Computational Creativity in Meme Generation: A Multimodal Approach

José P. Lopes, João M. Cunha and Pedro Martins
University of Coimbra,
Centre for Informatics and Systems of the University of Coimbra, Department of Informatics Engineering
{joselopes,jmacunha,pjmm}@dei.uc.pt

Abstract

In this paper, we explore computational creativity in internet meme generation, employing a multimodal framework that integrates natural language processing, image generation and visual understanding. Specifically, we utilise Llama2, a language model trained on 7 billion parameters, FAISS vector storage, and LangChain to process user input and select meme formats from a database. By leveraging FAISS and LangChain, we contextualise the input to facilitate template selection and generate descriptions of meme concepts. We then employ Stable-Diffusion1.5 and ControlNet for image-to-image generation, utilising predefined image macros linked to chosen templates, and using a generated image description as prompt. Lastly we use a multimodal model capable of visual and language understanding, Llava1.5, to create captions for the generated memes using the input from the first step, and the concept given by the Llama2 model. We show some ouptuts generated by the proposed system, discuss its strengths and weaknesses, and outline potential avenues for future research and improvement.

Introduction

Coined by Dawkins (1976), the term "meme" derives from the Greek word "mimema", indicating "something imitated". Much like genes propagate through reproductive cells, memes spread through imitation, moving from one mind to another. They are also susceptible to variation and distortion which leads to the creation of entirely new and unique combinations of existing ideas, beliefs, and expressions (Wiggins and Bowers 2015). Internet memes exemplify this evolutionary process, harnessing the internet's power to disseminate jokes, rumours, videos, and websites from person to person (Milner 2012; Soleh et al. 2021). This idea is reinforced by the tendency for internet users to follow existing memes and adapt them to convey their own message, despite their freedom to create memes in various ways, as we can see in Figure 1. Such patterns can be considered as "meme genres", which are defined as "socially recognised types of communicative action" (Shifman 2014). Genres share not only structures and stylistic elements, but also topics and target audiences, meaning that crafting and comprehending memes may demand some previous "meme literacy" (Shifman 2014).

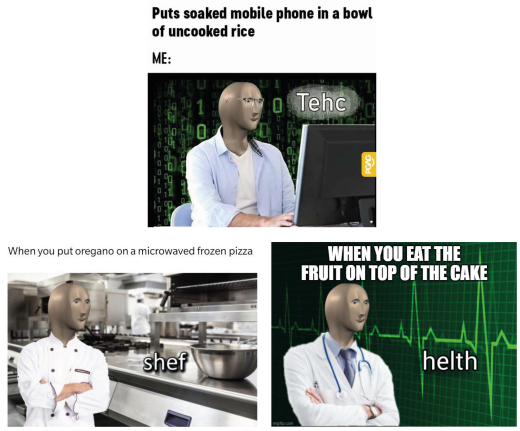

Figure 1: Example of memes in the Stonks format (Lopes et al., 2023)

Using these ideas as a foundation, we propose an approach for meme generation that, based on existing image macros, generates new images and captions for memes, following the memetic evolution concept of variation and distortion. We explored internet memes in computational creativity, by developing a modular proof of concept system, running on a local machine, using state-of-the-art generative models that can create and interpret text and images. Through the integration of open source models such as LLaMa-2 (Andrei 2024; Gerganov 2024; Touvron et al. 2023), Stable-Diffusion (Luo et al. 2023; Rombach et al. 2022), ControlNet (Zhang, Rao, and Agrawala 2023), and LLaVa-1.5 (Andrei 2024; Liu et al. 2024), we introduce a modular approach to internet meme generation that leverages the principles of memetic evolution, embracing mutation and variation. Our system operates by receiving textual input and selecting from a pool of 18 meme templates to generate a meme related to the input. To generate the meme image itself, we use an image macro related to the chosen meme template as a starting point. This approach serves as a guide to keep the generated image within the template format. Lastly, a different module writes a caption that aligns with both the generated image and the input text. Additionally, the system generates a meme explanation, offering im-

portant context that helps both users and subsequent system modules in completing their respective tasks. This approach not only makes meme creation accessible to everyone, but may also allow experienced meme-makers to leverage their "meme knowledge" and get inspiration for new and unique ideas. Additionally, since meme comprehension may demand some "meme literacy", as previously stated, our system's explanation can help inexperienced users understand the generated meme.

Memes have been studied in different fields, but the majority of projects only prioritise two main aspects: meme template selection and caption generation, with image generation in meme creation being overlooked. Our approach contributes with an an artificial creative system that leverages the strengths of each generative model through explainability and communication between models.

Related Work

This section explores the field of meme generation and generative models. Internet memes propagate through imitation and evolve through variation, resulting in novel expressions of ideas and beliefs. They serve as participatory media, reflecting societal mindsets and fostering user engagement.

In our examination of meme generation systems, we discovered that the majority of projects prioritise two main aspects: meme template selection and subsequent caption generation. Additionally, while some projects leverage caption templates to seamlessly integrate customised information into an image macro caption, others prefer to generate captions specifically made for custom images. We can observe this in Vyalla and Udandarao (2020) and Wang et al. (2021), which both present a meme dataset and a meme generation system with a similar approach, since both use a convolutional neural network for image feature extraction, although using different models, and a transformer-based model, both using GPT-2, for caption generation. Here Vyalla and Udandarao (2020) focus on image caption generation according to predefined meme templates, while Wang et al. (2021) generate chinese captions for any kind of image. Liu et al. (2021) propose a different system, one that determines specific emotions in different inputs, leveraging text-emotion classifiers, to select macro images and uses a sequence-to-sequence network to convert input sentences into captions for memes. In a different project, Shimomoto et al. (2019) use a database containing images, catchphrases from different media, and news articles, where words and images are matched by representing them as sets of word vectors through the word2vec representation, enabling information retrieval across different media formats and generating the final output. Continuing with meme generation from news headlines,

Image generation in automatic meme generators is still relatively unexplored. Lopes, Cunha, and Martins (2023) propose a pipeline system composed of different modules, each with its own tasks. The system can only produce memes in one format, the Stonks format, but this limitation is leveraged by allowing the definition of a set of rules which are followed to produce the meme. In another approach, Chen et al. (2019) leverage Generative Adversarial Networks (GANs) with an attention module to generate meme-faces related to textual inputs, by training a model on a dataset of meme-faces and meme templates information collected from online sources. Using a multimodal framework, Wang and Lee (2024) leverage large language models and visual language models to create a meme generator, presenting an end-to-end pipeline that creates memes without manual intervention and advocates to specific social movements.

By approaching meme generation as a multimodal task, as in the work by Liu et al. (2021) and Wang and Lee (2024), we wanted to create a modular system to create internet memes, such as the Stonkinator (Lopes, Cunha, and Martins 2023), that could be updated overtime with new state-of-the-art models that could successfully complete the different tasks executed by our system. As it was done in Shimomoto et al. (2019), Vyalla and Udandarao (2020) and Wang et al. (2021), we also used a predefined database containing image templates and information about them to guide our modules in the completion of their tasks. Lastly, following the idea in Chen et al. (2019), we used an image-to-image approach to change the original image and create a new one that still fits the meme template.

Meme Generator Approach

This section of the paper presents the development of our meme generation system. Firstly, we report how we selected specific meme templates used by our system. We also describe the meme generator's development process and its pipeline. We will be using Figure 4 as an example, where we used two types of input prompt, one more objective and one more subjective. This approach enables us to evaluate the system's ability to operate in contexts varying in degrees of creative freedom, from more open-ended (subjective prompts) to more constrained environments (objective prompts).

Meme template selection

To develop our system we used different pre-trained open source models. First, we used Llama-2 (Touvron et al. 2023), a large language model (LLM) that stands out for its ability to generate text and answer questions. We were able to incorporate LangChain (Chase 2022) into the system workflow, a framework to help enhance LLMs by using different components from several modules to "chain" prompts and give them the needed context. We transformed the meme templates information into embeddings, using FAISS (Douze et al. 2024), and stored in a vector store. This allowed us to give the necessary context for the model to select a meme template and generate an image description and meme explanation related to the chosen template.

For the meme template selection, we also used an online internet meme database called "KnowYourMeme", where we obtained the available information about the first 35 most viewed memes in the "Confirmed Meme Entries"[1] section. From these 35 memes, we removed memes with a NSFW tag (Not Safe For Work) and memes representing unique

[1]https://knowyourmeme.com/categories/meme?sort=views

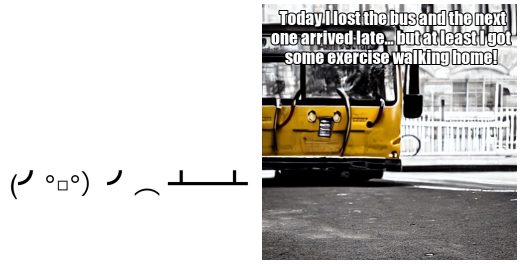

Figure 2: Example of output from an emoticon with the original image the left and the generated output on the right

concepts that do not have a specific image or image element that represents them. Later, while testing the system, we also found that memes represented by unicode character symbols do not work using our approach, therefore, we also removed emoticons (Figure 2). In total, our system is capable of generating and explaining memes in 18 templates.

In Figure 4, for the input prompts "A long road" and "When you play bowling for the first time and win", our system chose the templates "Big Chungus", which is used to humorously exaggerate the size or importance of something, and "Overly Attached Girlfriend", which humorously depict clingy or possessive behaviour in relationships, respectively.

System pipeline

To generate memes, we divided our system into three main modules, each focused on a specific task: the concept generation, the image generation, and the caption generation, as seen in Figure 3. For the concept generation we used three models: LLaMa-2 trained with 7 billion parameters (Andrei 2024; Gerganov 2024; Touvron et al. 2023), LangChain (Chase 2022) and FAISS (Douze et al. 2024). Using LangChain's "WebBaseLoader" function, we download information about the selected meme templates from the online database and we divide that information into chunks with a LangChain function, creating and storing embeddings in a vector store using FAISS. With the use of LangChain and FAISS, the system is capable of incorporating the necessary context in the input prompt, allowing it to use custom data in the selection of the meme template and in the explanation of the meme. It also allows us to request the source documents, which we can use to help debug any problems we may encounter. With this module the system is capable of selecting a meme template, generate an image description, and create an explanation of the conceptualised meme. For each prompt in Figure 4, the generated image descriptions were, respectively, "An image of Big Chungus sitting on a long, winding road with a disapproving expression on his face." and "A cartoonish representation of the Overly Attached Girlfriend character with a bowling ball and strike symbol on her forehead, giving a manic grin while holding a trophy or balloons.", and the generated meme explanations were "Used to humorously exaggerate the length or difficulty of a journey." and "Represents feeling of excitement or elation after achieving something unexpectedly, such as winning at bowling for the first time.".

This information is then used by the image generation

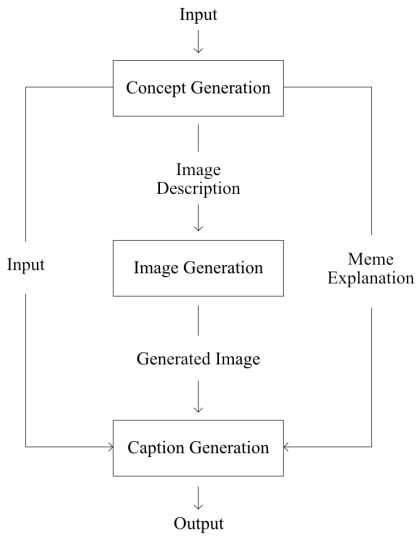

Figure 3: System framework

module. We employed a Stable Diffusion model[2] (Luo et al. 2023; Rombach et al. 2022) converted to run efficiently on local Apple Silicon computer (Orhon, Siracusa, and Wadhwa 2022). The model conversion repository[3] also supports the conversion of ControlNet (Zhang, Rao, and Agrawala 2023) models, which we used[4] to guide our system into generating memes based on the selected image macro. The model (Rombach et al. 2022) obtains an image description from Llama-2, which then produces an image with aid from ControlNet (Zhang, Rao, and Agrawala 2023) to guide the image generation and maintain the meme template structure.

Lastly, the generated image is sent to the final module, which completes the image caption task. We use a LLaVa-1.5 (Andrei 2024; Liu et al. 2024) model trained on 7 billion parameters[5] to receive the image, the input prompt, and the meme explanation. LLaVa-1.5 is a multimodal model capable of processing visual and textual data for tasks such as image captioning and answering questions based on images and text. To write the generated caption on the top of the generated image, we use OpenCV (Bradski 2000) and Pillow (Clark 2015), writing it with the Impact font, widely used in internet meme images (Brideau and Berret 2014).

Discussion

The research outlined in this paper explores computational creativity in internet meme generation, utilising a framework that integrates natural language processing, image generation, and visual understanding. This section discusses the importance of the proposed approach, places it in the context of existing literature, and suggests potential areas for future research and improvement.

The proposed method introduces a modular system for internet meme generation that aligns with the principles of

[2]https://huggingface.co/runwayml/stable-diffusion-v1-5
[3]https://github.com/apple/ml-stable-diffusion
[4]https://huggingface.co/lllyasviel/ControlNet
[5]https://huggingface.co/mys/ggml_llava-v1.5-7b

Figure 4: Output images for the prompts “A long road” and “When you play bowling for the first time and win”

memetic evolution, emphasising mutation and variation. By integrating open source generative models and tools such as LLaMa-2, LangChain, FAISS, Stable-Diffusion, ControlNet, and LLaVa-1.5, the system simplifies meme creation, providing users with a platform to generate memes according to their input prompts. Previous projects focused on meme template selection and caption generation tasks, and while these efforts have laid the groundwork for understanding meme creation processes, the omission of image generation in meme creation is notable. The proposed approach addresses this gap by offering a meme generation system that encompasses both textual and visual modalities, drawing inspiration from prior works, while following the principles of memetic evolution. These principles are demonstrated through the integration of Stable Diffusion and ControlNet, which allows the system to create new images from existing meme templates, as we can see[6] in Figures 4 and 5.

Another strength in our approach is its modular design, offering flexibility and adaptability to improve individual components of the system. By decomposing the process into distinct modules, we enable an easier integration of new advancements and updates, providing users with the tools to enhance specific aspects of meme generation as necessary. Moreover, our system operates locally, leveraging open-source models without relying on external APIs. This not only ensures user privacy and data security but also provides users with the freedom to choose between open-source and closed-source models based on their preferences.

While the proposed approach shows promise, there are avenues for future research and improvement. Firstly, in Figures 4 and 5, we can observe some bias in meme template selection towards the same three meme templates. In Figure 4 we may also notice that, although the final image and caption seem coherent, the selected template could be better suited. Additionally, adopting heavier models, particularly for image generation tasks, could enhance the system’s ability to produce higher quality meme images. This could also lead to a better meme template selection. Lastly, further research into co-creative approaches for meme generation systems could unlock new possibilities for collaborative meme creation, enabling users to participate actively in the creative process. Moreover, ongoing advancements in generative modelling techniques offer opportunities to refine and optimise the system’s performance, ensuring its continued relevance in the landscape of internet memes.

[6] github.com/ze-lopes/memeGeneratorMultimodalApproach

Figure 5: More output images from the system

Conclusion

This paper presents a new approach to internet meme generation, guided by the principles of memetic evolution. Leveraging a modular framework and state-of-the-art generative models to facilitate meme creation, the current system runs on local machines using open-source generative models, which ensures user privacy and data security, while its modular design gives the user freedom of choice on which models to use, providing users with the tools to enhance specific aspects of meme generation as necessary.

Consisting of three different modules, our system selects meme templates, generates new images guided by those meme templates, and creates captions for the generated meme. Our system also generates an image description and meme explanation, which can be leveraged by the system modules themselves, but also to help users understand the generated meme.

Despite the promising aspects of our approach, there are opportunities for further research and improvement. Addressing biases in meme template selection, enhancing image generation quality through the adoption of heavier models, and exploring co-creative approaches for meme generation systems are possible paths to improve the existing approach.

Acknowledgements

This work is funded by the FCT - Foundation for Science and Technology, I.P./MCTES through national funds (PIDDAC), within the scope of CISUC R&D Unit - UIDB/00326/2020 or project code UIDP/00326/2020 and also supported by the Portuguese Recovery and Resilience Plan (PRR) through project C645008882-00000055, Center for Responsible AI.

References

Andrei. 2024. abetlen/llama-cpp-python. original-date: 2023-03-23T09:30:33Z.

Bradski, G. 2000. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.

Brideau, K., and Berret, C. 2014. A Brief Introduction to Impact: 'The Meme Font'. *Journal of Visual Culture* 13(3):307–313. Publisher: SAGE Publications.

Chase, H. 2022. LangChain.

Chen, Y.; Wang, Z.; Wu, B.; Li, M.; Zhang, H.; Ma, L.; Liu, F.; Feng, Q.; and Wang, B. 2019. MemeFaceGenerator: Adversarial Synthesis of Chinese Meme-face from Natural Sentences. arXiv:1908.05138 [cs].

Clark, A. 2015. Pillow (PIL Fork) Documentation.

Dawkins, R. 1976. *The selfish gene*. New York: Oxford University Press.

Douze, M.; Guzhva, A.; Deng, C.; Johnson, J.; Szilvasy, G.; Mazaré, P.-E.; Lomeli, M.; Hosseini, L.; and Jégou, H. 2024. The Faiss library. _eprint: 2401.08281.

Gerganov, G. 2024. ggerganov/llama.cpp. original-date: 2023-03-10T18:58:00Z.

Liu, Z.; Sun, C.; Jiang, Y.; Jiang, S.; and Ming, M. 2021. Multi-modal application: Image Memes Generation. arXiv:2112.01651 [cs].

Liu, H.; Li, C.; Li, Y.; Li, B.; Zhang, Y.; Shen, S.; and Lee, Y. J. 2024. LLaVA-NeXT: Improved reasoning, OCR, and world knowledge.

Lopes, J. e. P.; Cunha, J. a. M.; and Martins, P. 2023. Stonkinator: An Automatic Generator of Memetic Images. In *Proceedings of the 14th International Conference on Computational Creativity, ICCC 2023, Waterloo, June 19-23, 2023*.

Luo, S.; Tan, Y.; Huang, L.; Li, J.; and Zhao, H. 2023. Latent Consistency Models: Synthesizing High-Resolution Images with Few-Step Inference. arXiv:2310.04378 [cs].

Milner, R. M. 2012. The World Made Meme: Discourse and Identity in Participatory Media. Accepted: 2012-10-28T15:14:25Z Publisher: University of Kansas.

Orhon, A.; Siracusa, M.; and Wadhwa, A. 2022. Stable Diffusion with Core ML on Apple Silicon.

Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; and Ommer, B. 2022. High-Resolution Image Synthesis with Latent Diffusion Models. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10674–10685. New Orleans, LA, USA: IEEE.

Shifman, L. 2014. *Memes in digital culture*. The MIT Press essential knowledge series. Cambridge, Massachusetts: The MIT Press.

Shimomoto, E. K.; Souza, L. S.; Gatto, B. B.; and Fukui, K. 2019. News2meme: An Automatic Content Generator from News Based on Word Subspaces from Text and Image. In *2019 16th International Conference on Machine Vision Applications (MVA)*, 1–6. Tokyo, Japan: IEEE.

Soleh, M. B.; Anisa, Y. H.; Absor, N. F.; Suswandari; and Edison, R. E. 2021. Differences of Visual Attention to Memes: An Eye Tracking Study:.

Touvron, H.; Martin, L.; Stone, K.; Albert, P.; Almahairi, A.; Babaei, Y.; Bashlykov, N.; Batra, S.; Bhargava, P.; Bhosale, S.; Bikel, D.; Blecher, L.; Ferrer, C. C.; Chen, M.; Cucurull, G.; Esiobu, D.; Fernandes, J.; Fu, J.; Fu, W.; Fuller, B.; Gao, C.; Goswami, V.; Goyal, N.; Hartshorn, A.; Hosseini, S.; Hou, R.; Inan, H.; Kardas, M.; Kerkez, V.; Khabsa, M.; Kloumann, I.; Korenev, A.; Koura, P. S.; Lachaux, M.-A.; Lavril, T.; Lee, J.; Liskovich, D.; Lu, Y.; Mao, Y.; Martinet, X.; Mihaylov, T.; Mishra, P.; Molybog, I.; Nie, Y.; Poulton, A.; Reizenstein, J.; Rungta, R.; Saladi, K.; Schelten, A.; Silva, R.; Smith, E. M.; Subramanian, R.; Tan, X. E.; Tang, B.; Taylor, R.; Williams, A.; Kuan, J. X.; Xu, P.; Yan, Z.; Zarov, I.; Zhang, Y.; Fan, A.; Kambadur, M.; Narang, S.; Rodriguez, A.; Stojnic, R.; Edunov, S.; and Scialom, T. 2023. Llama 2: Open Foundation and Fine-Tuned Chat Models.

Vyalla, S. R., and Udandarao, V. 2020. Memeify: A Large-Scale Meme Generation System. In *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*, 307–311. Hyderabad India: ACM.

Wang, H., and Lee, R. K.-W. 2024. MemeCraft: Contextual and Stance-Driven Multimodal Meme Generation. arXiv:2403.14652 [cs].

Wang, L.; Zhang, Q.; Kim, Y.; Wu, R.; Jin, H.; Deng, H.; Luo, P.; and Kim, C.-H. 2021. Automatic Chinese Meme Generation Using Deep Neural Networks. *IEEE Access* 9:152657–152667.

Wiggins, B. E., and Bowers, G. B. 2015. Memes as genre: A structurational analysis of the memescape. *New Media & Society* 17(11):1886–1906.

Zhang, L.; Rao, A.; and Agrawala, M. 2023. Adding Conditional Control to Text-to-Image Diffusion Models. Publication Title: IEEE International Conference on Computer Vision (ICCV).