Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

# AdMeme: An Agentic AI Framework for Automated Meme Advertisement Generation and Virality Prediction using Vision-Language Models

Harsh Mahesh Antarkar, Pranav Karve, Ankit Yadav, Krishna Patil
Department of Artificial Intelligence and Data Science
Vivekanand Education Society's Institute of Technology
Mumbai, India

***Abstract*—This paper presents AdMeme, an agentic artificial intelligence framework designed for automated generation and evaluation of meme-based advertisements. The system integrates retrieval-augmented generation, workflow orchestration, and multimodal reasoning to enable scalable and context-aware meme creation.**

**The framework operates on a curated dataset of approximately 10,000 meme images collected from publicly available sources. To evaluate the effectiveness of generated memes, a hybrid virality prediction model is proposed, combining psychological engagement principles, visual memorability features, and content diffusion characteristics.**

**Unlike conventional approaches, the proposed system provides end-to-end automation, including caption generation, template rendering, semantic evaluation, and virality scoring. Qualitative evaluation demonstrates that the system is capable of producing coherent, contextually relevant, and marketing-oriented meme content. The results highlight the potential of agentic multimodal AI systems in digital marketing applications.**

## I. Introduction

The rise of social media platforms has significantly transformed the way information is created and shared. Among various forms of digital content, memes have emerged as a highly effective medium due to their simplicity, humor, and ability to convey messages rapidly.

In the context of marketing, memes serve as a powerful tool for engaging audiences, increasing brand visibility, and promoting viral dissemination of content. However, creating effective memes is a complex process that requires creativity, cultural awareness, and an understanding of evolving trends.

Traditional meme generation relies heavily on human intervention, making it difficult to scale content production for large-scale marketing campaigns. Additionally, evaluating the effectiveness of memes in terms of engagement and virality remains a challenging task.

To address these limitations, this paper introduces AdMeme, an agentic AI framework that automates the process of meme generation, evaluation, and ranking. The system leverages advancements in large language models and vision-language models to understand and generate multimodal content.

A key innovation of this work is the integration of a structured virality prediction model that captures multiple dimensions of meme effectiveness, including emotional impact, visual distinctiveness, and audience alignment. By combining these elements into a unified pipeline, the proposed system aims to provide a scalable solution for meme-based advertising.

## II. Contributions

The primary contributions of this work are summarized as follows:

- Design and implementation of an end-to-end automated meme advertisement generation system.
- Integration of multimodal reasoning using a fine-tuned Llama 3.2 Vision-Instruct model.
- Development of a hybrid virality prediction model incorporating 13 engagement-related parameters.
- Deployment of a real-world pipeline using n8n workflow automation, Ollama-based LLMs, and external APIs.
- Qualitative evaluation demonstrating the feasibility of automated meme creation for marketing applications.

## III. Related Work

Early research in meme analysis primarily focused on unimodal approaches. Text-based models such as BERT demonstrated strong capabilities in sentiment analysis and language understanding, while image-based models such as convolutional neural networks were effective in object recognition tasks.

However, these approaches fail to capture the interplay between textual and visual elements, which is essential for understanding meme semantics. Memes often rely on the interaction between image context and textual overlays to convey meaning, making unimodal approaches insufficient.

To address this limitation, multimodal models such as CLIP and VisualBERT were introduced. These models learn joint representations of image and text, enabling improved cross-modal understanding. Despite these advancements, early-fusion architectures often struggle with complex semantic relationships, particularly in cases involving sarcasm or contextual ambiguity.

Recent advancements in vision-language models have introduced more sophisticated mechanisms for integrating visual and textual information. Instruction-tuned models allow for contextual reasoning and interpretation of multimodal inputs,

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

making them suitable for tasks such as meme generation and evaluation.

The AdMeme framework builds upon these developments by combining multimodal reasoning with workflow automation and structured virality modeling.

## IV. Methodology

### A. Dataset

A multimodal dataset consisting of approximately 10,000 meme images was constructed using publicly available sources. The dataset includes a wide variety of meme templates and captions, covering different humor styles and contextual scenarios.

To ensure consistency, all images were resized to a fixed resolution and annotations were standardized. The dataset was split into training and testing subsets using an 80:20 ratio, enabling evaluation on unseen samples.

### B. System Implementation Pipeline

The AdMeme framework is implemented as a modular pipeline that integrates language models, workflow orchestration, and external APIs.

**Conceptualization Phase:** The system utilizes a large language model to generate context-aware meme captions based on input prompts such as product descriptions and campaign objectives. Multiple candidate captions are generated to provide diversity in outputs.

**Data Transformation Phase:** The generated captions are processed to ensure compatibility with meme templates. This involves mapping textual content to template-specific regions and validating input structure.

**Rendering Phase:** The processed captions are passed to an image rendering API, which embeds the text into predefined meme templates to generate final images.

**Post-Processing and Enrichment:** Generated meme outputs are filtered to remove unsuccessful or low-quality results. Valid outputs are prepared for further evaluation using multimodal models.

**Virality Scoring:** Each generated meme is evaluated using a structured scoring mechanism that considers contextual relevance, readability, and template effectiveness.

$$V_s = (C_w \cdot R) + (T_{score} \cdot 0.4) + (F_{match} \cdot 0.2) \quad (1)$$

### C. Virality Prediction Model

To estimate meme virality, a hybrid model is proposed that integrates psychological engagement principles with visual and contextual features.

$$\begin{aligned} V = {} & w_1 SC + w_2 TR + w_3 EM + w_4 PB + w_5 PV + w_6 ST \\ & + w_7 IM + w_8 VD + w_9 OP + w_{10} SU \\ & + w_{11} AA + w_{12} CN + w_{13} SP \end{aligned} \quad (2)$$

This formulation captures multiple dimensions influencing meme propagation, including emotional impact, visual distinctiveness, and audience alignment.

## V. Training Configuration

The multimodal model is fine-tuned using a parameter-efficient LoRA approach applied to the Llama 3.2 Vision-Instruct model. The use of LoRA enables efficient adaptation of large models without modifying all parameters.

Training is conducted using a learning rate of $2 \times 10^{-4}$, a batch size of 4, and 600 training steps. Gradient accumulation is used to improve training stability, while mixed precision training is employed to optimize computational efficiency.

## VI. Experimental Results

### A. System-Level Evaluation

The system was evaluated using real-world input scenarios to assess its ability to generate relevant and coherent meme advertisements. The pipeline successfully produces multiple outputs aligned with input context.

### B. Qualitative Results

The generated memes demonstrate strong alignment between visual templates and textual captions. This indicates that the model effectively captures multimodal relationships required for meme understanding.


Fig. 1. Sample generated meme demonstrating contextual relevance and alignment between visual and textual elements.

### C. Discussion of Results

The results highlight the capability of the system to generate context-aware meme content. The integration of virality scoring provides additional interpretability by enabling ranking of generated outputs.

## VII. Discussion and Limitations

The AdMeme framework demonstrates the feasibility of automated meme generation using multimodal AI systems. However, meme interpretation is influenced by cultural context and rapidly evolving trends, which remain challenging for automated models.

Additionally, the virality prediction model relies on heuristic parameters and does not incorporate real-world engagement data, which may affect prediction accuracy.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

## VIII. Conclusion and Future Scope

This paper presented AdMeme, an agentic AI framework for automated meme advertisement generation and evaluation. The system integrates multimodal reasoning, workflow automation, and structured virality modeling to enable scalable meme creation.

The qualitative evaluation demonstrates the effectiveness of the proposed approach in generating relevant and engaging meme content. Future work will focus on incorporating real-world engagement data, expanding datasets, and extending the framework to support video-based memes.

## References

[1] A. Radford et al., "Learning Transferable Visual Models," ICML, 2021.
[2] D. Kiela et al., "Hateful Memes Challenge," NeurIPS, 2020.
[3] E. Hu et al., "LoRA," ICLR, 2022.
[4] J. Devlin et al., "BERT," NAACL, 2019.
[5] K. He et al., "ResNet," CVPR, 2016.
[6] J. Berger, "Contagious," 2013.
[7] A. Khosla et al., "Image Memorability," ICCV, 2015.
[8] D. Watts, "Global Cascades," 2008.