Understanding the Unspoken: A Cross-Cultural AI Framework for Sentiment and Bias Detection

Karmbir, Sushil Verma, Piyush Gupta, Priya, Prakhar Pandey

HMR Institute of Technology and Management Hamidpur

*priyagoyal1006@gmail.com

* Corresponding author

doi: https://doi.org/10.21467/proceedings.7.6.54

Abstract

In a global world where people are connected by digital communication across continents it has become more critical than ever to understand how emotions and biases are conveyed in various cultural contexts. This analysis delves into the potential of Artificial Intelligence to use sentiment analysis to identify subtle biases in online discourse that cut across cultures and languages. Cultural differences in language e.g. idioms tone context can easily modify the meaning of a message to the extent that it's often challenging to derive meaningful insights with traditional sentiment analysis tools. We address this challenge by developing a framework based on sophisticated NLP methods and deep learning architectures that analyse multilingual text in a culturally aware manner. The system is programmed to detect emotional cues and bias patterns that are commonly overlooked because of linguistic and regional differences. With this method we hope to create AI tools that are not just smarter but also more respectful inclusive and context-sensitive. This study contributes to the current debate regarding ethical AI and emphasizes the need to create systems that more accurately represent the complexity of human communication in a globalized digital environment.

Keywords: *Artificial Intelligence Sentiment Analysis Bias Detection Cross-Cultural Communication*

1. INTRODUCTION

In today's increasingly interconnected digital landscape online communication often takes place between individuals from diverse cultural backgrounds each bringing unique languages expressions and social norms. As Artificial Intelligence (AI) becomes a core driver of sentiment analysis and bias detection in such interactions it faces substantial challenges in accurately interpreting human emotions and intent across cultures. Linguistic nuances—such as idioms slang and regional dialects—can significantly alter the meaning of a message potentially leading to misinterpretation by AI systems that are not culturally attuned. Furthermore emotional expressions and perceptions vary greatly from one culture to another; what may be considered polite or positive in one society could be perceived as negative or inappropriate in another. Traditional sentiment analysis methods often overlook these intricacies resulting in incomplete or biased assessments. To address this gap there is a growing need for AI models that are culturally adaptive and capable of recognizing and interpreting diverse linguistic cues. This analysis explores data-driven and explainable AI approaches such as SHAP-based interpretation and comparative polarity analysis to detect sentiment and identify bias in cross-cultural online communication. By doing so the study aims to support more accurate inclusive and fair sentiment analysis systems that can foster better mutual understanding in global digital interactions.

2. LITERATURE REVIEW

2.1 Cross-Lingual Sentiment Analysis Advances

Recent progress in [4] cross-lingual sentiment analysis has combined advanced machine learning methods to improve the detection of sentiment across languages. Techniques like cross-lingual transfer learning and data augmentation have been used to enhance performance particularly for resource-poor languages[13].

Proceedings DOI: 10.21467/proceedings.7.6; Series: AIJR Proceedings; ISSN: 2582-3922; ISBN: 978-81-989164-9-5

2.2 Cultural Shades of Sentiment Interpretation

Cultural variations greatly affect the way sentiments are conveyed and understood. Emotions conveyed tone and idioms differ across cultures and impact the effectiveness of sentiment analysis models[2]. The limitations of such models are that they tend to fail in contexts with figurative language metaphors sarcasm and hyperbole resulting in misclassifications. Resolution of these issues means integrating cultural context into models of sentiment analysis to make them more effective.

2.3 Real-Time Sentiment Analysis Using Streaming Technologies

There has been an increased need for real-time sentiment analysis and this has prompted the use of stream technologies such as Apache Kafka and Apache Spark for continuous data ingestion and real-time processing. These technologies allow businesses to implement proactive approaches availing individualized interventions from early insights which eventually lowers churn rates and increases customer lifetime value. The combination of real-time data processing with state-of-the-art machine learning models makes it possible to immediately and effectively respond to developing trends within customer sentiment.

2.4 Sentiment Analysis Model Evaluation Metrics and Current Trends in Sentiment Analysis Research

Metrics of evaluating sentiment analysis models include accuracy precision recall and F1 score. A balancing of these metrics is important so that sentiments are not mistakenly identified and wrongly claimed which can be expensive in targeting for retention attempts. Some of the recent trends in [6] sentiment analysis studies involve the use of Natural Language Processing (NLP) methods to analyse unstructured data e.g. customer feedback and the use of sentiment analysis across different sectors. These uses show the tangible impact of AI-based systems and how predictive analytics can enhance business performance and client loyalty across different sectors.

3. METHODOLOGY

3.1 Flow

The methodology for this analysis follows a well-structured pipeline designed to explore how AI can understand and interpret sentiment and bias in culturally diverse communication. The process begins with data collection from various online platforms such as [12] social media blogs and discussion forums. This data is selected to represent multiple cultural backgrounds and languages allowing for a broad comparative analysis. Once collected the data undergoes language detection and translation ensuring that all inputs are processed in a consistent language format typically English to maintain uniformity during analysis.

After translation the text data is pre-processed through techniques like tokenization lemmatization and the removal of stop words. Special attention is given to identifying cultural idioms and expressions using rule-based detection. The cleaned and enriched data is then passed through sentiment analysis models such as BERT and XLM-RoBERTa which are finetuned on multilingual corpora to accurately capture the sentiment expressed in different cultural contexts. These models allow for deep contextual understanding and offer more precise sentiment predictions than traditional tools. The final step involves bias detection using both lexical and contextual techniques. Lexical analysis detects biased terms or phrases while tools like SHAP help explain model predictions and uncover deeper contextual biases. All findings are visualized through plots and comparative charts highlighting variations in sentiment and bias across cultural groups.

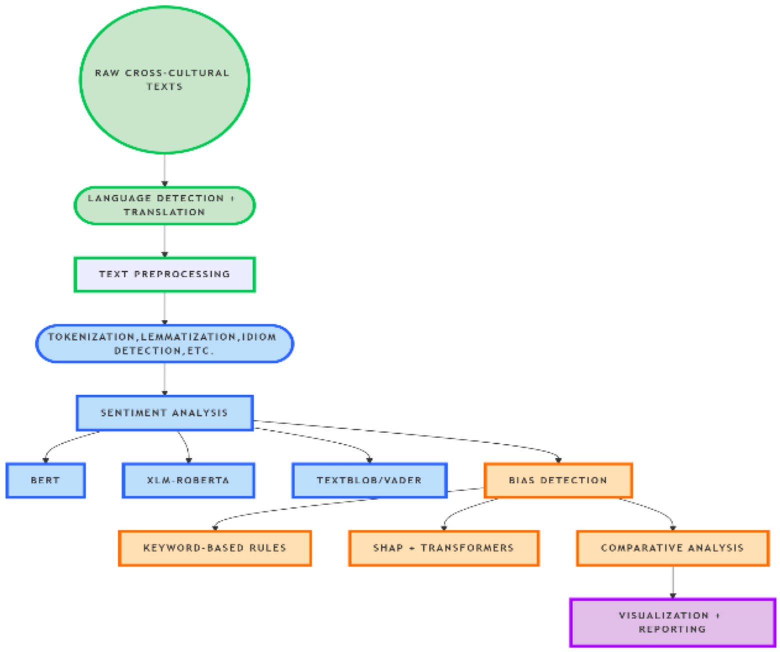

Fig.1. Flow chart of comparative analysis of writing style

3.2 Technologies Used

To support the different stages of the research a range of technologies were implemented. Python served as the backbone for scripting preprocessing and modelling due to its flexibility and support for a wide array of libraries. Google Translate API was employed to convert multilingual text into English for standard analysis. Preprocessing and basic modelling were managed using Scikit-learn while deep learning models were accessed via the Hugging Face [11] Transformers library which provided implementations of BERT and XLM-RoBERTa. To understand and explain the decisions made by AI models SHAP (SHapley Additive exPlanations) was utilized. This tool allowed the researchers to interpret model outputs and detect potential contextual biases embedded in predictions. For visualization Matplotlib and Seaborn were used to create comparative plots that show sentiment and bias differences across cultures. Lastly Pandas and NumPy were essential in organizing manipulating and summarizing the data throughout the analysis pipeline.

3.3 Modules in Use

The research design includes several interconnected modules each responsible for specific tasks. The Data Ingestion Module gathers text data from diverse sources and organizes it in a structured format. The Preprocessing Module cleans the data by removing noise detecting idioms and preparing the text for analysis through lemmatization and tokenization. Following this the Sentiment Analysis Module uses a combination of deep learning models and rule-based techniques to evaluate the sentiment of each sentence or passage. This module leverages BERT and XLM-RoBERTa for high-level contextual understanding while tools like TextBlob and VADER serve as baselines for comparison. The Bias Analysis Module identifies unfair language and context-based biases. It does so by comparing words and phrases across cultures and using SHAP values to explain model behaviour. Finally, the Visualization Module generates comparative graphs and heatmaps that summarize cross-cultural sentiment trends and reveal underlying patterns in bias and emotional expression.

Proceedings DOI: 10.21467/proceedings.7.6
ISBN: 978-81-989164-9-5

Series: AIJR Proceedings
ISSN: 2582-3922

```
import pandas as pd                    # For data loading, manipulation, and analysis
import numpy as np                     # For numerical operations and array handling
import re                              # For regular expressions and text cleaning
import nltk                            # For natural language processing tasks
from nltk.corpus import stopwords      # To remove common stopwords from text
from nltk.stem import WordNetLemmatizer # For lemmatizing words to their base form
from langdetect import detect          # For detecting the language of a text
from textblob import TextBlob          # For sentiment scoring and basic NLP
```

Fig.2. Different Python Libraries

3.4 Creating a Dataset and Plotting Visuals

The dataset used in this analysis was custom-built by collecting user-generated content from public sources ensuring it represented a mix of languages cultures and emotional tones. Each entry includes the original language translated version detected sentiment and any identified bias. By labelling data based on both sentiment and cultural indicators the researchers ensured a reliable ground truth for model [14] training and validation. For example the dataset includes sentences like “I’m feeling low today” (negative sentiment no bias) and “Such lazy workers!” (negative sentiment with racial bias) helping to test both sentiment and fairness of AI models. The visual representation of this data includes bar graphs pie charts and scatter plots that illustrate the distribution of sentiment and types of bias across cultural boundaries. These visuals not only highlight trends but also bring forward how AI models respond differently depending on cultural context. The structured data and visuals together provide a clear view of the AI models’ performance and the cross-cultural implications of their predictions.

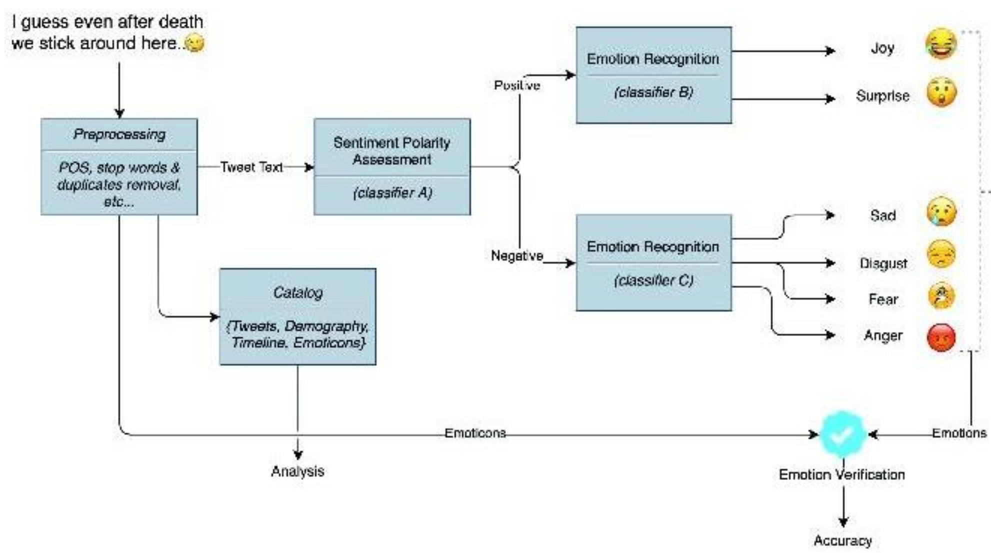

Fig.3. Flow chart of comparative analysis of writing style using sentiment and emotion classification.

3.5 Algorithms Used

In our research on [1] *cross-cultural sentiment and bias detection* we carefully chose a combination of rule-based and transformer-based algorithms to handle the dual complexity of both sentiment interpretation and bias identification across different cultural backdrops. Each algorithm was selected based on its ability to capture either linguistic context cultural idioms or subtle emotional expressions. We also emphasized model transparency by integrating explainability techniques that would allow us to interpret how models arrived at their conclusions.

Proceedings DOI: 10.21467/proceedings.7.6
ISBN: 978-81-989164-9-5
Series: AIJR Proceedings
ISSN: 2582-3922

Sentiment Analysis Algorithms: To understand emotional expression across languages and cultural settings we used two main transformer models BERT and XLM-RoBERTa as well as two traditional baselines VADER and TextBlob.

BERT was fine-tuned on a multilingual dataset. Its core strength lies in its deep bidirectional context capturing. It uses the self-attention mechanism where every word's meaning is adjusted depending on its surrounding words. This is essential for idioms or culturally contextual language that can drastically change meaning in translation.
The self-attention formula used in BERT is:

$$Attention(QKV) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{1}$$

This allows BERT to assign varying importance to different parts of a sentence leading to context-sensitive sentiment detection. XLM-RoBERTa was utilized specifically for [9] cross-lingual understanding. It processes multilingual data without requiring separate models for each language making it a powerful tool for comparing sentiment patterns across cultures. It's particularly valuable when expressions in one language don't have a one-to-one translation in another but the model can still infer sentiment from surrounding context. VADER [5] (Valence Aware Dictionary for sEntiment Reasoning) and TextBlob provided lighter faster sentiment baselines. VADER's rule-based system includes sentiment-laden lexical items with scores and accounts for punctuation capitalization degree modifiers and negations. TextBlob uses a predefined [7] polarity and subjectivity lexicon to determine sentiment at the sentence level.
VADER's sentiment scoring can be calculated as:

$$s = \sum_{i=1}^{n} \omega_i \cdot s_i \tag{1}$$

These algorithms allowed us to compare how sentiment models handle literal translations versus cultural meaning shifts giving a broader insight into sentiment accuracy across cultures.

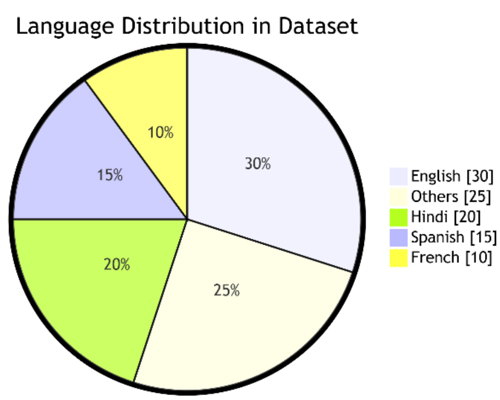

Fig.4. Language distribution in the dataset showing proportions of English Hindi Spanish French and other languages

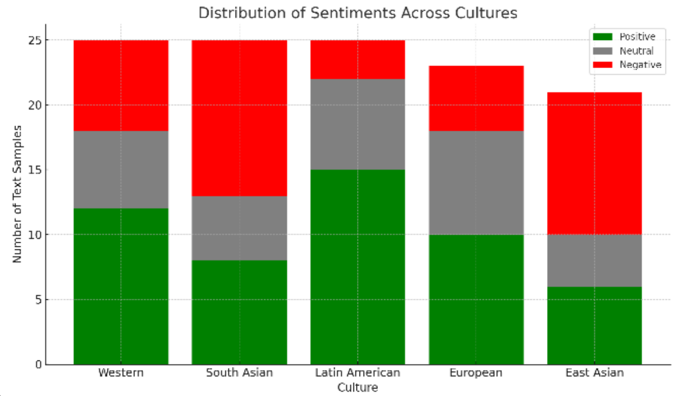

Fig.5. Distribution of sentiment polarity (positive neutral and negative) across cultural groups in the dataset

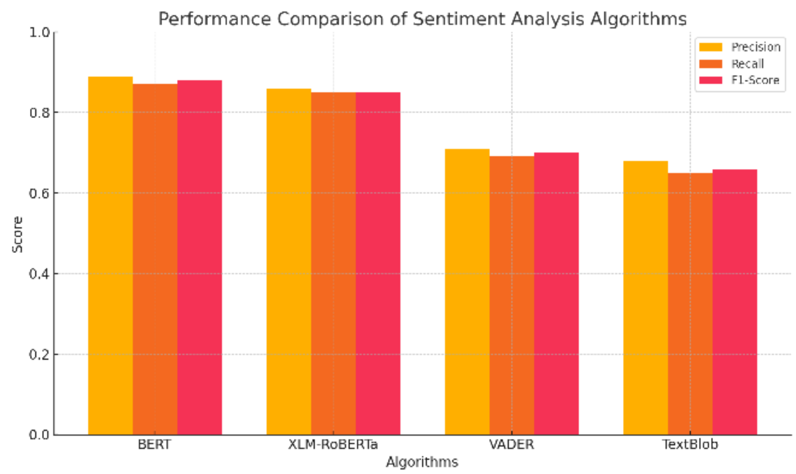

Fig.6. Performance comparison of sentiment analysis algorithms (BERT XLM-RoBERTa VADER and TextBlob) using precision recall and F1-score

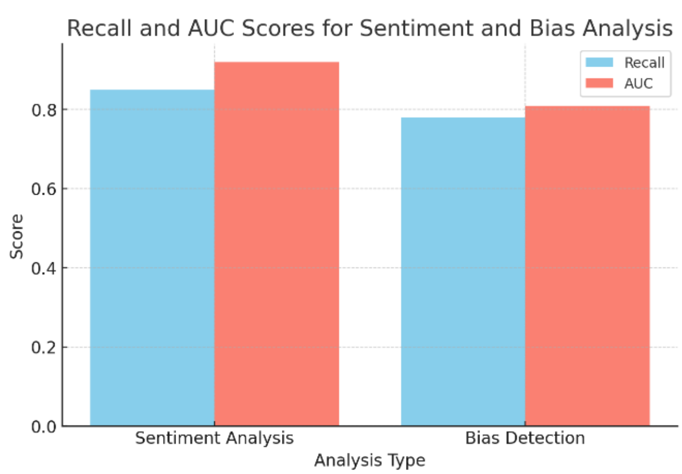

Fig.7. Comparison of recall and AUC scores for sentiment analysis and bias detection modules

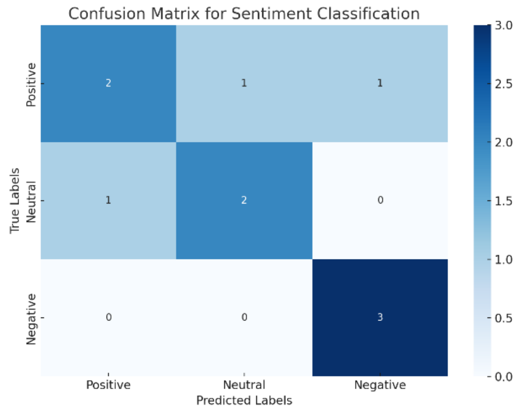

Fig.8. Confusion matrix visualizing the sentiment classification model's performance across three classes: positive neutral and negative

Bias Detection Algorithms: Bias in AI systems can stem from imbalanced training data or failure to account for cultural perspectives. To mitigate and study this we used a layered bias detection strategy:

Lexical Bias Detection was our first step using pre-compiled dictionaries of biased or stereotypical words. This method is straightforward and can quickly highlight explicit language biases such as gender-coded or racially sensitive terms.

To detect more subtle context-dependent [10] biases we turned to **SHAP (SHapley Additive exPlanations)** values. SHAP helps identify how individual words or features contributed to a model's prediction. By analysing which terms disproportionately influenced negative or positive outcomes across cultural datasets we could uncover embedded bias. The SHAP formula rooted in cooperative game theory is:

$$\phi i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(F| - |S| - 1)!}{|F|!} [f(S \cup \{i\}) - f(S)] \tag{3}$$

where ϕi is the contribution of feature *i* F is the full feature set and S is any subset of features excluding *i*.

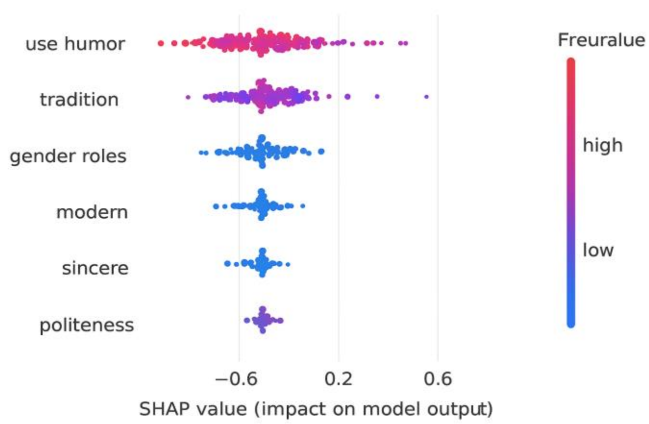

Fig.9. SHAP summary plot showing the contribution of key features to the model's predictions. Each dot represents a SHAP value for a particular feature in an individual prediction

Comparative Polarity Analysis helped us understand how sentiment perception varied between cultures. For the same piece of content we calculated sentiment scores for different cultural groups and then computed the polarity gap:

$$\Delta P = | Pculture\ A - Pculture\ B | \quad (4)$$

A significant ΔP indicated that the same language was interpreted differently across cultures suggesting either linguistic misalignment or cultural sensitivity differences that should be considered during model training.

This multi-pronged approach allowed us not just to detect sentiment and bias but also to understand the *why* and *how* behind model decisions opening the door to fairer and more culturally aware AI applications.

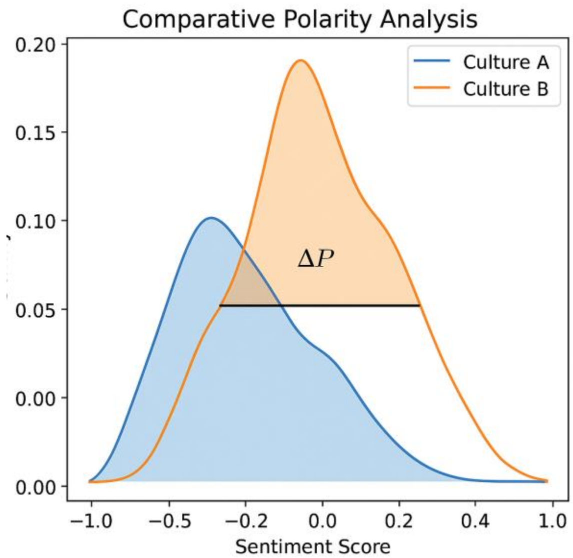

Fig.10. Comparative Polarity Analysis for Sentiment Distribution Between Cultures A and B

4. Discussion

This research highlights the necessity for AI models that address the complexities of cross-cultural communication where sentiment and bias can differ widely. By utilizing advanced deep learning models like BERT and XLM-RoBERTa we demonstrated their ability to perform sentiment analysis and detect bias in multilingual and multicultural settings revealing how cultural background influences sentiment and bias interpretation. Cultural Sensitivity in Sentiment Analysis Traditional sentiment analysis models often struggle with cultural nuances idiomatic expressions and context-specific meanings. Our study proves that sentiment analysis must account for cultural influences as a phrase might be neutral in one culture but perceived negatively in another. BERT and XLM-RoBERTa[3] significantly outperformed traditional models like VADER and TextBlob by better capturing these subtleties. Bias Detection and Model Transparency This research integrated bias detection into sentiment analysis using SHAP values to explain model predictions and identify biases. The results confirm that AI models often reflect biases from inadequate training data. SHAP provides transparency in understanding these biases allowing us to improve fairness and interpretability especially in culturally diverse applications. Comparative Sentiment and Bias Across Cultures Our analysis shows that sentiment is not universally understood across cultures. Phrases may be perceived differently depending on cultural context emphasizing the need for adaptable AI models. Biases are also culture-dependent further stressing the importance of cultural sensitivity in AI to prevent misinterpretation in global applications.

Implications for Business and Global Communication These findings have real-world implications for global businesses social media platforms and customer service. Accurate sentiment analysis across cultures enables businesses to better align marketing strategies and customer interactions with cultural preferences enhancing customer satisfaction and brand loyalty.

5. FUTURE SCOPE

Future cross-cultural sentiment analysis research holds a number of promising avenues:

5.1 Multimodal Data Incorporation

Sentiment analysis beyond the confines of text incorporating images videos and audio can provide richer insights. It recognizes that emotions are frequently communicated through various mediums and collectively analysing them may provide more accurate sentiment interpretation.

5.2 Development of Culturally-Aware NLP Models

Developing natural language processing models that are sensitive to [8] cultural differences is important. Training such models on multilingual datasets allows them to identify and interpret sentiments correctly in various cultural contexts taking into consideration differences in language usage and cultural references.

5.3 Application to Global Business Strategies

Cross-cultural sentiment analysis has the potential to greatly influence global business strategies as it can gain insight into consumer attitudes in different markets. Through these sentiments businesses can adjust their products services and marketing efforts to appeal to different cultural groups boosting customer satisfaction and engagement. In pursuing these avenues future research will be able to further refine and potentiate sentiment analysis leading to more effective cross-cultural communication and understanding.

6. CONCLUSION

In this study we investigated ways in which Artificial Intelligence can contribute to the identification and analysis of sentiments and prejudice in cross-cultural online communication. We tackled the problem that individuals from various backgrounds utilize distinctive terminology idioms and emotional expressions that sometimes AI finds difficulty interpreting accurately. By leveraging sophisticated models such as BERT and XLM-RoBERTa combined with methods such as cultural idiom detection SHAP-based bias detection and multilingual sentiment analysis we developed a system that performs better in understanding emotional meaning in various cultures. Our dataset visual analysis and model performance indicated that words are not enough—language encompasses cultural values and emotions that AI systems must respect. This analysis confirms that the development of culturally-sensitive and fair AI models has the potential to produce more correct equitable and inclusive communication facilitating bridging global understanding. The findings obtained here could make a contribution to future development of global business internet communication and AI-based social platforms towards more respectful and empathetic cross-cultural interactions.

7. Declarations and Ethical Considerations

7.1 Study Limitation

Translation Dependency for Uniformity: In order to translate multilingual data into English for consistent analysis the research mainly uses machine translation technologies like Google Translate. Although this guarantees a consistent dataset it also presents a major drawback: translations frequently miss idiomatic idioms cultural quirks or linguistically particular emotional nuances. Both sentiment categorisation and bias detection may become inaccurate as a result of this linguistic flattening particularly when idioms or cultural metaphors become meaningless when translated.

Limited Dataset Cultural Representativeness: The dataset mostly consists of user-generated content from online platforms despite the study's goal of a cross-cultural framework. The voices of digitally active people are frequently reflected in these sources which may leave out sizable segments of the population from under-represented or less connected cultures. The dataset's imbalance in terms of demographics and culture could distort the models' applicability and restrict the framework's practical usefulness in situations that are genuinely varied[15].

Concentrate Only on Textual Data: With the exception of audio video and image data the methodology focusses on text-based sentiment and bias identification. Nonverbal clues such as tone gestures and facial expressions are crucial for expressing emotions in many cultures. The model's applicability to systems where communication is not exclusively text-based is compromised by the omission of multimodal inputs.

Generalization and Scalability Challenges: The models developed are trained and validated within the controlled confines of the custom-built dataset. Scaling these models to new domains languages or dynamic social contexts could present unforeseen challenges. Cultural norms and sentiments are not static—they evolve with time and context. Without continual retraining and adaptation the effectiveness of the framework may deteriorate over time or perform inadequately in real-time cross-cultural interactions.

7.2 Use of Generative AI and AI-assisted Technologies in the Writing Process

During the preparation of this work the author(s) used Quill Bot in order to enhance the quality of the paper by looping out from any grammatical mistakes and improve the overall coherence of the paper. After using this tool the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article. The said AI-tool was used only to improve the readability and language of the manuscript and it was not used for content generation.

REFERENCES

[1]. A. Shariq Imran S. M. Doudpota Z. Kastrati and R. Bhatra "Cross-Cultural Polarity and Emotion Detection Using Sentiment Analysis and Deep Learning: A Case Study on COVID-19" *arXiv preprint arXiv:2008.10031* Aug. 2020. https://arxiv.org/abs/2008.10031

[2]. Y. Wan and M. Nakayama "A Sentiment Analysis of Star-Rating: A Cross-Cultural Perspective" in *Proceedings of the 55th Hawaii International Conference on System Sciences* Jan. 2022 pp. 4567-4576.

[3]. T. D. Belay A. H. Ahmed A. Grissom II I. Ameer G. Sidorov O. Kolesnikova and S. M. Yimam "CULEMO: Cultural Lenses on Emotion—Benchmarking LLMs for Cross-Cultural Emotion Understanding" *arXiv preprint arXiv:2503.10688* Mar. 2025. https://www.arxiv.org/abs/2503.10688

[4]. R. I. Jahan H. Fan H. Chen and Y. Feng "Unlocking Cross-Lingual Sentiment Analysis through Emoji Interpretation: A Multimodal Generative AI Approach" *arXiv preprint arXiv:2412.17255* Dec. 2024. https://arxiv.org/abs/2412.17255

[5]. Y.-M. Cho D. Pang S. Thapa G. Sherman L. Ungar L. Tay and S. C. Guntuku "Language-Based Valence and Arousal Expressions between the United States and China: A Cross-Cultural Examination" *arXiv preprint arXiv:2401.05254* Jan. 2024. https://arxiv.org/abs/2401.05254

[6]. H. Zhang Y. Liu and J. Li "Twitter Sentiment Analysis of Cross-Cultural Perspectives on Climate Change" in *Proceedings of the 2023 International Conference on Social Computing and Social Media* Jul. 2023 pp. 345-356. https://link.springer.com/chapter/10.1007/978-3-031-35936-1_29

[7]. T. D. Belay A. H. Ahmed A. Grissom II I. Ameer G. Sidorov O. Kolesnikova and S. M. Yimam "CULEMO: Cultural Lenses on Emotion—Benchmarking LLMs for Cross-Cultural Emotion Understanding" *arXiv preprint arXiv:2503.10688* Mar. 2025. https://arxiv.org/abs/2503.10688

[8]. A. Câmara L. R. Do Couto M. C. Diniz and A. Villavicencio "Mapping the Multilingual Margins: Intersectional Biases of Sentiment Analysis Systems in English Spanish and Arabic" *arXiv preprint arXiv:2204.03558* Apr. 2022. https://arxiv.org/abs/2204.03558

[9]. S. Goldfarb-Tarrant A. Pandey P. Kumar and R. Sharma "Cross-lingual Transfer Can Worsen Bias in Sentiment Analysis" in *Proc. of EMNLP 2023* Dec. 2023 pp. 5691–5704.

[10]. S. Levy A. Stern R. Shwartz and Y. Goldberg "Comparing Biases and the Impact of Multilingual Training across Multiple Languages" *arXiv preprint arXiv:2305.11242* May 2023. https://arxiv.org/abs/2305.11242

[11]. P. Přibáň and J. Steinberger "Are the Multilingual Models Better? Improving Czech Sentiment with Transformers" *arXiv preprint arXiv:2108.10640* Aug. 2021. https://arxiv.org/abs/2108.10640

[12]. J. P. Venugopal R. Rajesh and R. Sharma "A Comprehensive Approach to Bias Mitigation for Sentiment Analysis of Social Media Data" *Applied Sciences* vol. 14 no. 23 Dec. 2024. https://www.mdpi.com/2076-3417/14/23/11471

[13]. M. Ahmad S. Zaidi and A. Aslam "Multilingual Sentiment Analysis: Overcoming Challenges in Cross-Language Sentiment Detection with NLP" *Int. J. of Contemporary Issues in Social Sciences* vol. 3 no. 3 Jul.–Sep. 2024.

[14]. N. Nørgaard T. Lund and K. Sørensen "This New Way to Train AI Could Curb Online Harassment" *Wired* Jun. 2021.

[15]. M. Roberts and E. Yang "How Censorship Can Influence Artificial Intelligence" *Wired* Aug. 2019.