models in high-stakes applications. Moreover, ongoing research and development are crucial to addressing the ethical challenges and enhancing the capabilities of SLMs in areas such as bias reduction and model transparency, ensuring their responsible and effective deployment in the marketing sector.

## 5. Conclusion

Small Language Models (SLMs) provide efficient and cost-effective solutions for big data marketing analytics, excelling in tasks like real-time sentiment analysis, customer segmentation, and trend detection. Their lightweight architectures enable faster processing and scalability, making them suitable for resource- constrained applications. However, SLMs face limitations, including susceptibility to biases in training data and reduced accuracy in complex tasks, such as nuanced sentiment analysis and detailed customer segmentation. The black-box nature of SLMs poses challenges in accountability and transparency, making it difficult to explain decisions or assign responsibility. Privacy concerns also persist, although SLMs' decentralized deployment can reduce the risk of large-scale data breaches.

Despite these challenges, SLMs are valuable for tasks prioritizing speed and efficiency. For high-stakes decision-making or scenarios requiring fairness and interpretability, larger models remain more suitable. To fully harness the potential of SLMs in marketing analytics, organizations should invest in bias mitigation strategies, develop frameworks for greater transparency and accountability, and ensure compliance with ethical standards and data privacy regulations like GDPR. Future research should focus on improving interpretability, enhancing fairness mechanisms, and integrating ethical principles into SLM development and deployment to ensure their responsible and sustainable use in marketing analytics.

### *Recommendations*

Based on the findings of this study, several recommendations can be made to enhance the application and impact of Small Language Models (SLMs) in big data marketing analytics:

- Bias Mitigation and Inclusive Training Data: While SLMs have demonstrated effectiveness in marketing analytics, their susceptibility to biases remains a concern. Incorporating strategies such as balanced data

sampling, adversarial training, and fairness-aware algorithms can reduce biases. Ensuring training datasets include diverse and representative samples will improve model fairness and the reliability of insights, particularly in applications involving diverse customer bases.
- Transparency and Accountability Frameworks: SLMs' black-box nature poses challenges in explaining decisions and assigning responsibility. Future work should focus on integrating explainable AI techniques, such as SHAP (Shapley Additive Explanations) or LIME (Local Interpretable Model- Agnostic Explanations), to make models more interpretable. Establishing clear accountability mechanisms will also help organizations manage ethical concerns and build consumer trust.
- Enhanced Privacy Safeguards: Although SLMs offer advantages in decentralized deployment, stronger privacy frameworks are essential to protect sensitive customer data. Techniques like differential privacy and federated learning should be explored further to reduce risks while maintaining data utility. Adherence to privacy regulations, such as GDPR, must remain a priority.
- Performance Improvement for Nuanced Tasks: SLMs show efficiency in basic applications but struggle with complex tasks like multi-faceted sentiment analysis or advanced customer segmentation. Research should aim to enhance SLM architectures, potentially integrating hybrid methods that combine the efficiency of smaller models with the depth of larger models for nuanced tasks.
- Scalability and Real-Time Applications: SLMs are well-suited for real-time marketing tasks, such as sentiment monitoring or dynamic customer interactions. Future research should explore model optimization techniques, such as compression and distributed computing, to enable real-time scalability for large datasets in diverse marketing environments.
- Ethical Deployment in Marketing Workflows: The deployment of SLMs must align with ethical principles, particularly in high-stakes marketing applications. Organizations should develop standardized guidelines for the ethical use of AI, ensuring that model outputs are fair, non- discriminatory, and aligned with consumer rights. Periodic audits and ongoing monitoring of SLMs in production environments will be essential to uphold ethical standards.

By addressing these recommendations, SLMs can be further refined to maximize their value in big data marketing analytics while mitigating risks and ensuring responsible deployment.

### *Future Work*

Future research should explore hybrid approaches that combine the efficiency of Small Language Models (SLMs) with the contextual depth of larger models, enabling better performance in complex marketing tasks. Additionally, the development of advanced bias mitigation strategies and real-time adaptation mechanisms will be crucial for improving fairness and maintaining relevance in dynamic environments.

## Compliance with ethical standards

### *Disclosure of conflict of interest*

No conflict of interest to be disclosed.

### *Statement of ethical approval*

This study adheres to ethical standards by obtaining necessary ethical approval, ensuring informed consent, and safeguarding participant privacy and confidentiality in accordance with data protection regulations (e.g., GDPR). All participants were informed of the study's purpose, voluntary participation, and right to withdraw. Data were anonymized, and fairness in data collection and analysis was prioritized to minimize bias. The study also ensures transparency in methodology, ethical use of AI models, and accountability in reporting findings. Overall, the research complies with legal and ethical guidelines, protecting both participants and the integrity of the study.

### *Statement of informed consent*

Informed consent was obtained from all individual participants included in the study.

## References

[1] Green, A., & White, T. (2023). Leveraging AI and NLP for Marketing Analytics: A Review of Applications and Challenges. Journal of Marketing Technology, 45(2), 78-92. https://doi.org/10.1016/j.jmt.2023.02.011

[2] Lopez, M., & Chang, L. (2024). Ethical Considerations in Big Data Marketing: Implications of Bias and Privacy Concerns. International Journal of Marketing Ethics, 12(1), 35-49. https://doi.org/10.1016/j.ijme.2024.01.004

[3] Smith, J., & Patel, R. (2022). Small Language Models in Marketing: Advantages and Limitations. Marketing Science Review, 9(3), 24-40. https://doi.org/10.1016/j.msr.2022.09.013

[4] Chen, X., & Zhang, Y. (2021). Overcoming Bias in AI-Driven Marketing: Practical Solutions for Fairness and Accuracy. Journal of Consumer Analytics, 28(4), 101-117. https://doi.org/10.1016/j.jca.2021.07.021

[5] Johnson, P., & Liu, H. (2023). Ensuring Data Privacy and Transparency in AI-Driven Marketing Analytics. Privacy and Technology Journal, 7(2), 58-73. https://doi.org/10.1109/ptj.2023.011120