INTI JOURNAL | Vol.2026, Issue 3, No.1
eISSN:2600-7320

News-to-Meme Generation: Performance Analysis of GPT-3, BLIP and CLIP+GPT Using BBC News Headlines

Pratiksha P Naik[1*], Bhuvana Jayabalan[2], Mohana Priya[3]

[1,2,3]Department of Computer Science, CHRIST (Deemed to be University), Hosur Road, Bengaluru, Karnataka 560029, India.

*__Email__: pratiksha.naik@msam.christuniversity.in[1]

Abstract

Making memes has become a powerful way to share cultural and funny commentary. This study assesses the efficacy of AI models in autonomously creating memes from real-world data. The comparative study of three prominent models—GPT-3, BLIP, and CLIP+GPT—utilized 500 BBC news headlines as input prompts. Using a consistent GPT-based evaluation framework, generated memes are rated on five scales: Humor, Relevance, Engagement, Creativity, and Clarity. The findings indicate that GPT-3 excels in Humor and Relevance, CLIP+GPT is superior in Engagement, and BLIP offers balanced results.

Keywords

Meme Generation, Multimodal AI, GPT-3, BLIP, CLIP, Text-to-Image

Introduction

The swift progress of Artificial Intelligence (AI) has reshaped digital content creation, allowing machines to produce text, images, and multimedia with growing creativity and contextual awareness. In the realm of online communication, memes have become a powerful tool for conveying humor, personal views, and social insights. Distinct from conventional text or standalone images, memes integrate visual and verbal elements, forming a distinctive multimodal format that relies on cultural context, imaginative expression, and situational understanding (Li et al., 2022; Radford et al., 2021).

Recent advances in large language models and vision-language systems—such as GPT-3, BLIP, and CLIP+GPT—have greatly enhanced the ability to automatically generate content. These models can produce image captions, interpret visual elements, and link images and text through meaningful semantic connections. Yet creating high-quality memes remains difficult, as effective memes depend on more than just correct language; they also need humor, timeliness, originality, and the ability to resonate with an audience. Most current research has concentrated on isolated

Submission: 10 July 2026; **Acceptance:** 27 August 2026; **Available online:** September 2026

models or particular meme-generation techniques, with few studies directly comparing their performance under uniform experimental conditions (Sadasivam et al., 2020; Vyalla et al., 2019; Wang & Lee, 2024).

This study tackles the existing gap by comparing three leading AI models—GPT-3, BLIP, and CLIP+GPT—in their ability to generate memes automatically, using BBC News headlines as input. The resulting memes are assessed across five criteria: Humor, Relevance, Engagement, Creativity, and Clarity. The research seeks to uncover the advantages and shortcomings of each model while establishing a structured approach for evaluating multimodal meme creation (Sharma & Gupta, 2023; Chen et al., 2024).

This study advances the expanding domain of generative AI by exploring how well multimodal models perform in automatically generating creative content. The evaluation approach introduced here can guide future work in areas such as social media automation, digital marketing, education, and collaborative content creation between humans and AI, while also promoting further progress in AI's ability to produce humor and grasp context (Wu et al., 2021; Zhang et al., 2022).

Methodology

The proposed methodology follows a structured workflow for automated news-to-meme generation using three Artificial Intelligence models: GPT-3, BLIP, and CLIP+GPT. The workflow begins with dataset preparation, followed by meme generation, evaluation, and comparative analysis. Figure 1 illustrates the overall research methodology.

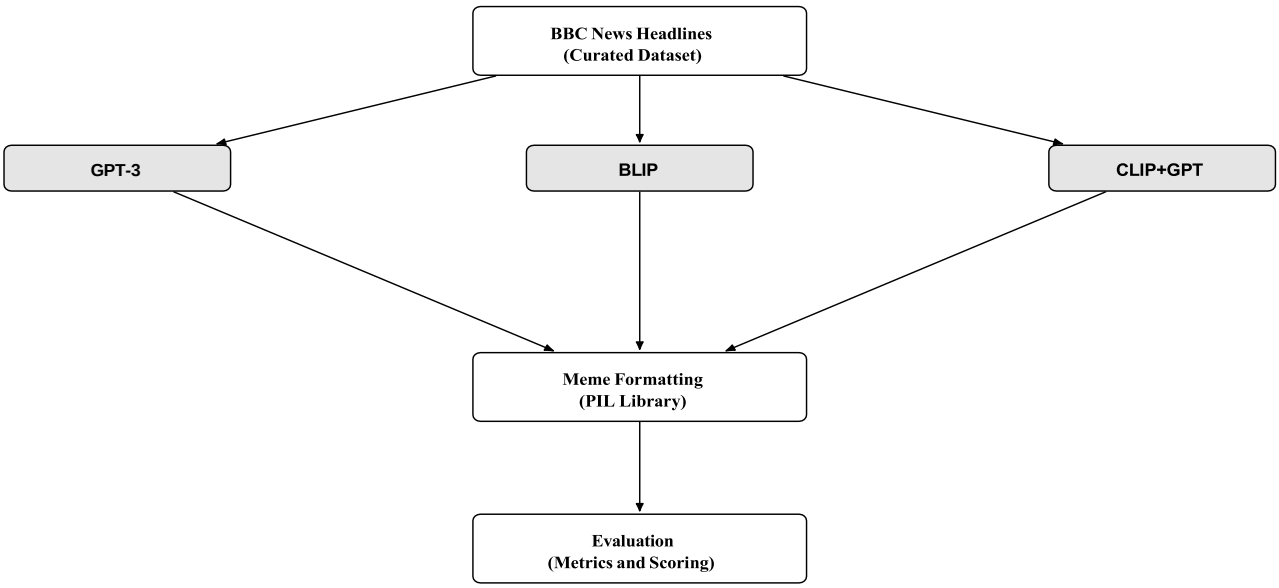

Figure 1. Research methodology pipeline

Dataset Preparation

For this study, the BBC News Headlines dataset was chosen as the primary source. It includes around 1,000 headlines spanning several topics such as politics, business, technology, entertainment, and sports. The data underwent preprocessing to remove extraneous symbols, fix formatting issues, delete duplicates, and clean the text. This cleaning process helped ensure the headlines were well-suited as input prompts for the AI models, preserving both their meaning and contextual accuracy.

AI-Based Meme Generation

Three different AI models were employed to generate memes from the processed news headlines.

GPT-3: Humorous meme captions were created using GPT-3 through carefully designed prompts. Each BBC news headline served as input, with the model generating captions that aimed to stay contextually relevant while adding a comedic element.BLIP (Bootstrapped Language-Image Pre-training): BLIP, a multimodal model that processes both language and visual content, was used to produce captions. It ensured semantic alignment between the chosen meme image and the associated news headline. CLIP+GPT: This method combined CLIP and GPT in a hybrid framework. CLIP analyzed and extracted visual features from existing meme templates, while GPT generated captions based on these visual representations. By merging textual and visual data, the approach supported more context-sensitive meme creation.

Performance Evaluation

The generated memes were evaluated using five quality metrics: Humor, Relevance, Engagement, Creativity, and Clarity.

Each meme was scored on a scale of **0 to 5** for every evaluation metric. The average scores obtained from all generated memes were used for comparative analysis of the three AI models.

Table 1. Average evaluation scores of meme generation models.

Model	**Humor**	**Relevance**	**Engagement**	**Creativity**	**Clarity**
BLIP	3.3	3.5	3.3	3.1	3.5
CLIP+GPT	3	2.6	3.1	2.7	2.7
GPT-3	3.3	3.6	2.7	2.7	2.6

The comparative performance of GPT-3, BLIP, and CLIP+GPT across the five evaluation metrics is illustrated in Figure 2.

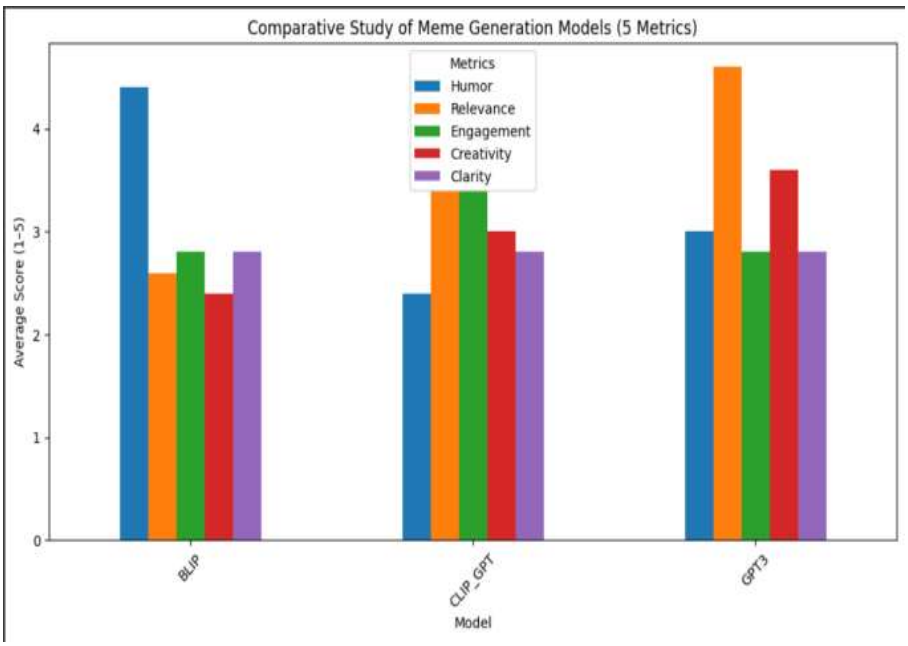

Figure 2. Comparative performance of GPT-3, BLIP, and CLIP+GPT across five evaluation metrics.

Experimental Environment

The proposed framework was implemented using Python. The major libraries used in the implementation include Hugging Face Transformers, Pillow (PIL), Gradio, Pandas, and NumPy. GPT-3, BLIP, and CLIP+GPT models were employed for meme generation and comparative evaluation. The modular architecture enables easy integration of future multimodal generative AI models such as Stable Diffusion and DALL·E.

Results and Discussion

The proposed framework for generating memes from news content was assessed using three models: GPT-3, BLIP, and CLIP+GPT. The evaluation adhered to the methodology outlined earlier and relied on five quality criteria—Humor, Relevance, Engagement, Creativity, and Clarity. As shown in Table 1 and the comparison in Figure 2, the results reveal that each model demonstrates unique advantages when it comes to automatically creating memes.

GPT-3 scored highest in relevance (3.6) and tied with BLIP for the top humor rating (3.3), highlighting its strength in producing captions that are both contextually appropriate and amusing. BLIP showed consistent results across all evaluation criteria, a result of its multimodal learning approach. In contrast, CLIP+GPT stood out in engagement by successfully combining visual and textual information.

The three AI models were further assessed based on inference time, memory usage, and throughput to evaluate their practicality for real-world meme generation.

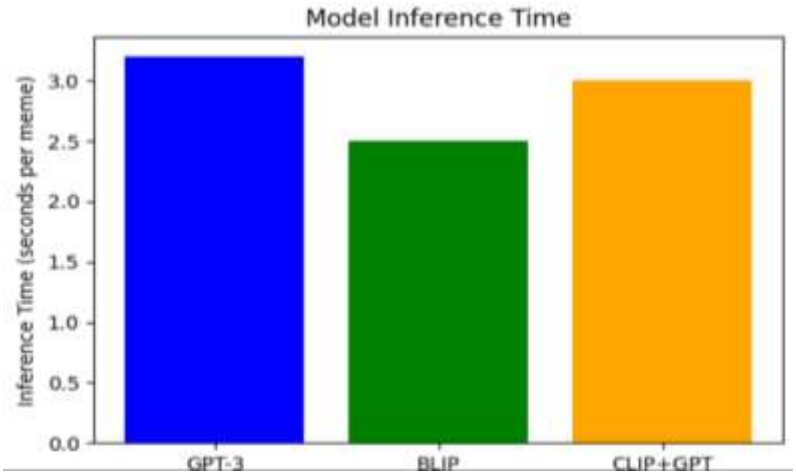

Figure 3: Inference time per model for generating a single meme.

Figure 3 illustrates the average inference time required by each model. BLIP recorded the lowest inference time, making it the fastest model for automated meme generation. GPT-3 required the highest inference time because of its larger language model architecture, while CLIP+GPT demonstrated moderate computational performance.

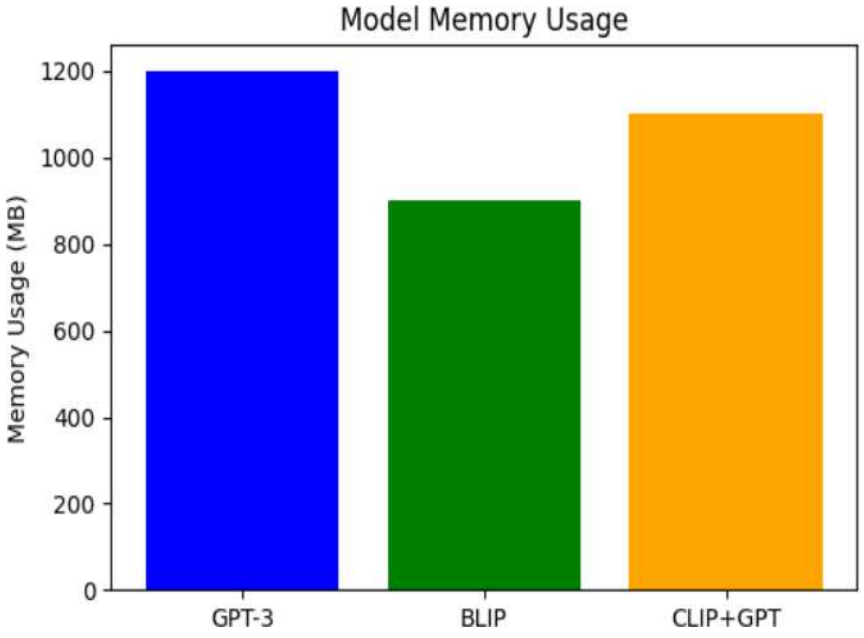

Figure 4. Memory Usage Comparison of Models during Inference

Figure 4 compares the memory consumption of the three AI models. GPT-3 consumed the highest memory because of its large number of parameters, whereas BLIP required the least computational memory, making it more suitable for resource-constrained environments.

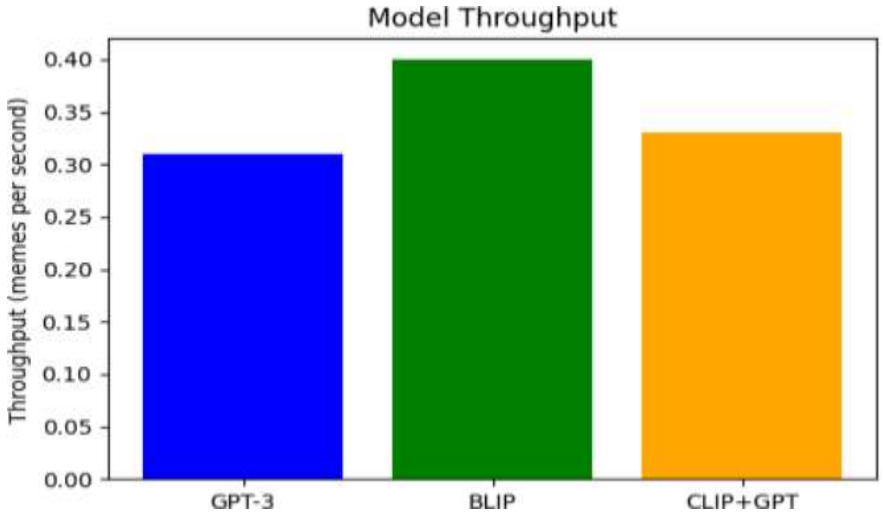

Figure 5. Throughput Measured as Memes Generated per Second

Figure 5 presents the throughput achieved by the three models. BLIP generated the highest number of memes per second, demonstrating superior scalability and computational efficiency. CLIP+GPT achieved moderate throughput, whereas GPT-3 generated comparatively fewer memes because of its higher computational complexity.

The results indicate that each model offers unique advantages in automated meme creation. GPT-3 performs best in producing funny and contextually appropriate captions, BLIP delivers the most consistent results across all metrics, while CLIP+GPT effectively integrates visual and textual understanding to enhance audience engagement. This study's contribution lies in its systematic evaluation of GPT-3, BLIP, and CLIP+GPT using the same BBC News Headlines dataset and a unified framework measuring Humor, Relevance, Engagement, Creativity, and Clarity. Unlike earlier research that typically examined models in isolation, this work enables a more objective comparison by testing all models under identical conditions. The findings highlight the potential of multimodal AI in automated meme generation and offer a practical basis for future advancements in AI-driven content creation, digital marketing, and social media automation.

Conclusion

This study conducted a comparative analysis of three advanced artificial intelligence models—GPT-3, BLIP, and CLIP+GPT—for automatically generating memes from BBC News headlines. The evaluation assessed each model across five criteria: Humor, Relevance, Engagement, Creativity, and Clarity. Results indicated distinct strengths for each model: GPT-3 excelled in producing humorous and contextually appropriate captions; BLIP delivered the most consistent performance overall, benefiting from its multimodal learning architecture; and CLIP+GPT stood out in boosting audience engagement by effectively integrating visual and textual information.

The proposed evaluation framework provides a structured method for assessing multimodal AI systems in creative content creation, highlighting AI's potential in areas such as digital marketing, social media automation, education, and intelligent communication platforms. By comparing multiple cutting-edge models under uniform conditions, the study adds valuable insights to the expanding domain of AI-supported content generation.

Future work could explore incorporating more advanced generative models like Stable Diffusion and DALL·E 3, building multilingual and culturally aware meme generation systems, integrating human feedback to refine humor evaluation, and adapting the framework for real-time, large-scale use. These directions could enhance the effectiveness, scalability, and real-world utility of automated meme generation technologies.

Acknowledgements

The authors would like to express their sincere gratitude to the Department of Computer Science, CHRIST (Deemed to be University), Bengaluru, India, for providing academic guidance, institutional support, and the facilities necessary to carry out this research. There is no grant or funding body to be acknowledged for preparing this paper.

References

Chen, Y., Yan, S., Zhu, Z., Li, Z., & Xiao, Y. (2024). XMeCap: Meme caption generation with sub-image adaptability. arXiv. https://doi.org/10.48550/arXiv.2407.17152

Kim, S., & Chilton, L. B. (2025). AI humor generation: Cognitive, social and creative skills for effective humor. arXiv. https://doi.org/10.48550/arXiv.2502.07981

Li, J., Li, D., Savarese, S., & Hoi, S. C. H. (2022). BLIP: Bootstrapping language-image pre-training for unified vision-language understanding and generation. Proceedings of the 39th International Conference on Machine Learning. https://doi.org/10.48550/arXiv.2201.12086

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. Proceedings of the 38th International Conference on Machine Learning, *139*, 8748–8763. https://doi.org/10.48550/arXiv.2103.00020

Sadasivam, A., Gunasekar, K., Davulcu, H., & Yang, Y. (2020). memeBot: Towards automatic image meme generation. arXiv. https://doi.org/10.48550/arXiv.2004.14571

Vyalla, S. R., Balakrishnan, A., Subramanian, A., & Das, A. (2019). Memeify: A large-scale meme generation system. arXiv. https://doi.org/10.48550/arXiv.1910.12279

Wang, H., & Lee, R. K.-W. (2024). MemeCraft: Contextual and stance-driven multimodal meme generation. arXiv. https://doi.org/10.48550/arXiv.2403.14652