International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 04 | April - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

SENSYS: A MULTIMODEL SENTIMENT ANALYSIS ON SOCIAL MEDIA DATA

Rifa Fathima K K, Shilpamol S, Sidharth P Vinod, Vaishnav M, Anjana C V
Department of Computer Science and Engineering
College of Engineering Kidangoor, Kottayam, Kerala, India
Email: rifahftm@gmail.com, shilpashaji13@gmail.com,
sidharth.vinod1@outlook.com, vaishnav6544@gmail.com,
anjanachennothu@gmail.com

***Abstract*—Social media platforms generate vast amounts of multimodal content, including text, images, and videos, each conveying rich emotional information; however, most existing sentiment analysis methods focus primarily on text, overlooking critical context from visual and audiovisual data. This paper presents SENSYS (Sentiment Analysis System), a multimodal framework designed to analyze sentiment across text, image, and video content within a unified pipeline. The system operates on a simulated social media environment populated by AI-driven bot users, enabling scalable and privacy-preserving experimentation. SENSYS integrates BLIP for visual caption generation and the Groq-hosted LLaMA 3.3 70B model for sentiment classification, while a React.js-based dashboard provides real-time visualization of sentiment trends. Experimental evaluation demonstrates an overall accuracy of 88.4%, with text, image, and video achieving 91.2%, 87.6%, and 84.3%, respectively, ensuring reproducibility, scalability, and effective multimodal sentiment monitoring.**

***Index Terms*—Multimodal Sentiment Analysis, Simulated Social Media, Vision-Language Models, BLIP, LLaMA 3.3 70B, Zero-Shot Classification, Bot-Driven Data Generation, Cross-Modal Learning, Sentiment Visualization Dashboard**

I. INTRODUCTION

Social media platforms such as Twitter, Instagram, Reddit, YouTube, and Facebook have experienced rapid growth, transforming the internet into a continuously evolving repository of public opinion. Billions of user-generated posts are created daily, reflecting perspectives on products, politics, trends, and cultural phenomena. For organizations, researchers, and policymakers, the ability to automatically analyze this large-scale data has become essential for understanding public sentiment and supporting decision-making processes [2], [10].

Sentiment analysis focuses on identifying and extracting emotions, opinions, and attitudes from textual or multimedia data. Early approaches relied on lexicon-based methods such as SentiWordNet and VADER, which assign sentiment scores to words and aggregate them [2]. While computationally efficient, these methods struggle with linguistic complexities such as negation, sarcasm, and informal expressions commonly found in social media content [7].

The introduction of deep learning significantly improved sentiment analysis performance. Models such as Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) capture contextual dependencies in sequential data, while Convolutional Neural Networks (CNNs) effectively identify local patterns [7]. More recently, transformer-based architectures such as BERT have achieved state-of-the-art performance through large-scale pretraining and fine-tuning [3]. However, text-only approaches remain limited, as they fail to capture crucial contextual information embedded in images, videos, and other modalities.

To address these limitations, multimodal sentiment analysis has emerged as a promising direction by integrating textual, visual, and auditory information [1]. Research indicates that combining multiple modalities improves sentiment prediction accuracy compared to unimodal approaches [4], [5]. Nevertheless, several challenges persist, including the scarcity of labeled multimodal datasets, high computational requirements, and practical deployment constraints such as privacy concerns and platform restrictions [1], [5].

This paper introduces SENSYS (Sentiment Analysis System), a multimodal framework designed to address these challenges through a simulation-based approach. The proposed system constructs a synthetic social media environment populated by AI-driven agents that generate and interact with multimodal content, including text, images, and videos. Visual data is processed using BLIP for caption generation [11], while sentiment classification is performed using the Groq-hosted LLaMA 3.3 70B model [12]. The outputs are presented through a React.js-based dashboard that enables real-time sentiment visualization and analysis [6]. By operating in a simulated environment, SENSYS eliminates privacy constraints and enables controlled, reproducible experimentation.

Although synthetic environments may not fully capture the complexity of real-world social interactions, they provide a scalable and flexible platform for evaluating multimodal sentiment analysis systems and addressing deployment challenges in a controlled setting.

The key contributions of this work are as follows:

- A simulated social media platform with configurable AI bot users exhibiting emotion profiles, topic biases, and temporal behavioral patterns.
- A unified multimodal sentiment analysis pipeline integrating text, image, and video processing using a common LLM-based backend.
- Integration of BLIP for converting visual content into textual representations, effectively bridging computer vision

and natural language processing.
- A real-time React.js dashboard with dynamic visualization, filtering capabilities, and per-post sentiment analysis.
- Experimental evaluation demonstrating 88.4% overall accuracy on a dataset of 3,200 simulated multimodal social media posts.

The remainder of this paper is organized as follows. Section II presents the related work. Section III describes the system methodology and architecture. Section IV discusses implementation details and experimental results. Section V concludes the paper and outlines future research directions.

II. Related Work

A substantial body of research underpins the design of SENSYS. This section reviews prior work across five thematic areas.

A. Multimodal Fusion Architectures

Recent advances in multimodal sentiment analysis have focused on effective fusion strategies. Zhang *et al.* proposed the Multimodal Mixture of Low-Rank Experts (MMoLRE), which decomposes modality-specific representations into low-rank expert networks and integrates them through a learned gating mechanism [4]. This approach achieves state-of-the-art performance on benchmark datasets such as CMU-MOSEI, CMU-MOSI, and MELD, while maintaining parameter efficiency through low-rank factorization.

Similarly, Li *et al.* introduced the Hierarchical Representation Learning Framework (HRLF), which is designed to handle missing modalities during inference [5]. By constructing hierarchical modality-aware representations, the framework maintains robust performance even when certain modalities are absent, making it suitable for real-world deployment scenarios.

Ren proposed a multimodal framework combining textual embeddings from BERT with visual features extracted using ResNet, followed by an attention-based fusion mechanism [3]. This approach demonstrates improved performance over unimodal baselines on datasets such as Twitter, MVSA, and Flickr, highlighting the effectiveness of cross-modal attention in aligning semantic information across modalities.

B. Visual-Textual Sentiment Analysis

Al-Tameemi *et al.* presented a comprehensive survey on visual-textual sentiment analysis in social media, categorizing fusion strategies into early, late, and hybrid approaches [1]. The study highlights key challenges, including the scarcity of labeled multimodal datasets, domain adaptation issues, and the influence of cultural context on visual interpretation.

Kumar *et al.* provided a taxonomy of sentiment analysis techniques across aspect-level, sentence-level, and document-level tasks [2]. Their work emphasizes the growing importance of transformer-based models such as BERT and their variants, as well as the potential of large language models for transfer learning and zero-shot classification.

C. Deep Learning for Social Media

Deep learning models have significantly advanced sentiment analysis in social media contexts. Gothane *et al.* evaluated CNN, LSTM, and GRU models, demonstrating that hybrid CNN-LSTM architectures achieve superior performance by capturing both spatial and sequential features [7].

Benrouba and Boudour explored emotion-aware sentiment analysis using an SVM-LSTM hybrid model, focusing on mental health-related content in social media data [8]. Their approach improves the detection of emotionally sensitive content by incorporating nuanced emotional cues.

Hamed *et al.* proposed a CNN-LSTM-based framework for fake news detection, integrating sentiment analysis of news content with emotion analysis of user comments [9]. Their findings indicate that combining multiple signals enhances classification performance in detecting misinformation.

D. Sentiment Visualization and Explainability

Jain *et al.* developed a visualization framework that integrates attention mechanisms with LSTM and transformer-based models to improve interpretability of sentiment predictions [6]. This approach enables users to understand model decisions through visual explanations.

Suhaimin *et al.* conducted a comprehensive review of sentiment analysis applications in public security, identifying key challenges such as adversarial robustness and real-time scalability [10]. These challenges significantly influence the design considerations of modern sentiment analysis systems.

E. Research Gaps and Motivation

Despite significant progress, several limitations remain in existing works. First, most systems rely on real-world social media data, raising concerns related to privacy, reproducibility, and annotation effort. Second, few approaches effectively integrate all three modalities—text, image, and video—within a unified framework. Third, real-time interactive visualization of multimodal sentiment outputs is rarely addressed in research prototypes. Finally, the use of large language models, such as LLaMA 3.3 70B, for zero-shot multimodal sentiment classification remains underexplored [12].

To address these challenges, SENSYS proposes a simulation-based multimodal sentiment analysis framework that ensures privacy, scalability, and reproducibility while integrating advanced vision-language models such as BLIP [11].

III. Methodology

A. System Overview

SENSYS is designed as a six-stage pipeline: (1) content creation, (2) temporal dynamics and bot behavior, (3) content engagement, (4) data extraction, (5) multimodal sentiment analysis, and (6) visualization and alerting. The overall system architecture is illustrated in Fig. 1.

The system comprises two primary subsystems: the *Social Media Simulator* and the *Sentiment Analyzer*, interconnected through an API-driven data collection and storage layer. The

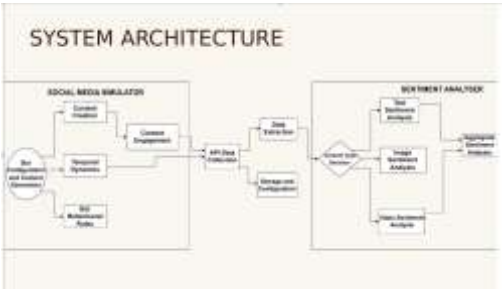

Fig. 1. SENSYS System Architecture

backend is implemented using Python 3.11 with FastAPI for RESTful services, while SQLite with SQLAlchemy ORM is used for data management. Multimodal inference is performed using PyTorch and Hugging Face Transformers for BLIP, and OpenCV is employed for video frame extraction. Sentiment classification is carried out using the Groq-hosted LLaMA 3.3 70B model. The frontend dashboard is developed using React.js with Chart.js for real-time data visualization.

B. Simulation Environment and Dataset

SENSYS utilizes a controlled simulation environment instead of real-world social media data. This design ensures (i) independence from third-party API limitations, (ii) availability of ground-truth sentiment labels for evaluation, and (iii) reproducibility of experiments.

The platform enables user registration, post creation (text, image, and video), and bot management. Each bot is characterized by three parameters: (i) an emotional profile (positive, negative, or neutral), (ii) a topic bias (e.g., technology, entertainment, politics), and (iii) an engagement frequency (active or passive).

Posts include metadata such as unique identifiers, author references, timestamps, hashtags, and engagement metrics (likes, dislikes, comments). Temporal patterns of post creation follow a non-uniform distribution to mimic real-world activity peaks during morning and evening hours.

C. Content Creation Module

Both users and bots generate multimodal posts consisting of text, images, and videos. Each post is associated with metadata, including timestamps, hashtags, and content type labels.

Text generation for bots is performed using the Groq-hosted LLaMA API, conditioned on the bot's emotional profile and topic bias. This ensures semantically coherent and sentiment-aligned content generation.

D. Temporal Dynamics and Bot Behavior

The temporal dynamics module simulates realistic user activity by distributing posts and interactions across time intervals. Each bot follows predefined behavioral rules based on its emotional profile, topic bias, and engagement frequency. This design introduces diversity and variability, enabling realistic simulation of social media ecosystems.

E. Content Engagement Module

The content engagement module manages interactions between users, bots, and posts. For text posts, responses are generated using the LLaMA API. For image posts, BLIP generates descriptive captions, which are then processed by the LLM. Video posts are decomposed into keyframes using OpenCV, and each frame is captioned using BLIP.

The generated captions or summaries are passed to the LLM for sentiment-aware responses. Additionally, reaction media such as GIFs are retrieved using external APIs based on predicted sentiment.

F. Data Extraction Module

This module acts as a bridge between the simulation environment and the sentiment analysis pipeline. It serializes posts along with their metadata and forwards them to the appropriate processing pipeline. Extracted data is stored in the SQLite database and made accessible to the frontend via FastAPI-based REST endpoints.

G. BLIP: Visual Content Understanding

BLIP serves as the vision-language backbone for processing visual content [11]. It is trained on large-scale image-text datasets using three objectives: Image-Text Contrastive (ITC), Image-Text Matching (ITM), and Language Modeling (LM). The architecture consists of an image encoder (ViT-based), an image-grounded text encoder, and a text decoder for caption generation. In SENSYS, BLIP converts visual inputs into textual descriptions, enabling unified text-based sentiment analysis.

For video inputs, OpenCV extracts keyframes at regular intervals (default: every 2 seconds). Each frame is captioned using BLIP, and the resulting captions are aggregated into a temporal summary for downstream processing.

H. LLaMA 3.3 70B: Sentiment Classification

Sentiment classification is performed using the Groq-hosted LLaMA 3.3 70B model [12]. This instruction-tuned large language model supports zero-shot and few-shot classification tasks with high accuracy.

A structured prompt is used to classify input text into positive, negative, or neutral categories, along with a reasoning trace. The output is formatted in JSON to ensure reliable parsing. The Groq API enables low-latency inference (typically under 500 ms per request).

I. Multimodal Sentiment Pipeline

The sentiment analysis pipeline dynamically routes inputs based on content type. Text posts are directly processed by the LLM. Image posts undergo caption generation using BLIP before classification. Video posts are processed through frame extraction, caption generation, and temporal summarization before sentiment classification.

All outputs are stored in the database and made available for visualization via API endpoints.

J. Visualization and Alert Module

A React.js-based dashboard provides real-time visualization of sentiment data. Key features include a live feed with sentiment labels, pie charts showing sentiment distribution, time-series graphs for temporal trends, and topic-wise sentiment analysis.

An alerting mechanism notifies users when negative sentiment for a specific topic exceeds a predefined threshold, enabling proactive monitoring.

K. Hardware and Software Requirements

The system requires a minimum hardware configuration of an Intel Core i5 processor, 8 GB RAM, 250 GB SSD storage, and a standard display resolution.

The software stack includes Windows 10/11 or Ubuntu 22.04, Python 3.11, Node.js 18+, React.js, FastAPI, SQLite with SQLAlchemy, PyTorch, Hugging Face Transformers, OpenCV, and modern web browsers such as Google Chrome or Microsoft Edge.

IV. Implementation Details

A. Development Timeline

SENSYS was developed in eight structured phases from December 2025 to March 2026. Phase 1 (Dec. 10) focused on problem formulation and literature review. Phase 2 (Dec. 20) addressed model and technology selection. Phase 3 (Jan. 5) involved backend development of the social media simulator. Phase 4 (Jan. 20) implemented bot user generation and content creation. Phase 5 (Feb. 5) integrated the multimodal sentiment analysis pipeline. Phase 6 (Feb. 15) focused on system optimization and performance evaluation. Phase 7 (Mar. 5) delivered the visualization dashboard and system integration. Finally, Phase 8 (Mar. 20) completed documentation and final validation.

B. Experimental Setup

The simulation environment consisted of 500 bot users, evenly distributed across three emotional profiles (positive, negative, neutral) and five topic domains (technology, product reviews, entertainment, politics, and lifestyle). Over a 14-day simulation period, 3200 posts were generated, including 1450 text posts, 770 image posts, and 980 video posts. Ground-truth sentiment labels were derived from bot configurations and generation prompts.

C. System Interface — Registration and Login

The registration and login system uses authentication managed through JWT tokens provided by the Python-JOSE library. For registration, a new user must enter their username, email, full name, and password. Once logged in, a session token is generated, allowing full interaction with the platform.

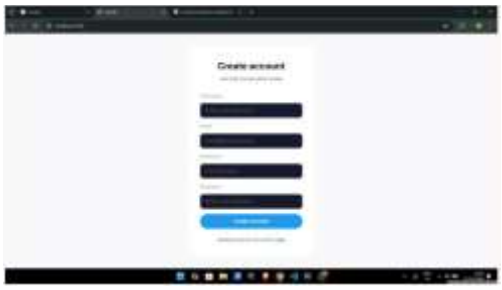

Fig. 2. Registration Screen

Fig. 3. Login Screen

D. System Interface — Homepage and User Profile

The homepage (Fig. 4) presents the live post feed, a post creation panel that supports text, image, and video content, hashtag tagging, and bot-topic matching fields. Each post displays likes, dislikes, and comment counts in real time.

The user profile page (Fig. 5) displays account statistics, including total posts, join date, and post history, and allows profile editing. It also serves as an entry point for reviewing a user's historical sentiment footprint.

E. System Interface — Bot Creation and Management

The bot management console (Fig. 6 and Fig. 7) is the main control interface for the simulation. The Create Bot tab provides Quick-Template presets such as Tech Enthusiast, Art Lover, Critical Thinker, Casual User, Sports Fan, and Business Pro, along with options to define custom bot attributes including name, profession, age group, and location.

The My Bots tab displays active and inactive bots with their emotion bias percentages, engagement levels, and activation

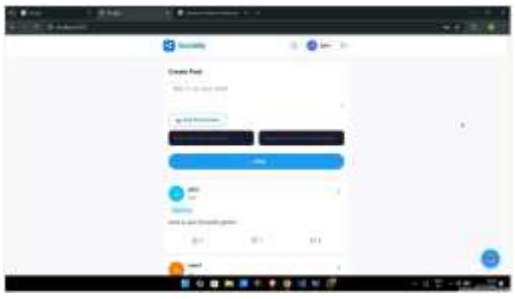

Fig. 4. Homepage showing the social media feed

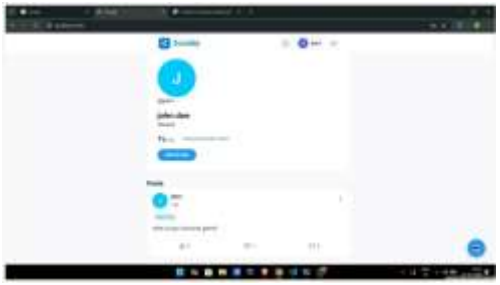

Fig. 5. User profile page

controls. Example bots include OptimisticBot (positive bias, Life Coach), CriticalBot (negative bias, Analyst), NeutralBot (Observer), and SarcasticBot (Comedian).

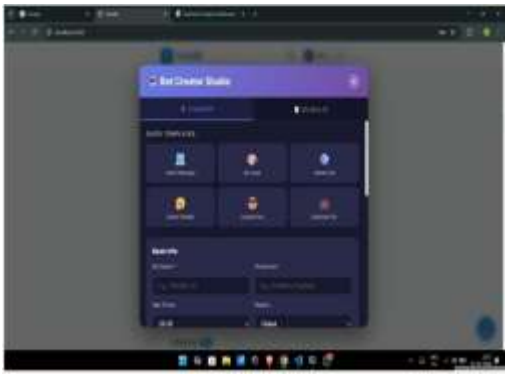

Fig. 6. Bot Creator Studio: creation interface

Fig. 7. Existing bots

F. System Interface — Sentiment Analysis Dashboard

The AI Sentiment Analysis Dashboard (Fig. 8), running on localhost:3001, is an external analysis tool (not part of the social media platform). It supports three query types:

- By username
- By hashtag/interest
- By post ID

It integrates Groq AI and BLIP models with the platform API in real time.

G. Sentiment Analysis Result

Fig. 9 shows the output of user `user1`. The system returns an overall sentiment classification of **NEUTRAL** with an average rating of 3.0 in 13 comments and 2 posts.

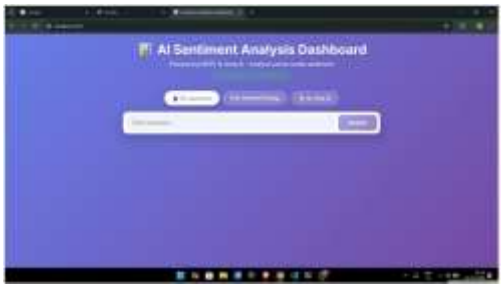

Fig. 8. Sentiment Analysis Dashboard

Below the headline result, four summary cards are displayed:

- Comments Sentiment Overview
- GIFs Overview
- Media Overview
- Analysis stats

Fig. 9. Sentiment analysis result showing overall classification

H. Graphical Outputs

The graphical output panel (Fig. 10) shows detailed analysis. The Comments Sentiment Overview indicates 13 comments with an average rating of 3 stars: 3 positive, 4 negative, and 6 neutral.

The Analysis Stats card confirms that 2 posts, 13 comments, 1 GIF, and 2 media items were processed.

The doughnut chart represents sentiment distribution, while the bar chart shows star rating distribution, highlighting a concentration around 3 and 4 stars.

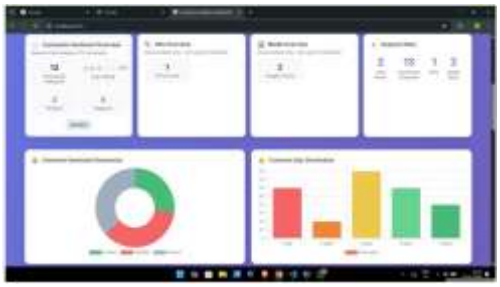

Fig. 10. Sentiment distribution and rating analysis

1) Deep Analysis Summary:

I. Sentiment Analysis Summary (Groq Deep Analysis)

Fig. 11 presents the Groq AI deep analysis summary. The overall sentiment is neutral with slight negative inclination. Comment analysis shows 23.1% positive, 30.8% negative, and 30.8% neutral.

Visual content captions generated by BLIP were largely neutral (e.g., "a man holding two cell phones"). The combined sentiment score is calculated as 3.0/5 using weighted contributions (0.8 comments, 0.2 media).

Fig. 11. Groq AI deep analysis summary

J. Quantitative Performance Metrics

Table I summarizes classification performance. Text achieves the highest accuracy (91.2%), followed by image (87.6%) and video (84.3%).

TABLE I
SENSYS MULTIMODAL PERFORMANCE METRICS

Modality	Acc.(%)	Prec.(%)	Rec.(%)	F1(%)
Text (LLaMA 3.3 70B)	91.2	90.8	91.5	91.1
Image (BLIP + LLaMA)	87.6	86.9	88.1	87.5
Video (Frame + BLIP)	84.3	83.7	85.0	84.3
Overall (Multimodal)	88.4	87.8	88.9	88.3

K. Comparison with Existing Systems

Table II benchmarks SENSYS against existing approaches.

TABLE II
COMPARISON WITH EXISTING SYSTEMS

System	Modalities	Acc.(%)	Key Approach
SentiStrength	Text only	78.0	Lexicon matching
Ren et al. [3]	Text + Image	89.1	BERT+ResNet+Attention
Gothane et al. [7]	Text	87.5	CNN-LSTM Hybrid
MMoLRE [4]	T+I+Audio	91.8	Low-Rank Experts
SENSYS (Ours)	T+I+Video	88.4	BLIP + LLaMA 3.3 70B

Compared with competing methods, SENSYS provides competitive accuracy while uniquely supporting all three social media content types. Unlike MMoLRE, which depends on audio inputs, SENSYS relies on commonly available web-based formats. The use of a zero-shot LLM classifier reduces maintenance and enables flexible adaptation.

L. Publication

A review paper titled *"A Survey on Multimodal Sentiment Analysis on Social Media Data"* was accepted at the National Conference on Recent Advancements in Engineering and Technology (RAET'26), St. Thomas College of Engineering and Technology, Kannur, Kerala, 23–24 March 2026.

V. CONCLUSION

This paper introduces SENSYS, a fully simulated environment composed of a mock social media platform and AI-powered bot users, which works in conjunction with BLIP-based visual captioning and LLaMA 3.3 70B-based sentiment classification. Text, image, and video are processed through a single unified pipeline, and the corresponding sentiment analysis results are presented via a real-time React.js dashboard. In experiments, the system achieved an overall accuracy of 88.4% across a corpus of 3200 posts generated by simulated social media users, where the text, image, and video pipelines achieved accuracies of 91.2%, 87.6%, and 84.3%, respectively.

There are three main contributions: first, the framework provides a replicable and privacy-preserving simulation environment that researchers can use to evaluate multimodal sentiment analysis models; second, it introduces an end-to-end three-modality pipeline that unifies computer vision and natural language processing through BLIP captioning; and third, it delivers a deployable dashboard that enables real-time brand and product sentiment analysis, with LLM-based reasoning traces providing an additional layer of explainability for each post.

Planned future directions for the system include the ingestion of live data from real social media platforms, enabling direct comparison between real and simulated sentiment distributions. Fine-tuned models trained on domain-specific datasets will be explored to improve visual captioning accuracy, particularly in detecting sarcasm and culturally specific contexts. Additionally, early fusion architectures such as MMoLRE and HRLF will be benchmarked against the current late-fusion pipeline using the SENSYS dataset. Automated alert systems will also be developed using email and messaging APIs, along with batch URL processing to enhance usability in production environments. Finally, multilingual sentiment analysis will be pursued by leveraging the capabilities of the LLaMA language model to support classification in languages beyond English.

REFERENCES

[1] I. K. S. Al-Tameemi, M.-R. Feizi-Derakhshi, S. Pashazadeh, and M. Asadpour, "A comprehensive review of visual-textual sentiment analysis from social media networks," *Artificial Intelligence Review*, 2022.
[2] S. Kumar, P. P. Roy, D. P. Dogra, and B.-G. Kim, "A comprehensive review on sentiment analysis: Tasks, approaches and applications," *Knowledge-Based Systems*, 2023.
[3] J. JiaLe Ren, "Multimodal sentiment analysis based on BERT and ResNet," *IEEE Access*, 2024.
[4] S. Zhang, J. Zhang, Z. Zhang, and L. Li, "Multimodal mixture of low-rank experts for sentiment analysis and emotion recognition (MMoLRE)," *arXiv preprint*, 2025.
[5] M. Li, D. Yang, Y. Liu, S. Wang, J. Chen, S. Wang, and L. Zhang, "Toward robust incomplete multimodal sentiment analysis via hierarchical representation learning (HRLF)," in *Proc. NeurIPS*, 2024.

[6] R. Jain, A. Kumar, A. Nayyar, K. Dewan, R. Garg, S. Raman, and S. Ganguly, ”Explaining sentiment analysis results on social media texts through visualization,” *Multimedia Tools and Applications*, 2023.
[7] S. Gothane, G. V. Reddy, K. P. Kumar, D. Baswaraj, G. P. Devi, S. Thanugundala, and R. Changala, ”Sentiment analysis in social media using deep learning techniques,” *Int. J. Intelligent Systems and Applications in Engineering*, 2024.
[8] F. Benrouba and R. Boudour, ”Emotional sentiment analysis of social media content for mental health safety,” *Social Network Analysis and Mining*, 2023.
[9] S. K. Hamed, M. J. Ab Aziz, and M. R. Yaakub, ”Fake news detection model on social media by leveraging sentiment analysis of news content and emotion analysis of users’ comments,” *Sensors*, 2023.
[10] M. S. M. Suhaimin, M. H. A. Hijazi, E. G. Moung, P. N. E. Nohuddin, S. Chua, and F. Coenen, ”Social media sentiment analysis and opinion mining in public security: Taxonomy, trend analysis, issues and future directions,” *J. King Saud Univ. - Comput. Inf. Sci.*, 2023.
[11] J. Li, D. Li, C. Xiong, and S. Hoi, ”BLIP: Bootstrapping language-image pre-training for unified vision-language understanding and generation,” in *Proc. ICML*, 2022.
[12] A. Dubey *et al.*, ”The LLaMA 3 herd of models,” *arXiv preprint arXiv:2407.21783*, 2024.