中山醫學大學
Chung Shan Medical University

IEEE

iFUZZY 2023

Cognitive-Inclusive Communication: An Integrated Text and Graphics Generator for Improved Readability

Presenter：Yong-Xiang Chang

Authors：Yong-Xiang Chang, Chiun-Li Chin, Zhong-Ting Zhemg, Geng-Kai Wong, Su-Juan Chen, Guan-Tsen Liu, Pei-Hsin Chang, Che-Cheng Liu, Pei-Chen Huang, Pei-Xin Ye

Outline

01 Introduction

02 Related Works

03 Method

04 Results

05 Conclusion

01

Introduction

Introduction(1/3)

When it comes to individuals with intellectual disabilities, reading often poses challenges due to their restricted reading abilities.

For these individuals, all we need is an accessible publication that can serve as a valuable resource.

- Easy-to-read publications would be a door-opener and a useful training resource.

Introduction(2/3)

In another part, we found that 「the Transformer Model」 is used in the fields of natural language processing and computer vision.

We assumed that it could combine text and images by using multi-modal processing capabilities, thereby enhancing the readability and expressiveness of information.

Introduction(3/3)

We selected LLaMA 2 13B to be the Transformer framework.

LLaMA 2 is superiority compared to other large-scale language models, and LLaMA 2's open-source nature and adjustable model features make it suitable for integration with diffusion models designed to generate images from text.

02

Related Works

Related Works (1/3)

In 2010

Nomura et al. mentions that easy-to-read publications should have a solid scientific foundation and be able to learn from new research findings.

This research could include different disciplines like linguistics and education, as well physical and intellectual or cognitive disabilities.

- Through adding Artificial Intelligence with computer science, it could enable easy-to-read to advance and contribute to another subject area.

Related Works (2/3)

In 2021

Ilya et al. show that while convolutions and attention are both sufficient for good performance, neither of them are necessary.

In 2022

Rombach et al. turn diffusion model into powerful and flexible generators for general conditioning inputs such as text or bounding boxes and high-resolution synthesis becomes possible in a convolutional manner by introducing cross-attention layers into the model architecture.

In 2023

Touvron et al. develop and release Llama 2, a collection of pretrained and fine-tuned large language models (LLMs) ranging in scale from 7 billion to 70 billion parameters.

Related Works (3/3)

We mainly propose a novel way by introducing MLP, the overall process will be optimized and followed end-to-end.

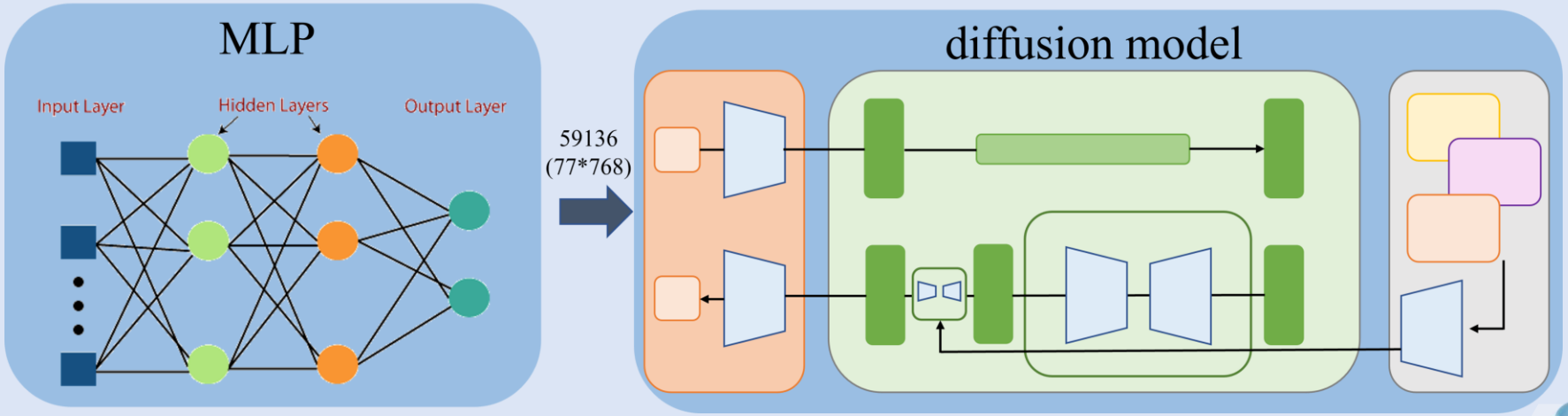
- Our method will through an efficient innovative way, by inserting MLP Layers between LLaMA 2 and Diffusion Model to increase generative image's effective and accuracy.

03

Method

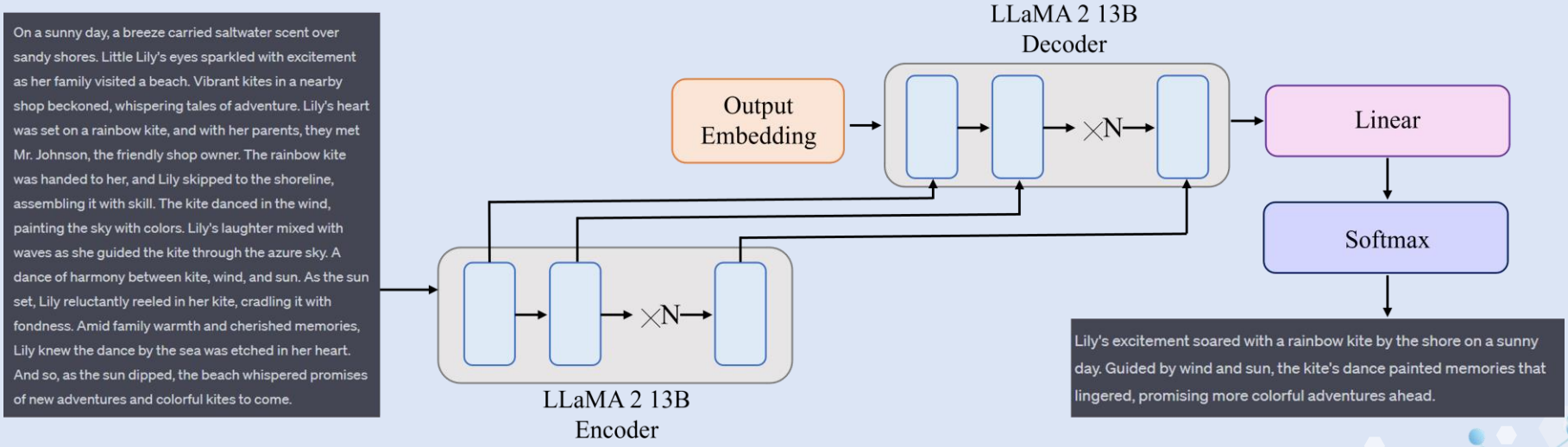
Method(1/4)

Our target is to achieve the simplification of lengthy text in an end-to-end manner. This process will include diminishing the dimensionality through MLP layer, resulting in a size appropriate for input into the diffusion model.



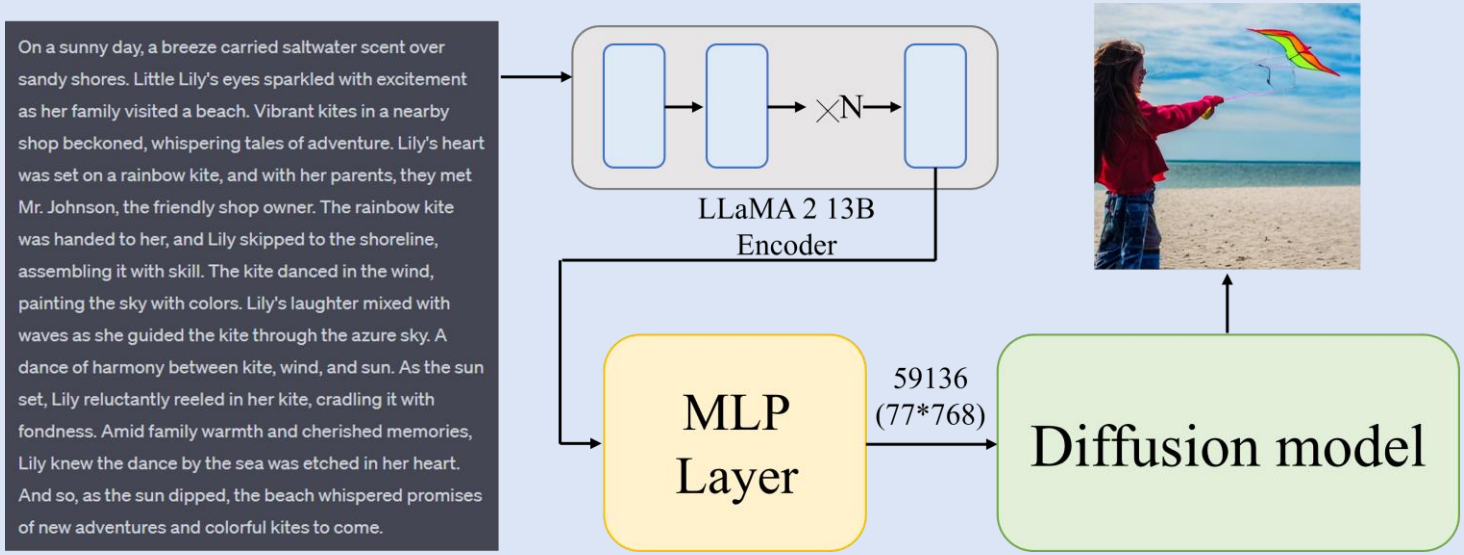
Method(2/4)

Our task is to present an image that best corresponds to the original text to achieve an easy-to-read effect. The lengthy text as input and opt to utilize the LLaMA 2 13B model as our foundational model.



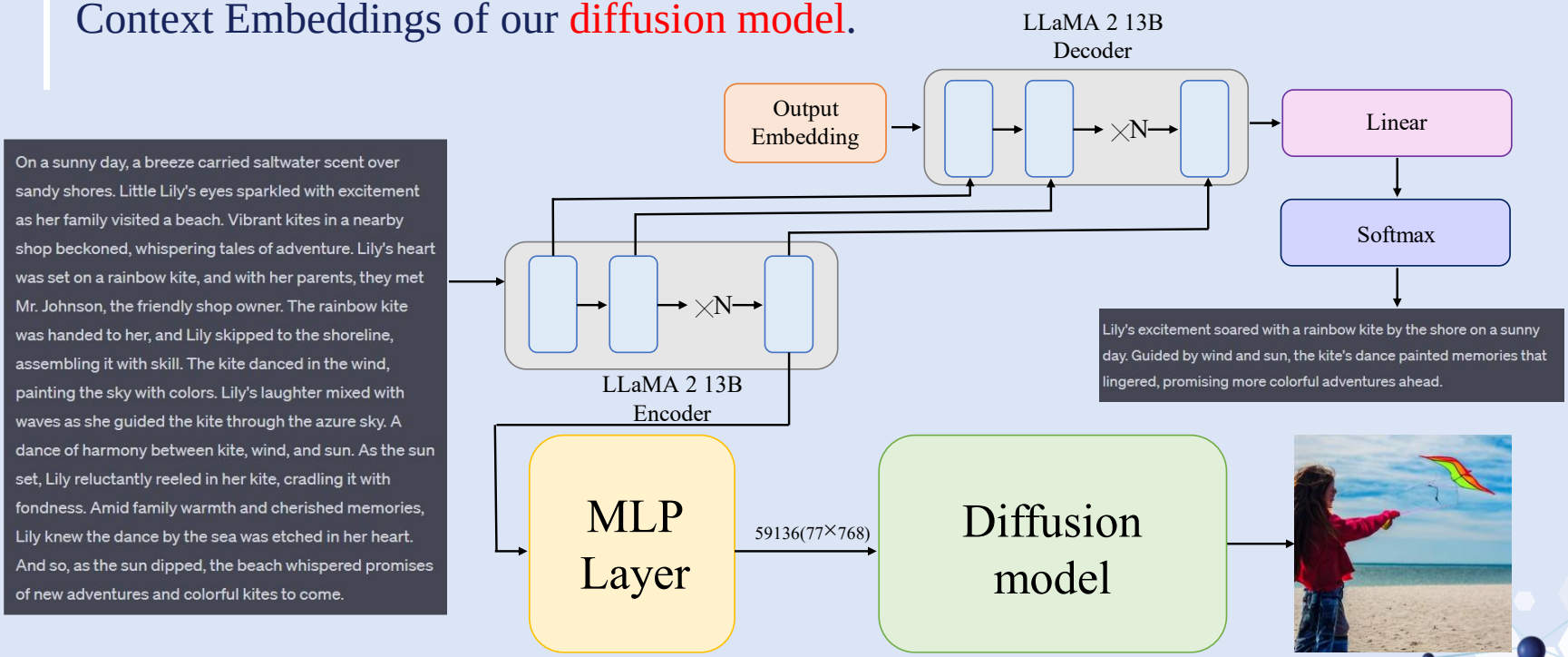
Method(3/4)

The results of the last layer of the encoder are taken out and fed into the MLP layer. The core focus of MLP layer is for reducing the last attention normalization output of LLaMA 2 to 59136(77×768) text embeddings.



Method(4/4)

This represents the closest feature vector to the concise text that matches the Context Embeddings of our diffusion model.



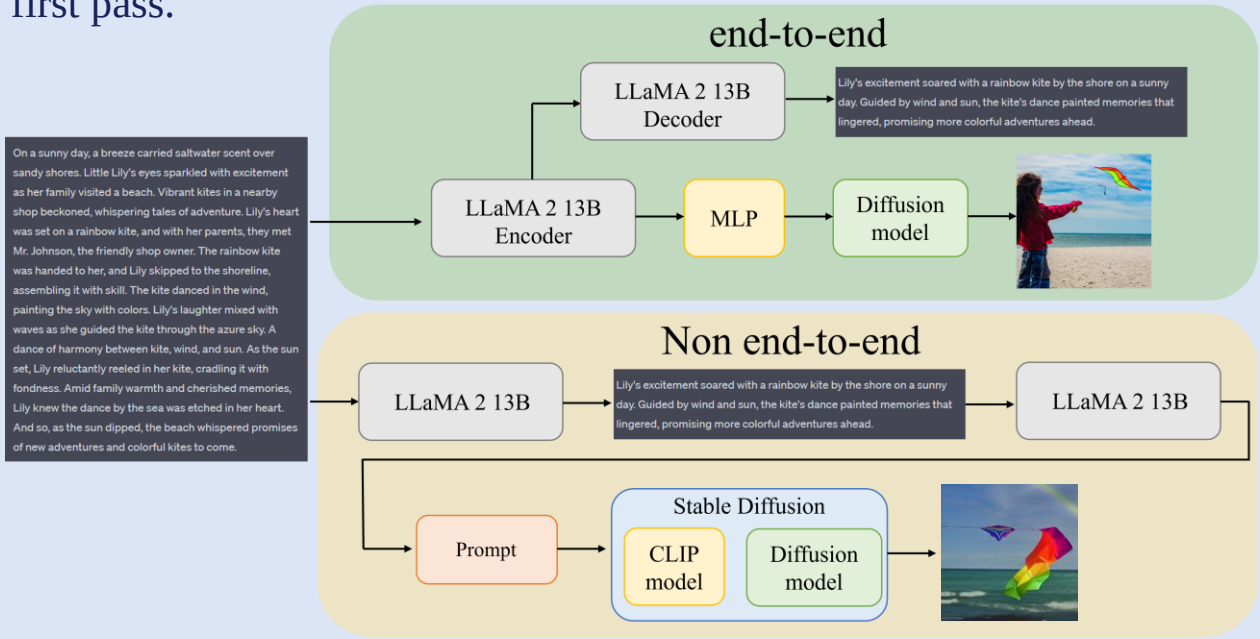
04

Result

Result(1/2)

There are two different ways to accomplish the same lengthy text to generate an image.

- The first approach is using our above subject, which achieves the end-to-end process.
- The second approach involves subjecting the lengthy text to two passes of the LLaMA 2 13B model to generate a prompt, which stems from the concise text produced during the first pass.



Lengthy text

On a sunny day, a breeze carried saltwater scent over sandy shores. Little Lily's eyes sparkled with excitement as her family visited a beach. Vibrant kites in a nearby shop beckoned, whispering tales of adventure. Lily's heart was set on a rainbow kite, and with her parents, they met Mr. Johnson, the friendly shop owner. The rainbow kite was handed to her, and Lily skipped to the shoreline, assembling it with skill. The kite danced in the wind, painting the sky with colors. Lily's laughter mixed with waves as she guided the kite through the azure sky. A dance of harmony between kite, wind, and sun. As the sun set, Lily reluctantly reeled in her kite, cradling it with fondness. Amid family warmth and cherished memories, Lily knew the dance by the sea was etched in her heart. And so, as the sun dipped, the beach whispered promises of new adventures and colorful kites to come.

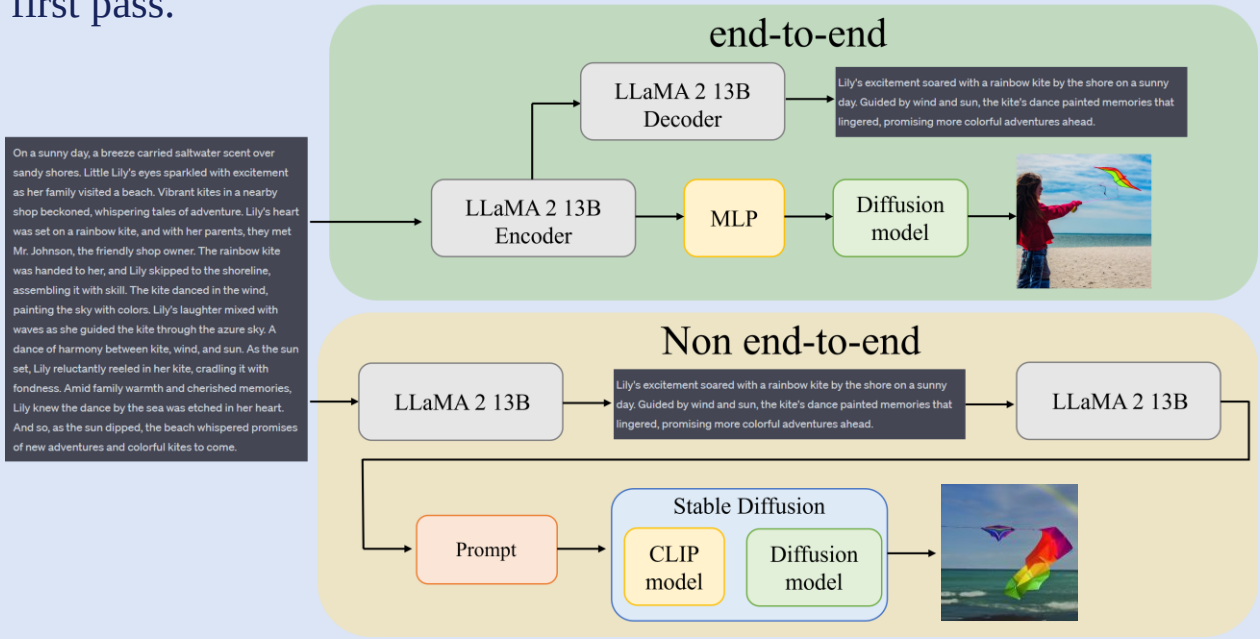
Concise text

Lily's excitement soared with a rainbow kite by the shore on a sunny day. Guided by wind and sun, the kite's dance painted memories that lingered, promising more colorful adventures ahead.

Result(1/2)

There are two different ways to accomplish the same lengthy text to generate an image.

- The first approach is using our above subject, which achieves the end-to-end process.
- The second approach involves subjecting the lengthy text to two passes of the LLaMA 2 13B model to generate a prompt, which stems from the concise text produced during the first pass.



Result(2/2)

This prompt is then fed into the stable diffusion model to generate corresponding images.

- It would submit the images generated in both ways to multiple rounds of blind testing and assessment by image recognition experts.

The evaluation process will focus on the degree of correspondence between the images generated by the two methods and the original concise text. The non-end-to-end way is much inferior matching due to probability weight distribution.

Opinion statics table		
Result \ Approach	end-to-end	Non end-to-end
Effectiveness	83.28%	67.71%
Accuracy	90.63%	77.18%
Degree of loss	9.55%	16.35%
Image-comprehensive	87.36%	72.64%

05

Conclusion

Conclusion

In this study, the primary objective to approach cognitive-inclusive communication combines transformer models and image generation to enhance accessibility for individuals with intellectual disabilities. Through streamlined processes and the integration of MLP layer, it achieves easy-to-read functions.

Different from the existing non-end-to-end methods, it could connect the graph with concise text very well. Our proposed model superiority over the diffusion model in blind testing underscores its potential in generating images closely aligned with concise text.

—iFUZZY 2023—

THANK YOU!